Journal of Official Statistics, Vol. 15, No. 2, 1999, pp. 305–327

Calibration as a Standard Method for Treatment of Nonresponse

Sixten Lundström[1] and Carl-Erik Särndal[2]

This article suggests a simple and unified approach to the use of auxiliary information for reducing both the sampling error and the nonresponse bias in a survey. We propose a point estimator based on calibration and a corresponding variance estimator. They are general in regard to both the sampling design and the form of the auxiliary information. The calibration procedure generates final weights which are as close as possible to specified initial weights (the design weights), while respecting known auxiliary population totals or unbiased estimates of these totals. When population totals are used, the resulting point estimators are consistent in the sense that the final weights give perfect estimates when applied to each auxiliary variable. A clear tendency in our empirical findings is that an increased auxiliary information content will reduce both variance and nonresponse bias of the point estimator. Despite some residual bias, the coverage rate of our confidence intervals comes close to a nominal 95%.

Key words: Nonresponse adjustment; nonresponse bias; nonresponse variance; auxiliary information; variance estimation.

1. Introduction

Both producers and users of statistics are aware that nonresponse can impair the quality of the estimates, and therefore considerable resources are spent on data collection procedures aiming at preventing nonresponse from occurring. Nevertheless, we have to accept some missing values. Since we can be practically certain that the nonresponse is not the result of a simple random selection mechanism, we try to adjust for the selection bias at the estimation stage. Furthermore, we know that the nonresponse creates an additional component of variance.

The literature provides two standard methods for treating nonresponse: imputation and weighting. In the first method, missing values are replaced by proxies; in the second, the design weights are multiplied by adjustment weights aiming at nonresponse bias reduction. This article is concerned with weighting.

To fix ideas we introduce some notation. Consider the finite population of N elements $U = \{1, \ldots, k, \ldots, N\}$. We wish to estimate the total $Y = \Sigma_U y_k$, where y_k is the value of a typical study variable, y, for the kth element. For short, Σ_A will be used for $\Sigma_{k \in A}$, where $A \subseteq U$ is an arbitrary set. Let s be a sample of size n drawn from U with the probability $p(s)$. The inclusion probabilities are then $\pi_k = \Sigma_{s \ni k} p(s)$ and $\pi_{kl} = \Sigma_{s \ni \{k,l\}} p(s)$. Let $d_k = 1/\pi_k$ denote

[1] Statistics Sweden, S-70189 Örebro, Sweden.
[2] Statistics Canada, Ottawa, Ontario, K1A0T6 Canada and Statistics Sweden, S-70189 Örebro, Sweden.
Acknowledgments: The authors are grateful to three anonymous referees and an associate editor for helpful comments which led to improvements in the original manuscript.

the *design weight* of element k, and let $d_{kl} = 1/\pi_{kl}$. However, nonresponse occurs, and the response set r of size m is obtained, where $r \subseteq s$ and $m \leq n$.

Strong auxiliary information is a prerequisite for a successful reduction of both the sampling error and the nonresponse bias. We assume that there exists an auxiliary vector, $\mathbf{x}$, containing such strong information. Its value for the kth element is denoted $\mathbf{x}_k$. We define the two ''information levels,'' called Info-S and Info-U, to be examined in the following sections:

i) Info-S: $\mathbf{x}_k$ is known for all $k \in s$

and

ii) Info-U: $\Sigma_U \mathbf{x}_k$ is known and moreover $\mathbf{x}_k$ is known for all $k \in s$.

In case (ii) the information goes ''up to the population level'' and is more extensive than in case (i), where it goes ''up to the sample level'' only.

Before presenting our approach we comment on some conventional methods. Commonly a two-phase approach is used, where the response mechanism is considered as the second phase. Let us for the time being assume that the response distribution $q(r|s)$ is known with the corresponding known response probabilities denoted $\Pr(k \in r|s) = \theta_k$ and $\Pr(k\&l \in r|s) = \theta_{kl}$ (we assume these probabilities to be independent of the realized sample s). Under these conditions Särndal, Swensson, and Wretman (1992) suggest, in their Chapter 9 on two-phase sampling, the following estimators at Info-S and Info-U, respectively:

$$\hat{Y}_{SSW,s\theta} = \Sigma_r d_k g_{sk\theta} y_k / \theta_k \tag{1.1}$$

where

$$g_{sk\theta} = 1 + q_k(\Sigma_s d_k \mathbf{x}_k - \Sigma_r d_k \mathbf{x}_k/\theta_k)'(\Sigma_r d_k q_k \mathbf{x}_k \mathbf{x}_k'/\theta_k)^{-1}\mathbf{x}_k \tag{1.2}$$

and

$$\hat{Y}_{SSW,U\theta} = \Sigma_r d_k g_{Uk\theta} y_k / \theta_k \tag{1.3}$$

where

$$g_{Uk\theta} = 1 + q_k(\Sigma_U \mathbf{x}_k - \Sigma_r d_k \mathbf{x}_k/\theta_k)'(\Sigma_r d_k q_k \mathbf{x}_k \mathbf{x}_k'/\theta_k)^{-1}\mathbf{x}_k \tag{1.4}$$

However, in practice the response probabilities are never known and so they have to be replaced by some proxies $\hat{\theta}_k$. Recognizing this, the literature suggests that a relevant response model be first formulated, then that its unknown response probabilities be estimated. The conventional estimator of Y is then

$$\hat{Y} = \Sigma_r d_k v_{1k} v_{2k} y_k \tag{1.5}$$

where $v_{1k} = 1/\hat{\theta}_k$ and v_{2k} is equal to the g-weights given either by (1.2) or by (1.4), where θ_k is replaced by $\hat{\theta}_k$.

Most of the response models considered so far in the literature are simple, for example that the given sample s can be partitioned into groups such that all elements within the same group are assumed to have the same response probability. Särndal, Swensson, and Wretman (1992) call these groups *response homogeneity groups*, (RHGs). Usually in practice the weights v_{2k} are also simple, as when they are poststratification weights.

The commonly used process is well expressed by Kalton and Kasprzyk (1986, p. 4): ''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.'' Other references in the same spirit are Binder, Michaud, and Poirier (1994) and Singh, Wu, and Boyer (1995).

Examples of more complex response models include, for example those in Politz and Simmons (1949) and Thomsen and Siring (1983). Some authors estimate the response probabilities from logistic regression models, as in Little (1986) and Ekholm and Laaksonen (1991), where the dependent variable is dichotomous (response, nonresponse) and the explanatory variables are selected from the set of available auxiliary variables.

In the Scandinavian countries, much information is available from registers; for example, Statistics Sweden's registers for individuals contain variables such as sex, age, nationality, income, education and field of occupation, and also geographical variables (address). All these data are known at the population level. Moreover, the unique personal identification number makes it easy to transfer data to the response set. However, little of this abundance of information seems to be used in practice, probably because the traditional methods are viewed by many as time-consuming, complex and otherwise inconvenient. Thus, to promote increased use of available auxiliary information it is essential to offer the users a wide choice of simple and convenient, but still effective, methods within a common framework. It should be possible within such a framework to access a wide class of estimators via a well-designed computer software. We present in this article an approach that is in our opinion simpler than the best of the current possibilities, but equally effective. In our approach a weight which corresponds to the product of v_{1k} and v_{2k} is calculated in one single step. No explicit response model is needed, this in contrast to the conventional methods, where some auxiliary information is used in modeling and deriving the v_{1k} and some in deriving the v_{2k} in (1.5). We include all the relevant auxiliary information in the vector $\mathbf{x}_k$, with the dual purpose of reducing both the sampling error and the nonresponse bias.

The general features of our approach are as follows. After having specified the auxiliary information, we compute *calibrated weights*, denoted w_k, and construct the estimator $\hat{Y}_w = \Sigma_r w_k y_k$ of Y. We call it the *w-estimator*. The weights w_k are ''as close as possible'' to the d_k, and they also satisfy a *calibration equation* given for Info-S by

$$\Sigma_r w_k \mathbf{x}_k = \Sigma_s d_k \mathbf{x}_k \tag{1.6}$$

and for Info-U by

$$\Sigma_r w_k \mathbf{x}_k = \Sigma_U \mathbf{x}_k \tag{1.7}$$

In Section 2 we describe the w-estimators for each of the two information levels. In Section 3 we develop variance estimators for the w-estimators. These take into account the increased variance caused by nonresponse. In Section 4 an expression for the nonresponse bias is presented and discussed. We derive in Section 5 the expression taken by the general point estimators for four specifications of the auxiliary vector corresponding to common types of auxiliary information. In Section 6 we conduct an empirical analysis.

An advantage with our approach is that it requires only a simple yet general one-step procedure. Users (nonspecialists in statistical science) can easily carry out the computations by means of well-designed computer software. At Statistics Sweden, the general purpose software CLAN is currently being extended to also manage auxiliary information in the second phase of a regression estimator for two-phase sampling. Hence, in the near future it will be possible to use CLAN to compute the variance estimators proposed in this paper for surveys with nonresponse.

2. The Point Estimators Derived From Calibration

Calibration estimators in the full response case are described in Deville and Särndal (1992). They seek an estimator of the form $\hat{Y}_{DS} = \Sigma_s w_k^o y_k$ with weights w_k^o as close as possible to the design weights d_k while respecting the calibration equation $\Sigma_s w_k^o \mathbf{x}_k = \Sigma_U \mathbf{x}_k$. They discuss the merits of different metrics for the distance between w_k^o and d_k (see also Dupont 1994).

In this article we use the calibration technique, but since y_k-values are observed for the response set only, rather than for the full sample, we seek new weights w_k that satisfy the calibration equation (1.6) or (1.7). The distance function to be minimized is

$$\Sigma_r (w_k - d_k)^2 / d_k q_k \tag{2.1}$$

where the q_k are specified positive factors. In the case of full response ($r = s$), this distance function leads to the generalized regression estimator (see Expression (6.4.1) in Särndal, Swensson, and Wretman 1992).

The proof of the following proposition follows easily by the Lagrange multiplier method.

Proposition 2.1: *For Info-U, minimization of the distance (2.1) under the constraint (1.7) leads to the w-estimator*

$$\hat{Y}_{wU} = \Sigma_r d_k v_{Uk} y_k \tag{2.2}$$

where

$$v_{Uk} = 1 + q_k \left(\Sigma_U \mathbf{x}_k - \Sigma_r d_k \mathbf{x}_k \right)' \left(\Sigma_r d_k q_k \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \mathbf{x}_k \text{ for } k \in r \tag{2.3}$$

If the auxiliary total $\Sigma_U \mathbf{x}_k$ is unknown, we can instead calibrate on the unbiased estimate $\Sigma_s d_k \mathbf{x}_k$. In this case the Lagrange multiplier method leads to the following proposition:

Proposition 2.2: *For Info-S, minimization of the distance (2.1) under the constraint (1.6) leads to the w-estimator*

$$\hat{Y}_{ws} = \Sigma_r d_k v_{sk} y_k \tag{2.4}$$

where

$$v_{sk} = 1 + q_k \left(\Sigma_s d_k \mathbf{x}_k - \Sigma_r d_k \mathbf{x}_k \right)' \left(\Sigma_r d_k q_k \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \mathbf{x}_k \text{ for } k \in r \tag{2.5}$$

One can argue, for both of these calibrations, that $\{d_k : k \in r\}$ is inappropriate as a set of initial weights, since they are on average too small. In many cases this is, however, not a serious flaw, as the following proposition shows.

Proposition 2.3: *For Info-U, suppose that*

(i) final weights w_k, $k \in r$, are sought to satisfy the calibration equation (1.7);
(ii) the distance to minimize is

$$\Sigma_r (w_k - d_k^*)^2 / d_k^* q_k \tag{2.6}$$

where the d_k^* are arbitrary initial weights;

(iii) $q_k = 1/\mu' \mathbf{x}_k$ *for all k* (2.7)

where μ is a column vector of the same dimension as $\mathbf{x}_k$ and independent of k.

Let c be an arbitrary positive constant. Then the initial weights $\{d_k^*; k \in r\}$ and the initial weights $\{c d_k^*; k \in r\}$ give exactly the same final weights, namely

$$w_k = d_k^* q_k \left(\Sigma_U \mathbf{x}_k'\right) \left(\Sigma_r d_k^* q_k \mathbf{x}_k \mathbf{x}_k'\right)^{-1} \mathbf{x}_k \tag{2.8}$$

For Info-S, (1.7) is replaced by (1.6). Assuming that (ii) and (iii) remain unchanged, the final weights are

$$w_k = d_k^* q_k \left(\Sigma_s d_k^* \mathbf{x}_k'\right) \left(\Sigma_r d_k^* q_k \mathbf{x}_k \mathbf{x}_k'\right)^{-1} \mathbf{x}_k \tag{2.9}$$

irrespective of the choice of c.

The proof of Proposition 2.3 is given in the Appendix.

Remark 2.1: Suppose the population is divided into P groups $U_1, \ldots, U_p, \ldots, U_P$ and that the group total $\Sigma_{U_p} \mathbf{x}_k$ is known for $p = 1, \ldots, P$, and used in the calibration. Let c_p, $p = 1, \ldots, P$, be arbitrary positive constants. Then the initial weights $d_k^* = d_k$ and the initial weights $d_k^* = c_p d_k$, for $k \in r_p$, give exactly the same final weights when (ii) is minimized and (iii) holds. The proof of this is similar to that of Proposition 2.3. The same conclusion can be drawn for Info-S when $\Sigma_{s_p} d_k \mathbf{x}_k$, $p = 1, \ldots, P$, is used in the calibration equations.

3. Variance Estimation

3.1. The mean squared error of the calibrated estimator

We start by examining the mean squared error (MSE). Let $\hat{Y}_w$ stand for the point estimator, so that $\hat{Y}_w$ is either $\hat{Y}_{ws}$ or $\hat{Y}_{wU}$, depending on whether the information level is S or U. Both levels are covered simultaneously in the following discussion. Further, let $\hat{Y}_s$ denote the estimator implied by $\hat{Y}_w$ if all elements in the sample s respond, that is, when we set $r = s$ in (2.2) or in (2.4). It is easy to see that $\hat{Y}_{ws}$ reduces to the (unbiased) Horvitz-Thompson estimator

$$\hat{Y}_s = \Sigma_s d_k y_k \tag{3.1}$$

and that $\hat{Y}_{wU}$ reduces to the (approximately unbiased) generalized regression estimator

$$\hat{Y}_s = \Sigma_s d_k g_k y_k \tag{3.2}$$

where

$$g_k = 1 + q_k\left(\Sigma_U \mathbf{x}_k - \Sigma_s d_k \mathbf{x}_k\right)'\left(\Sigma_s d_k q_k \mathbf{x}_k \mathbf{x}_k'\right)^{-1}\mathbf{x}_k \tag{3.3}$$

A decomposition of the MSE is given in the following proposition.

Proposition 3.1: *Jointly under the sampling design p(s) and the response distribution* $q(r|s)$, *the mean squared error of* $\hat{Y}_w$ *is*

$$MSE_{pq}\left(\hat{Y}_w\right) = V_{SAM} + V_{NR} + 2Cov_p\left(\hat{Y}_s, B_{NR|s}\right) + E_p\left(B_{NR|s}^2\right) \tag{3.4}$$

where $V_{SAM} = V_p(\hat{Y}_s)$ *is the sampling variance,* $V_{NR} = E_p V_q(\hat{Y}_w|s)$ *is the nonresponse error variance,* $B_{NR|s} = E_q(\hat{Y}_w - \hat{Y}_s|s)$ *is the nonresponse bias (conditionally on s), and* $Cov_p(\hat{Y}_s, B_{NR|s})$ *is the covariance of* $\hat{Y}_s$ *and* $B_{NR|s}$ *under the sampling design.*

The proof of this proposition is given in the Appendix.

Remark 3.1: If the condition

$$B_{NR|s} = 0 \text{ for all } s \tag{3.5}$$

is verified, then (3.4) becomes

$$V_{pq}(\hat{Y}_w) = V_{pq}^0(\hat{Y}_w)$$

where

$$V_{pq}^0(\hat{Y}_w) = V_{SAM} + V_{NR} \tag{3.6}$$

There exists virtually no survey such that the condition (3.5) is exactly satisfied. Inevitably, whenever there is nonresponse, some bias is introduced. But calibration on strong auxiliary information may go a long way toward eliminating the conditional nonresponse bias $B_{NR|s}$, and when this is the case, the inferences (confidence intervals, and so on) made by acting as if (3.5) is true will still be approximately valid. We shall work under the assumption that (3.5) holds approximately, so that $V_{pq}(\hat{Y}_w) \approx V_{pq}^0(\hat{Y}_w) = V_{SAM} + V_{NR}$.

Our approach will consist in estimating each of the two components of $V_{pq}^0(\hat{Y}_w)$, V_{SAM} and V_{NR}, and to use the sum of the two component estimates, $\hat{V}_{pq}^0(\hat{Y}_w)$, as an estimator of $V_{pq}(\hat{Y}_w)$. As for 95% confidence intervals, we propose to use

$$\left[\hat{Y}_w - 1.96\sqrt{\hat{V}_{pq}^0(\hat{Y}_w)},\ \hat{Y}_w + 1.96\sqrt{\hat{V}_{pq}^0(\hat{Y}_w)}\right]$$

The real confidence level of this interval is not exactly 95%, for three contributing reasons: (a) $\hat{Y}_w$ is not an unbiased point estimator, (b) $\hat{V}_{pq}^0(\hat{Y}_w)$ is not an unbiased variance estimator, and (c) the normal distribution assumption that motivates the score 1.96 holds at best approximately. However, if both $\hat{Y}_w$ and $\hat{V}_{pq}^0(\hat{Y}_w)$ have modest bias, and the sample size is not too small, the suggested confidence interval will be approximately valid at the 95% level. (This is also confirmed by the simulation results in Table 4.)

The complexity (except in the very simplest cases) of our point estimators $\hat{Y}_{ws}$ and $\hat{Y}_{wU}$ makes it difficult to obtain (even approximate) expressions for their variances. It is even less evident how variance estimates should be constructed. Nevertheless, our objective is to obtain variance estimators for $\hat{Y}_{ws}$ and $\hat{Y}_{wU}$ that will work reasonably well. Moreover, they should be simple to calculate. Finally, we impose the condition that the variance estimators must be model free: they must be constructed without recourse to a nonresponse mechanism model. A strength of our approach is precisely the fact that point estimators are derived without appealing to such models.

In order to obtain variance estimators for $\hat{Y}_{ws}$ and $\hat{Y}_{wU}$ we shall exploit a similarity that these point estimators have with the two-phase theory point estimators $\hat{Y}_{SSW,s\theta}$ and $\hat{Y}_{SSW,U\theta}$, given by (1.1) and (1.3), respectively.

3.2. A useful analogy with estimation under two-phase sampling

Särndal, Swensson, and Wretman (1992) suggested variance estimators for $\hat{Y}_{SSW,s\theta}$ and $\hat{Y}_{SSW,U\theta}$. Let us adapt these to the situation with sampling followed by nonresponse, assuming for simplicity that the sample elements respond independently, so that

$$\Pr(k\&l \in r|s) = \theta_{kl} = \theta_k\theta_l \text{ for all } k \neq l \tag{3.7}$$

Under this assumption the suggested variance estimator for $\hat{Y}_{SSW,s\theta}$, equation (9.7.28) in Särndal, Swensson, and Wretman (1992), gives, after some algebra,

$$\hat{V}(\hat{Y}_{SSW,s\theta}) = \Sigma_r\Sigma(d_k d_l - d_{kl})\left(\frac{y_k}{\theta_k}\right)\left(\frac{y_l}{\theta_l}\right) - \Sigma_r d_k(d_k - 1)\left(\frac{y_k}{\theta_k}\right)^2(1-\theta_k) + \Sigma_r d_k^2(1-\theta_k)\left(\frac{g_{sk\theta}e_{k\theta}}{\theta_k}\right)^2 \tag{3.8}$$

where $\Sigma_r\Sigma$ denotes the double sum over $k \in r$ and $l \in r$, $g_{sk\theta}$ is given by (1.2),

$$e_{k\theta} = y_k - \mathbf{x}_k'\mathbf{B}_{r\theta} \tag{3.9}$$

and $\mathbf{B}_{r\theta} = (\Sigma_r d_k q_k \mathbf{x}_k \mathbf{x}_k'/\theta_k)^{-1}\Sigma_r d_k q_k \mathbf{x}_k y_k/\theta_k$.

Under the assumption (3.7), the variance estimator for $\hat{Y}_{SSW,U\theta}$, derived from equation (9.7.22) in Särndal, Swensson, and Wretman (1992), becomes

$$\hat{V}(\hat{Y}_{SSW,U\theta}) = \Sigma_r\Sigma(d_k d_l - d_{kl})(g_k e_{k\theta}/\theta_k)(g_l e_{l\theta}/\theta_l) - \Sigma_r d_k(d_k-1)\left(\frac{g_k e_{k\theta}}{\theta_k}\right)^2(1-\theta_k) + \Sigma_r d_k^2(1-\theta_k)\left(\frac{g_{sk\theta}e_{k\theta}}{\theta_k}\right)^2 \tag{3.10}$$

where g_k is given by (3.3), $g_{sk\theta}$ by (1.2) and $e_{k\theta}$ by (3.9).

Let $\hat{Y}_{SSW\theta}$ stand for either $\hat{Y}_{SSW,s\theta}$ or $\hat{Y}_{SSW,U\theta}$, depending on whether the information level is S or U. Then we can write the variance of $\hat{Y}_{SSW\theta}$ as

$$V(\hat{Y}_{SSW\theta}) = V_{SAM} + E_p V_q(\hat{Y}_{SSW\theta}|s) \tag{3.11}$$

where V_{SAM} is the variance of the estimator to which $\hat{Y}_{SSW\theta}$ reduces under full response. Note that V_{SAM} is exactly the same as in (3.6) because $\hat{Y}_{SSW\theta}$ and $\hat{Y}_w$ reduce to the same expression under full response, which can be seen from the following. The estimator implied by (1.1) when all units respond with probability $\theta_k = 1$, is the unbiased Horvitz-Thompson estimator, $\Sigma_s d_k y_k$, with the variance $V_{SAM} = V_p(\Sigma_s d_k y_k)$. Recall that under full response, $\hat{Y}_{ws}$ also reduces to $\Sigma_s d_k y_k$; see (3.1). The corresponding full response estimator, implied by (1.3) when all sampled units respond with probability $\theta_k = 1$, is the generalized regression estimator $\Sigma_s d_k g_k y_k$, where g_k is given by (3.3). Its variance is $V_{SAM} = V_p(\Sigma_s d_k g_k y_k)$. Recall that under full response, $\hat{Y}_{wU}$ also reduces to $\Sigma_s d_k g_k y_k$, as noted in (3.2).

However, the second components in (3.6) and (3.11) differ, because $\hat{Y}_{SSW\theta}$ and $\hat{Y}_w$ are constructed differently. The estimation of the nonresponse variance component V_{NR} in (3.6) would be simple if $\hat{Y}_{SSW\theta}$ and $\hat{Y}_w$ were identical, but since they are not, a special procedure, outlined in Sections 3.3 and 3.4, is required. (Note that although (3.11) is written as an equality, a slight approximation is actually involved. This is because the expectation over the second phase of $\hat{Y}_{SSW\theta} - \hat{Y}_s$, conditionally on the first phase sample s, is not exactly zero but a close approximation to zero.) Aided by the theory just reviewed, we now construct the desired variance estimators, first for $\hat{Y}_{ws}$ in Section 3.3, then for $\hat{Y}_{wU}$ in Section 3.4.

Remark 3.2: We can expect (3.8) and (3.10) to give some underestimation of the variance. This is a well-known feature in formulas such as these which are derived by Taylor linearization. Part of the reason is that the ideal ''population residuals'' are replaced with ''sample-based residuals'' (see Särndal, Swensson, Wretman 1992, p. 363). This entails a loss of degrees of freedom, and the resulting underestimation can be significant, particularly for smallish samples. In Section 3.3, we propose variance estimators patterned on (3.8) and (3.10) and such that $e_{k\theta}$ is replaced by $f_k e_{k\theta}$, where f_k adjusts for the number of degrees of freedom lost when parameters are estimated. In Section 6, we propose f_k-values that are appropriate for each of the four different auxiliary vectors $\mathbf{x}_k$ examined in that section.

3.3. A variance estimator for $\hat{Y}_{ws}$

Assuming that $B_{NR|s} \approx 0$ for all s, we now develop a variance estimator for $\hat{Y}_{ws}$. That is, we seek to estimate $V^0_{pq}(\hat{Y}_{ws})$ given by (3.6). This variance is determined jointly by the known sampling design $p(s)$ and the unknown response mechanism $q(r|s)$ with its unknown response probabilities θ_k. The analogue of $\hat{Y}_{ws}$ in the two-phase approach is $\hat{Y}_{SSW,s\theta}$. A variance estimator for $\hat{Y}_{SSW,s\theta}$, under condition (3.7), is given by (3.8), but it cannot be used for $\hat{Y}_{ws}$ as it stands. Not only are $\hat{Y}_{SSW,s\theta}$ and $\hat{Y}_{ws}$ different estimators, but in addition the θ_k in (3.8) are unknown. If (3.8) is to be of any help at all, we must first resolve the difficulty with the unknown θ_k. To this end, we argue as follows: If $\hat{Y}_{ws}$ and $\hat{Y}_{SSW,s\theta}$ were identical (that is, equal for all samples s), it would follow that their respective variances, $V_{pq}(\hat{Y}_w)$ and $V_{pq}(\hat{Y}_{SSW,s\theta})$, are equal. As already noted, these two variances share the same first component, $V_{SAM} = V_p(\Sigma_s d_k y_k)$. It follows that if $\hat{Y}_{SSW,s\theta}$ and $\hat{Y}_{ws}$ could be made identical, then the second components would also be identical. The route that we shall follow is therefore first to make the two estimators identical by a suitable choice of the θ_k, then to substitute the resulting ''estimates'' of the θ_k in the already available formula (3.8). This amounts to a parameter fitting technique: We fit the unknown θ_k to the data available for the respondents. Two questions arise: What fitted response probabilities, if any, will make $\hat{Y}_{SSW,s\theta}$ identical to $\hat{Y}_{ws}$? If we replace the unknown θ_k in (3.8) by such fitted response probabilities, will the result be a ''good'' variance estimator for $\hat{Y}_{ws}$? Proposition 3.2, which follows, answers the first question. The second question can only be answered by empirical studies performed on actual data. One such study is reported in Section 6.

Proposition 3.2: *Let v_{sk} be given by (2.5). When θ_k is replaced by $\hat{\theta}_k = v_{sk}^{-1}$, then $\hat{Y}_{SSW,s\theta} = \Sigma_r d_k g_{sk\theta} y_k / \theta_k$ becomes identical to $\hat{Y}_{ws} = \Sigma_r d_k v_{sk} y_k$. Moreover, the values $\hat{\theta}_k = v_{sk}^{-1}$ satisfy the reasonable condition $\Sigma_r (d_k \mathbf{x}_k / \hat{\theta}_k) = \Sigma_s d_k \mathbf{x}_k$. When θ_k is replaced by $\hat{\theta}_k$ in the weights $g_{sk\theta}$, given by (1.2), then the resulting weights are equal to unity for all k.*

The proof of Proposition 3.2 is given in the Appendix.

We now create our variance estimator for $\hat{Y}_{ws}$ by replacing the unknown θ_k in (3.8) by the computable quantities $\hat{\theta}_k = v_{sk}^{-1}$ and by inserting factors f_k that compensate for loss of degrees of freedom (see Remark 3.2). It is clear that $\hat{\theta}_k = v_{sk}^{-1}$ may fall outside the interval [0,1] suggested by viewing $\hat{\theta}_k$ as an estimated probability. But the possibility that some $\hat{\theta}_k$ fall outside the unit interval is not seen as a serious deficiency here, because our main objective is that the $\hat{\theta}_k$ for $k \in r$ perform well collectively, as substitutes for the true θ_k, when the variance is estimated.

These operations, and the fact that all $g_{sk\theta}$ become equal to unity when θ_k is replaced in (3.8) by $\hat{\theta}_k = v_{sk}^{-1}$, lead to the variance estimator $\hat{V}_{pq}^{0}(\hat{Y}_{ws})$ given in Proposition 3.3. The performance of this variance estimator is tested by the empirical studies in Section 6.

Proposition 3.3: *A variance estimator for $\hat{Y}_{ws}$ is given by*

$$\hat{V}_{pq}^{0}(\hat{Y}_{ws}) = \Sigma_r \Sigma (d_k d_l - d_{kl})(v_{sk} y_k)(v_{sl} y_l) - \Sigma_r d_k (d_k - 1) v_{sk} (v_{sk} - 1) y_k^2 + \Sigma_r d_k^2 v_{sk} (v_{sk} - 1) f_k^2 e_k^2 \tag{3.12}$$

where

$$e_k = y_k - \mathbf{x}_k' \mathbf{B}_{rv} \tag{3.13}$$

and $\mathbf{B}_{rv} = (\Sigma_r d_k v_{sk} q_k \mathbf{x}_k \mathbf{x}_k')^{-1} \Sigma_r d_k v_{sk} q_k \mathbf{x}_k y_k$

3.4. A variance estimator for $\hat{Y}_{wU}$

Working still under the assumption that $B_{NR|s} \approx 0$ for all s, we now develop a variance estimator for $\hat{Y}_{wU}$. In this case we have auxiliary information at the population level. Nevertheless, it seems reasonable to ''estimate'' the response probabilities in the same way as for Info-S (Section 3.3), since the only information that we can possibly obtain about the response behaviour comes from examining characteristics of respondents and nonrespondents in the particular sample that was drawn. Therefore, for Info-U, we use the same ''estimates'' for the θ_k as for Info-S, that is, $\hat{\theta}_k = v_{sk}^{-1}$, where v_{sk} is given by (2.5). Our proposal for a variance estimator for $\hat{Y}_{wU}$ is obtained by replacing the unknown θ_k in (3.10) by $\hat{\theta}_k = v_{sk}^{-1}$ and by inserting the factors f_k (see Remark 3.2). The result is given in Proposition 3.4. We test its performance in the empirical investigations in Section 6. (Note, however, that in this case $\hat{Y}_{SSW,U\theta}$ and $\hat{Y}_{wU}$ will not be identical when θ_k is replaced by $\hat{\theta}_k = v_{sk}^{-1}$.)

Proposition 3.4: *A variance estimator for $\hat{Y}_{wU}$ is given by*

$$\hat{V}_{pq}^{0}(\hat{Y}_{wU}) = \Sigma_r \Sigma (d_k d_l - d_{kl})(g_k v_{sk} f_k e_k)(g_l v_{sl} f_l e_l) - \Sigma_r d_k (d_k - 1) v_{sk} (v_{sk} - 1)(g_k f_k e_k)^2 + \Sigma_r d_k^2 v_{sk} (v_{sk} - 1) f_k^2 e_k^2 \tag{3.14}$$

where g_k and e_k are given, respectively, by (3.3) and (3.13).

4. A General Expression of the Nonresponse Bias

In this section we develop a general expression for the bias of the two estimators $\hat{Y}_{ws}$ and $\hat{Y}_{wU}$. Even if such an expression will contain some unknowns, it can be highly useful for identifying those auxiliary variables that are particularly powerful for reducing both nonresponse bias and sampling error. Bethlehem (1988) and Fuller, Loughin, and Baker (1994) also discuss expressions for the nonresponse bias.

Since the same bias expression is obtained for both information levels S and U, we let $\hat{Y}_w$ represent both $\hat{Y}_{ws}$ and $\hat{Y}_{wU}$.

Proposition 4.1: *For large response sets, the nonresponse bias of $\hat{Y}_w$ is*

$$B_{pq}(\hat{Y}_w) \approx -\Sigma_U(1-\theta_k)E_k^\theta \tag{4.1}$$

where $E_k^\theta = y_k - \mathbf{x}_k' \mathbf{B}_U^\theta$ *and* $\mathbf{B}_U^\theta = (\Sigma_U \theta_k q_k \mathbf{x}_k \mathbf{x}_k')^{-1} \Sigma_U \theta_k q_k \mathbf{x}_k y_k$

By introducing a weak restriction on the factors q_k, we obtain from (4.1) another expression that will be used in several examples in the following.

Proposition 4.2: *Suppose that $q_k = 1/\mu' \mathbf{x}_k$ for all $k \in U$, where μ is a column vector of the same dimension as $\mathbf{x}_k$ and not dependent on k. Then, for large response sets,*

$$B_{pq}(\hat{Y}_w) \approx -\Sigma_U E_k^\theta = \Sigma_U \mathbf{x}_k' \mathbf{B}_{UE}^\theta \tag{4.2}$$

where E_k^θ is defined in Proposition 4.1 and $\mathbf{B}_{UE}^\theta = (\Sigma_U \theta_k q_k \mathbf{x}_k \mathbf{x}_k')^{-1} \Sigma_U \theta_k q_k \mathbf{x}_k E_k$;

$$E_k = y_k - \mathbf{x}_k' \mathbf{B}_U \text{ and } \mathbf{B}_U = (\Sigma_U q_k \mathbf{x}_k \mathbf{x}_k')^{-1} \Sigma_U q_k \mathbf{x}_k y_k$$

The proofs of Proposition 4.1 and 4.2 are given in the Appendix.

In Section 6 we examine the bias expressions (4.1) and (4.2) in special cases corresponding to commonly used types of auxiliary information. However, let us at this point make two statements which are valid generally, when $q_k = 1/\mu' \mathbf{x}_k$ for all $k \in U$.

(1) When $E_k = 0$ for all k, $\mathbf{B}_{UE}^\theta = 0$, and consequently $B_{pq}(\hat{Y}_w) \approx 0$. The implication is that if the auxiliary information is strong, the residuals E_k are near zero, and so is the bias. Moreover, the sampling error will be small. Under those highly favourable conditions, whatever the relation between y_k and θ_k, the bias will be near zero.

(2) When each element responds with the same probability θ_0 it follows from (4.2) that the bias will be zero.

In most surveys, we cannot expect the conditions in either (1) or (2) to hold. However, in regularly repeated surveys there often exists considerable information about how the response propensity is correlated with different auxiliary variables. The following proposition can then be helpful in selecting auxiliary variables for inclusion in the auxiliary vector $\mathbf{x}_k$.

Proposition 4.3: *If there exists a constant column vector λ such that $\theta_k^{-1} = 1 + q_k \lambda' \mathbf{x}_k$ for $k \in U$, then $B_{pq}(\hat{Y}_w) \approx 0$.*

The proof of this proposition is given in the Appendix. The importance of Proposition 4.3 lies in the fact that we can specify the auxiliary vector $\mathbf{x}_k$ and the factors q_k as we like, within the bounds set by the total number of available auxiliary variables. This liberty of choice provides a tool for controlling the nonresponse bias.

5. Special Cases

By following the general procedure outlined in Sections 2 and 3 we can construct estimators that incorporate auxiliary information in an efficient manner. The purpose of this section is to promote the credibility of the procedure by showing that many ''conventional techniques'' are special cases. Therefore we derive in this section the expression taken by the general point estimators suggested in Section 2 for four specifications of the vector $\mathbf{x}_k$ corresponding to common types of auxiliary information. Because of space limitations, only four vectors $\mathbf{x}_k$ are considered. For the first three $\mathbf{x}_k$-vectors our point estimators coincide with conventional estimators. For the first two $\mathbf{x}_k$-vectors our variance estimators also confirm existing suggestions. But for the last two $\mathbf{x}_k$-vectors, although they are not particularly ''complex,'' the literature that we have examined does not provide explicit variance estimators. This is not surprising since closed formula variance estimation rapidly becomes intractable as the $\mathbf{x}_k$-vector expands. But computationally speaking, variance estimation remains simple when, as with our approach, one can rely on a general software.

In all cases we choose to take $q_k = 1$ for all k. For each of the specified vectors $\mathbf{x}_k$, this choice satisfies $q_k = 1/\boldsymbol{\mu}'\mathbf{x}_k$. Thus it follows from Proposition 2.3 that the calibration yields the same final weights whether the initial weights are specified as d_k or as cd_k, where c is an arbitrary constant. The findings in Remark 2.1 apply in some of the cases.

In all cases we assume that simple random sampling (SRS) is used, that is, $d_k = N/n$ for all k, where n is the size of s.

Example 5.1. No auxiliary information

By no auxiliary information, we mean that $\mathbf{x}_k = 1$ for all k. It is easily seen from (2.5) and (2.3) that in this case $v_{sk} = v_{Uk} = n/m$ for all k and thus, from (2.4) and (2.2) we get $\hat{Y}_{ws} = \hat{Y}_{wU} = N\Sigma_r y_k/m$, denoted $\hat{Y}_{EXP}$. This is the traditional *expansion estimator*.

We let $f_k = 1$ for all k (see Remark 3.2) in the variance estimators $\hat{V}^0_{pq}(\hat{Y}_{ws})$, given by (3.12), and $\hat{V}^0_{pq}(\hat{Y}_{wU})$, given by (3.14). The two results differ slightly, but when m is large, both are very close to $N^2(1 - m/N)\, S_r^2/m$. Now, this is the usual variance estimator for the expansion estimator for an SRS sample of size m. That our approach gives this result is reassuring, because in the absence of auxiliary information we have no grounds for postulating anything but equal response probabilities for all elements. Moreover, when n is large, the size m of the responding set, although random, has a very small variance. Thus, to approximate the selection procedure by an SRS of size m seems very reasonable.

Turning now to the bias expression given by (4.2), we find that it becomes, for large response sets,

$$B_{pq}(\hat{Y}_{EXP}) \approx N\left(\frac{\Sigma_U \theta_k y_k}{\Sigma_U \theta_k} - \overline{Y}\right) \tag{5.1}$$

where $\overline{Y} = \Sigma_U y_k/N$. It follows that when θ_k is constant for all k, the nonresponse bias of $\hat{Y}_{EXP}$ is approximately zero, a well-known result. The bias can be considerable if the θ_k

are not constant, as most survey statisticians are aware. But with the extremely weak auxiliary information in this case, one cannot expect anything else.

Example 5.2 One-way classification into P groups

In this case the auxiliary vector is $\mathbf{x}_k = \Gamma_k$, with $\Gamma_k = (\gamma_{1k}, \ldots, \gamma_{pk}, \ldots, \gamma_{Pk})'$, where, for $p = 1, \ldots, P$

$$\gamma_{pk} = \begin{cases} 1 & \text{if } k \in \text{group } p \\ 0 & \text{otherwise} \end{cases}$$

By the term ''group p'' we mean either the part of the sample s that falls in group p, denoted s_p, or the part of the population U that defines group p, denoted U_p. We assume that the groups are mutually exclusive and exhaustive and we have $s = \cup_{p=1}^{P} s_p$; $U = \cup_{p=1}^{P} U_p$. The components of the key vector totals are denoted as follows: $\sum_U \mathbf{x}_k = (N_1, \ldots, N_p, \ldots, N_P)'$; $\sum_s \mathbf{x}_k = (n_1, \ldots, n_p, \ldots, n_P)'$ and $\sum_r \mathbf{x}_k = (m_1, \ldots, m_p, \ldots, m_P)'$. For Info-S, $\mathbf{x}_k$ is known for each $k \in s$, and we can calculate $\sum_s d_k \mathbf{x}_k = (\hat{N}_{\pi 1}, \ldots, \hat{N}_{\pi p}, \ldots, \hat{N}_{\pi P})'$, where $\hat{N}_{\pi p} = \sum_{s_p} d_k = N n_p / n$. For Info-U, we also know the vector $\sum_U \mathbf{x}_k = (N_1, \ldots, N_p, \ldots, N_P)'$.

For Info-S, it follows from (2.5) that $v_{sk} = n_p / m_p$ for $k \in r_p$, so the point estimator (2.4) becomes

$$\hat{Y}_{ws} = \hat{Y}_{WCL} = \frac{N}{n} \sum_{p=1}^{P} n_p \bar{y}_{r_p} \tag{5.2}$$

where $\bar{y}_{r_p} = \Sigma_{r_p} y_k / m_p$

Known as the *weighting class estimator*, it is often discussed in the literature; see for example Oh and Scheuren (1983), Kalton and Kasprzyk (1986), Little (1986), Statistics Sweden (1980).

The corresponding variance estimator can be derived from the general form (3.12). The resulting formula bears no close resemblance to the expressions suggested in earlier literature for this case. Anyway, our approach to variance computation is not dependent on a particular algebraic expression; instead we rely on computer software based on the general form.

For Info-U it follows from (2.3) that the weights are $v_{Uk} = N_p / \sum_{r_p} d_k = N_p n / m_p N$ for $k \in r_p$. Thus the general point estimator (2.2) becomes

$$\hat{Y}_{wU} = \hat{Y}_{PST} = \sum_{p=1}^{P} N_p \bar{y}_{r_p} \tag{5.3}$$

It is sometimes called the *poststratified estimator*, but this term is not ideal since it conceals the fact that not one but two phases of selection are involved (the sampling phase and the response phase). However, Kalton and Kasprzyk (1986) are more explicit and distinguish the poststratified estimator used in the full response case (a single phase estimator) from what they call *the population weighting adjustment estimator*, that is, (5.3). The latter term implicitly recognizes a sampling phase followed by a nonresponse phase. Jagers (1986), Bethlehem and Kersten (1985), and Thomsen (1973, 1978) also discuss the estimator (5.3).

We now turn to the variance estimator $\hat{V}^0_{pq}(\hat{Y}_{PST})$ given by (3.14). When m_p is large, $p = 1, \ldots, P$, the result becomes approximately equal to the variance estimator for the pre-stratified estimator with m_p elements drawn with SRS from stratum p, namely

$$\sum_{p=1}^{P} N_p^2 \left(1 - \frac{m_p}{N_p}\right) \frac{1}{m_p} S^2_{r_p}$$

This is a reassuring result.

The general bias expression (4.2) becomes, under the present specifications,

$$B_{pq}(\hat{Y}_{WCL}) = B_{pq}(\hat{Y}_{PST}) \approx \sum_{p=1}^{P} N_p \left(\frac{\sum_{U_p} \theta_k y_k}{\sum_{U_p} \theta_k} - \overline{Y}_p \right) \tag{5.4}$$

where $\overline{Y}_p = \Sigma_{U_p} y_k / N_p$. It follows immediately from (5.4) that the WCL estimator and the PST estimator are approximately unbiased when every element in U_p responds with the same probability. The practical implication is that one should strive to identify such groups, if possible.

Example 5.3 One-way classification and a numerical variable

In the preceding section we considered auxiliary information permitting each element in the sample or in the population to be assigned to one out of P possible groups. To take a step further towards more extensive information we now assume that a numerical auxiliary variable, denoted x, is also available. We examine the case of the auxiliary vector $\mathbf{x}_k = (\Gamma'_k, x_k \Gamma'_k)'$.

For Info-S the calibration equation (1.6) becomes

$\Sigma_r w_k \mathbf{x}_k = (\hat{N}_{\pi 1}, \ldots, \hat{N}_{\pi p}, \ldots, \hat{N}_{\pi P}, \hat{X}_{\pi 1} \ldots, \hat{X}_{\pi p}, \ldots, \hat{X}_{\pi P})'$, where $\hat{N}_{\pi p} = N n_p / n$ and $\hat{X}_{\pi p} = N \sum_{s_p} x_k / n$, and for Info-U the calibration equation (1.7) becomes

$\Sigma_r w_k \mathbf{x}_k = (N_1, \ldots, N_p, \ldots, N_P, X_1, \ldots, X_p, \ldots, X_P)'$, where $X_p = \sum_{U_p} x_k$

The estimator $\hat{Y}_{ws}$, given by (2.4) becomes

$$\hat{Y}_{ws} = \sum_{p=1}^{P} \hat{N}_{\pi p} \left\{ \overline{y}_{r_p} + \left(\overline{x}_{s_p} - \overline{x}_{r_p} \right) B_p \right\} \tag{5.6}$$

where $\overline{x}_{s_p} = \frac{1}{n_p} \sum_{s_p} x_k$; $\overline{x}_{r_p} = \frac{1}{m_p} \sum_{r_p} x_k$; $B_p = \frac{Cov_{xyr_p}}{S^2_{xr_p}}$

with $Cov_{xyr_p} = \frac{1}{m_p - 1} \sum_{r_p} \left(x_k - \overline{x}_{r_p} \right) \left(y_k - \overline{y}_{r_p} \right)$ and $S^2_{xr_p} = \frac{1}{m_p - 1} \sum_{r_p} \left(x_k - \overline{x}_{r_p} \right)^2$

The estimator $\hat{Y}_{wU}$, given by (2.2), takes the well-known form of a separate regression estimator,

$$\hat{Y}_{wU} = \sum_{p=1}^{P} N_p \left\{ \overline{y}_{r_p} + \left(\overline{X}_p - \overline{x}_{r_p} \right) B_p \right\}$$

where $\overline{X}_p = \sum_{U_p} x_k / N_p$

Let us examine the bias of $\hat{Y}_{wU}$ (and $\hat{Y}_{ws}$) in the special case of a single group, that is, when $\mathbf{x}_k = (1, x_k)'$. With $q_k = 1$ for all k, the general formula (4.2) becomes

$$B_{pq}(\hat{Y}_{wU}) \approx \frac{1}{C}\left\{\Sigma_U \theta_k x_k \left(\frac{\Sigma_U \theta_k x_k^2}{\Sigma_U \theta_k x_k} - \overline{X}\right)\Sigma_U \theta_k E_k \right.$$
$$\left. + \Sigma_U \theta_k \left(\overline{X} - \frac{\Sigma_U \theta_k x_k}{\Sigma_U \theta_k}\right)\Sigma_U \theta_k x_k E_k \right\} \tag{5.7}$$

where $C = [\Sigma_U \theta_k \Sigma_U \theta_k x_k^2 - (\Sigma_U \theta_k x_k)^2]/N$ and $E_k = y_k - \overline{Y} - B(x_k - \overline{X})$ with

$$B = \Sigma_U (x_k - \overline{X})(y_k - \overline{Y})/\Sigma_U (x_k - \overline{X})^2, \ \overline{Y} = \Sigma_U y_k/N \text{ and } \overline{X} = \Sigma_U x_k/N$$

We already know from Section 4 that the bias is zero when each element responds with the same probability. In practice this is not likely to occur. But (5.7) shows that other conditions exist under which the bias is near zero, namely, if both $\Sigma_U \theta_k E_k$ and $\Sigma_U \theta_k x_k E_k$ are small. This will happen when all residuals are near zero (which is unlikely), but also if θ_k is uncorrelated, or nearly so, both with E_k and with $x_k E_k$. If one has several x-variables to choose from, but wants to select one of them, the question arises as to which one will come closest to making both sums close to zero. It is not apparent how this could be realized.

Example 5.4 Two-way classification

In practice it is common to have two or more categorical auxiliary variables. In this section we discuss the possibilities and problems that occur with a two-way classification. The principal problems would be similar in the case of a multi-way classification.

Consider two crossing classifications. Let the first of these be defined by a set of groups indexed $p = 1, \ldots, P$ as in Example 5.2. The second is defined by the indicators δ_h, $h = 1, \ldots, H$, such that $\delta_{hk} = 1$ if $k \in$ group h and $\delta_{hk} = 0$ otherwise. If all cell response counts are reasonably large, and information exists at the cell level, then we can let the $\mathbf{x}_k$ be specified as a vector of dimension PH composed of PH-1 entries ''0'' and a single entry ''1'' indicating the cell membership of element k. We need not examine this case here, because it reduces to that of a one-way classification as discussed earlier; the only difference is that we have PH classes instead of just P. However, in many situations a more appropriate formulation of the auxiliary vector is $\mathbf{x}_k = (\Gamma'_k, \Delta'_k)'$, where Γ_k is defined as in Example 5.2 and $\Delta_k = (\delta_{1k}, \ldots, \delta_{hk}, \ldots, \delta_{H-1,k})'$. (In order to avoid singularity of the matrix $\Sigma_r d_k q_k \mathbf{x}_k \mathbf{x}'_k$, the final h-group is excluded.)

The resulting point estimators $\hat{Y}_{ws}$ and $\hat{Y}_{wU}$ have no simple form and are not given here. They are, however, studied empirically in Section 6. The bias formula (4.2) also gives a complex expression in this case.

With this formulation of the $\mathbf{x}_k$-vector, an alternative is to use the raking ratio solution (see for example, Oh and Scheuren (1983)). Raking ratio requires iteration; our solution does not. Raking ratio always gives positive weights; our solution may give a few negative weights. As Lundström (1997) points out, the raking ratio derives from the minimization of a different distance function than (2.1), which is the basis for our solution. As for the point estimators, they will differ by very little, in most situations, whether raking ratio or our solution is used. The literature on raking ratio does not propose a simple estimator of variance; our solution to variance estimation, if not explicit, is at least computationally straightforward and used in the simulations in Section 6.

6. Simulations Based on an Empirical Data Set

In this section, Monte Carlo simulation is used to study how alternative specifications of the auxiliary vector $\mathbf{x}_k$ affect the quality of the estimators derived by the calibration technique in Sections 2 and 3. We compute quality measures such as relative bias, variance and coverage rate of confidence intervals. All simulations are based on one real data set, The Annual Economic Reports of Swedish Clerical Municipalities, referred to in the following as KYBOK, the Swedish acronym.

Throughout the section the only type of parameter estimated is a total for the entire population. The SRS design is used in the first phase, and in the second phase two different response distributions are used.

KYBOK is a census of the 965 clerical municipalities in Sweden. This census provides statistics on different economic variables such as incomes, expenses and investments. In the 1992 census, 832 of the 965 municipalities responded to the postal inquiries. The response rate was thus 86 per cent. The population used in the simulations consists of the 832 responding elements. The study variable y_k is a numerical variable measuring a certain expense item (''central förvaltning – externa driftskostnader''). The first auxiliary variable, denoted Γ_k, is categorical, indicating one out of four possible types of clerical municipality (''pastorat'', ''kyrkliga samfälligheter'', ''församlingar'', ''stiftssamfälligheter''). The second auxiliary variable, denoted x_k, is numerical and defined as the square root of preliminary revenues (''förskottsmedel''). One of our alternatives consists of a two-way classification involving Γ_k and a second categorical auxiliary variable, Δ_k, that points out one of four equal-sized groups indexed by $h = 1, \ldots, H$ that we created by size ordering the 832 elements from the smallest to the largest according to the value of x_k and in such a way that the first group consists of the first 208 elements of the size ordering, the second group of the next 208 elements, and so on.

Figure 1 shows a scatter plot of the study variable y_k and the numerical auxiliary variable

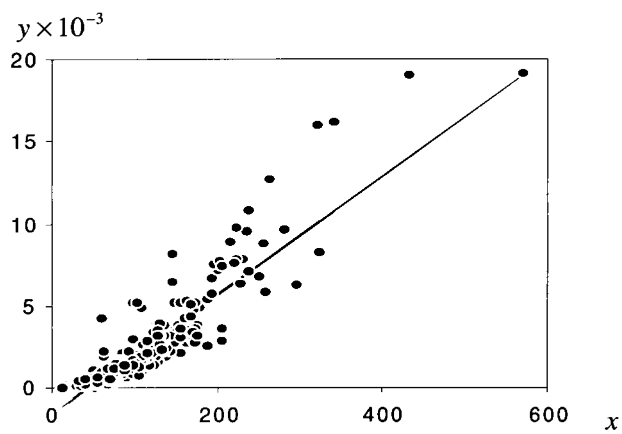

Fig. 1. The scatter plot of the study variable y and the auxiliary variable x. The population regression line is also shown

$x_k, k = 1, \ldots, N$. Also shown is the population regression line obtained by the least squares fit for $E(y_k) = \alpha + \beta x_k$; $V(y_k) = \sigma^2$.

Table 1 shows some key characteristics of the study variable y_k and its correlation with the auxiliary variable x_k.

From the population consisting of the 832 responding elements in the KYBOK census, repeated simple random samples are drawn. For each sample, two response sets are realized using two different response distributions. We refer to the first response distribution as the *logit response distribution*. It is based on a logistic model estimated from the KYBOK census with its 965 elements. For these elements we know the dichotomous variable ''responding/not responding,'' which was treated as the criterion variable in the logistic model fit. The predictor variables were x_k^2 (= preliminary revenues, element k) and two other numerical variables for which we also have data for all 965 elements. In the estimation of the logistic model parameters we assumed a single response model for the entire population, that is, the grouping indicated by the vector $\boldsymbol{\Gamma}_k$ was ignored. The estimated (predicted) response probabilities for the 832 responding units are then used as ''true'' response probabilities θ_k in the simulation studies. The average response probability, $(1/N)\Sigma_U \theta_k$, is 86%.

The second response distribution is artificially created, starting from a given mathematical form. That is, in contrast to the logit response distribution, it is not ''adjusted'' to the KYBOK population. The only thing that the two response distributions have in common is the average response probability of 86%. We call it the *increasing exponential response distribution*. It is defined by the response probabilities $\theta_k = 1 - \exp(-c_1 x_k)$, where we fixed $c_1 = 0.318$ in order to obtain the desired average response probability of 86%.

Our simulation studies involve drawing 10,000 samples of size 300 by SRS from the population of size 832. For each sample, two response sets (one for each response distribution) were realized by carrying out, for the sampled element k, a Bernoulli experiment with parameter θ_k. These experiments are independent. For response set $j (j = 1, \ldots, 10{,}000)$, $\hat{Y}_{w(j)}$ is calculated, where $\hat{Y}_{w(j)}$ is the value of the point estimator $\hat{Y}_w$ for response set j. Here, $\hat{Y}_w = \hat{Y}_{ws}$ given by (2.4) for Info-S, and $\hat{Y}_w = \hat{Y}_{wU}$ given by (2.2) for Info-U. We study two measures, namely (i) the simulation relative bias in percent, $RB_{SIM}(\hat{Y}_w) = 100[E_{SIM}(\hat{Y}_w) - Y]/Y$, where $E_{SIM}(\hat{Y}_w) = \sum_{j=1}^{10{,}000} \hat{Y}_{w(j)}/10{,}000$ is the simulation expectation of $\hat{Y}_w$, and (ii) the simulation variance, $V_{SIM}(\hat{Y}_w) = \sum_{j=1}^{10{,}000} [\hat{Y}_{w(j)} - E_{SIM}(\hat{Y}_w)]^2/9{,}999$.

In the simulations, we studied the point estimators generated by four different $\mathbf{x}_k$-vector

Table 1. Some key characteristics of the study variable y

Characteristics	Entire population	Group (type of clerical municipality)			
		1	2	3	4
Total	1,025,983	132,788	241,836	343,359	308,000
Mean	1,233	609	889	1,184	5,923
Number of elements	832	218	272	290	52
Variance $\times 10^{-3}$	3,598	228	566	1,683	20,369
Correlation coefficient between y and x (R_{yxU})	0.92	0.87	0.90	0.84	0.91

Table 2. Simulation relative bias and simulation variance for different response distributions and different point estimators. SRS with $n = 300$

Response distribution	Auxiliary vector $\mathbf{x}_k$	Info-level	$RB_{SIM}(\hat{Y}_w)$	$V_{SIM}(\hat{Y}_w) \times 10^{-6}$
logit	1	S or U	5.0	6,955
	Γ_k	S	2.2	5,936
		U	2.2	3,713
	$(\Gamma_k', x_k\Gamma_k')'$	S	−0.3	5,366
		U	−0.2	807
	$(\Gamma_k', \Delta_k')'$	S	0.6	5,667
		U	0.7	2,686
increasing exponential	1	S or U	9.3	7,013
	Γ_k	S	5.7	5,772
		U	5.7	3,634
	$(\Gamma_k', x_k\Gamma_k')'$	S	−0.9	5,289
		U	−0.8	743
	$(\Gamma_k', \Delta_k')'$	S	0.6	5,422
		U	0.6	2,519

specifications, namely, $\mathbf{x}_k = 1$ for all k, $\mathbf{x}_k = \Gamma_k$, $\mathbf{x}_k = (\Gamma_k', x_k\Gamma_k')'$, $\mathbf{x}_k = (\Gamma_k', \Delta_k')'$. The specification $\mathbf{x}_k = 1$ for all k, which produces the expansion estimator $\hat{Y}_{EXP} = N\Sigma_r y_k/m$, is of interest only as a benchmark with which the more bias-resistant specifications can be compared. It is a foregone conclusion that $\hat{Y}_{EXP}$ will be heavily biased. The results are displayed in Table 2.

For the variance estimator to be reasonably stable, the number of observations in a group must not be too small. In the simulations we required this number to be at least four. Since the last group in the partition defined by Γ_k (group 4) is very small, it will occur in our simulation, for a certain number of response sets, that this condition is not met. For these response sets, group 4 was collapsed with group 3.

We would expect the nonresponse bias to diminish with increasing amounts of auxiliary information, and this is confirmed by Table 2. A large simulation relative bias is observed for the uninformative specification $\mathbf{x}_k = 1$ for all k, namely 5.0% for the logit and 9.3% for the increasing exponential response distribution. The bias drops substantially when we go to the more informative auxiliary vector $\mathbf{x}_k = \Gamma_k$, but it still remains disturbingly high. However, the bias is nearly eliminated when the numerical variable x_k is also included in the auxiliary vector, either in its original form or categorized and represented by Δ_k.

We would also expect the variance to diminish with increasing amounts of auxiliary information. Table 2 confirms this. Not unexpectedly the reduction is much more striking for Info-U than for Info-S. The most extreme illustration of this occurs for the increasing exponential response distribution and $\mathbf{x}_k = (\Gamma_k', x_k\Gamma_k')'$, where the simulation variance is only 743×10^6 for Info-U as compared to $5,289 \times 10^6$ for Info-S. An interesting observation is that, for Info-U, $\mathbf{x}_k = (\Gamma_k', x_k\Gamma_k')'$ gives much smaller simulation variances than $\mathbf{x}_k = (\Gamma_k', \Delta_k')'$, but for both $\mathbf{x}_k$-vectors the simulation relative bias is small.

Table 3. The f_k-values

Auxiliary vector $\mathbf{x}_k$	1	Γ_k	$(\Gamma'_k, x_k\Gamma'_k)'$	$(\Gamma'_k, \Delta'_k)'$
f_k	1	$\sqrt{\frac{m_p-1}{m_p-2}}$ for $k \in r_p$	$\sqrt{\frac{m_p-1}{m_p-3}}$ for $k \in r_p$	$\sqrt{\frac{m-1}{m-P-H+1}}$ for $k \in r$

We now examine the variance estimator given by (3.12) for Info-S and by (3.14) for Info-U. Let us denote it by $\hat{V}^0_{pq}(\hat{Y}_w)$, where $\hat{Y}_w$ is either $\hat{Y}_{ws}$ or $\hat{Y}_{wU}$, and let us denote its value for sample j by $\hat{V}^0_{pq}(\hat{Y}_{w(j)})$. In Table 3 the factors f_k are defined, for each auxiliary vector, by considering the loss of degrees of freedom to be 0, 1, 2 and $P+H-2$, respectively.

We calculated (a) the simulation relative bias of $\hat{V}^0_{pq}(\hat{Y}_w)$, $RB_{SIM}\left[\hat{V}^0_{pq}(\hat{Y}_w)\right] = 100\left\{E_{SIM}\left[\hat{V}^0_{pq}(\hat{Y}_w)\right] - V_{SIM}(\hat{Y}_w)\right\}/V_{SIM}(\hat{Y}_w)$, where $E_{SIM}[\hat{V}^0_{pq}(\hat{Y}_w)] = \sum_{j=1}^{10{,}000} \hat{V}^0_{pq}(\hat{Y}_{w(j)})/10{,}000$, and (b) the simulation coverage rate for a nominal 95% confidence interval based on $\hat{V}^0_{pq}(\hat{Y}_w)$, $CR_{SIM}[\hat{V}^0_{pq}(\hat{Y}_w)] = \sum_{j=1}^{10{,}000} I_{(j)}/100$ with

$$I_{(j)} = \begin{cases} 1 & \text{if } [a_{1j}, a_{2j}] \text{ contains } Y \\ 0 & \text{otherwise} \end{cases}$$

where $a_{1j} = \hat{Y}_{w(j)} - 1.96[\hat{V}^0_{pq}(\hat{Y}_{w(j)})]^{1/2}$; $a_{2j} = \hat{Y}_{w(j)} + 1.96[\hat{V}^0_{pq}(\hat{Y}_{w(j)})]^{1/2}$. The simulation results are displayed in Table 4.

When $\mathbf{x}_k = 1$ for all k, the two variance estimators $\hat{V}^0_{pq}(\hat{Y}_{ws})$ and $\hat{V}^0_{pq}(\hat{Y}_{wU})$ are not identical, but very nearly equal. We found no numeric differences between them, to the degree of precision used in Table 4, so they appear on the same line in the table.

Some noteworthy results in Table 4 are: (1) The variance estimators that we propose, $\hat{V}^0_{pq}(\hat{Y}_{ws})$ and $\hat{V}^0_{pq}(\hat{Y}_{wU})$, perform well. In most cases, the relative bias is small, and the

Table 4. Simulation relative bias and simulation coverage rate for $\hat{V}^0_{pq}(\hat{Y}_w)$ for Info-S and Info-U, for different response distributions and different point estimators. SRS with $n = 300$

Response distribution	Auxiliary vector $\mathbf{x}_k$	Info-level	$RB_{SIM}[\hat{V}^0_{pq}(\hat{Y}_w)]$	$CR_{SIM}[\hat{V}^0_{pq}(\hat{Y}_w)]$
logit	1	S or U	3.5	94.1
	Γ_k	*S*	0.3	94.6
		U	0.6	94.1
	$(\Gamma'_k, x_k\Gamma'_k)'$	*S*	−1.3	93.2
		U	−5.9	92.9
	$(\Gamma'_k, \Delta'_k)'$	*S*	0.1	94.0
		U	−4.1	93.1
increasing exponential	1	S or U	4.6	85.9
	Γ_k	*S*	1.2	91.6
		U	1.9	87.1
	$(\Gamma'_k, x_k\Gamma'_k)'$	*S*	−2.6	92.3
		U	−6.2	91.6
	$(\Gamma'_k, \Delta'_k)'$	*S*	0.2	94.0
		U	−4.3	93.1

coverage rate is with few exceptions close to the nominal 95%. (2) The cases where the coverage rate drops considerably below the nominal 95% are, for obvious reasons, those where the bias of the point estimator remains relatively high, as Table 4 shows the case to be especially for the increasing exponential mechanism when $\mathbf{x}_k = 1$ for all k, and when $\mathbf{x}_k = \Gamma_k$. (3) As expected, there is no tendency that increased auxiliary information will reduce the relative bias of the variance estimator. (By contrast, such a tendency was predicted, and confirmed in Table 2, for the relative bias of the point estimator.) Although the relative bias is never very large, it is more noticeable for Info-U (where in several cases it lies between −4% and −6%) than for Info-S. This can perhaps be attributed to the fact that the variance (see Table 2) is substantially lower for Info-U than for Info-S. Since the variance consists of sampling variance and nonresponse variance, and the nonresponse variance is the same for both information levels, we will have for Info-U, that the nonresponse component is relatively more important. Any inaccuracy in estimating this component will therefore be more noticeable.

7. Concluding Remarks

We believe that our approach can provide a useful general tool for the methodologist faced with survey nonresponse. Questions beyond those discussed in this article need to be addressed to make the approach fully operational in a survey. For example, we have considered only the estimation of population totals, but most surveys involve other parameters of interest, such as means and ratios of totals. In addition, most surveys require estimation for (perhaps numerous) domains of interest. Also, coverage errors are commonly present in surveys. It would carry too far in this article to address these issues.

We have noted in Section 6 that a judicious use of auxiliary information can significantly reduce both the nonresponse bias and the sampling error. In some surveys, there is an abundance of auxiliary information, so that a selection of the most relevant variables would have to precede the start of the calibration process. We do not necessarily recommend that the totality of the information be used. To blindly add auxiliary information, over and beyond a set of crucial variables, might do more harm than good. These problems are indicated in Nascimento Silva and Skinner (1997) and Lundström (1997). The selection of an ''optimal'' set of auxiliary variables is thus not a trivial problem, and will in many cases require the judgement of an experienced survey statistician.

Some of the issues mentioned in this section require further theoretical work. Empirical studies are also important to gain further experience with the approach that we propose. Also, since many applications of our approach would use information taken from administrative registers, it is important to develop computerized systems that would facilitate the extraction and use of such information. Statistics Sweden is taking steps in this direction.

Appendix

A. Proof of Proposition 2.3.

When d_k^* is replaced by cd_k^* in (2.6) we have the development

$$\sum_r \frac{(w_k - cd_k^*)^2}{cd_k^* q_k} = \frac{1}{c}\sum_r \frac{(w_k - d_k^*)^2}{d_k^* q_k} + \frac{c^2-1}{c}\sum_r \frac{d_k^*}{q_k} - 2\frac{c-1}{c}\sum_r \frac{w_k}{q_k} \qquad \text{(A.1)}$$

But, using (2.7) and (1.7), $\sum_r w_k/q_k = \mu' \sum_r w_k \mathbf{x}_k = \mu' \sum_U \mathbf{x}_k = \sum_U q_k^{-1}$, which is a constant not dependent on the w_k. Likewise, the term $[(c^2-1)/c] \sum_r d_k^*/q_k$ is independent of the w_k. Therefore, to minimize $\sum_r (w_k - cd_k^*)^2/cd_k^* q_k$ under the constraint (1.7) gives exactly the same final weights w_k as minimizing $(1/c)\sum_r (w_k - d_k^*)^2/d_k^* q_k$ under the same constraint.

That these final weights are given by (2.8) is seen by the following. Insert $d_k = d_k^*$ into the expression (2.3), which gives

$$w_k = d_k^*\left[1 + q_k\left(\Sigma_U \mathbf{x}_k' - \Sigma_r d_k^* \mathbf{x}_k'\right)\left(\Sigma_r d_k^* q_k \mathbf{x}_k \mathbf{x}_k'\right)^{-1}\mathbf{x}_k\right] \tag{A.2}$$

That (A.2) simplifies into (2.8) is shown by the following argument: Multiply each term in the sum $\Sigma_r d_k^* \mathbf{x}_k'$ by $q_k \mu' \mathbf{x}_k$, which equals 1 for all k by virtue of (2.7). Then (A.2) can be written $w_k = d_k^*[1 + q_k(\Sigma_U \mathbf{x}_k' - \mu'\mathbf{T})\mathbf{T}^{-1}\mathbf{x}_k]$ where $\mathbf{T} = \Sigma_r d_k^* q_k \mathbf{x}_k \mathbf{x}_k'$. Since $q_k \mu' \mathbf{T}\mathbf{T}^{-1}\mathbf{x}_k = 1$ for all k, it follows that $w_k = d_k^* q_k(\Sigma_U \mathbf{x}_k')(\Sigma_r d_k^* q_k \mathbf{x}_k \mathbf{x}_k')^{-1}\mathbf{x}_k$, which is the desired expression (2.8).

For Info-S, a simple modification of the proof shows that the final weights in that case are given by (2.9).

B. Proof of Proposition 3.1.

The mean squared error can be written

$$MSE(\hat{Y}_w) = V_{pq}(\hat{Y}_w) + B_{pq}^2(\hat{Y}_w) \tag{B.1}$$

where $V_{pq}(\hat{Y}_w) = E_{pq}[\hat{Y}_w - E_{pq}(\hat{Y}_w)]^2$ and $B_{pq}(\hat{Y}_w) = E_{pq}(\hat{Y}_w) - Y$

Let us write $V_{pq}(\hat{Y}_w)$ as a sum of components. First, we note that

$$\begin{aligned} V_{pq}\left(\hat{Y}_w\right) &= E_p V_q\left(\hat{Y}_w|s\right) + V_p\left(B_{NR|s} + \hat{Y}_s\right) \\ &= E_p V_q\left(\hat{Y}_w|s\right) + V_p\left(B_{NR|s}\right) + V_p\left(\hat{Y}_s\right) + 2Cov_p\left(\hat{Y}_s, B_{NR|s}\right) \end{aligned} \tag{B.2}$$

However,

$$\begin{aligned} V_p\left(B_{NR|s}\right) &= E_p\left(B_{NR|s}^2\right) - \left[E_p\left(B_{NR|s}\right)\right]^2 \\ &= E_p\left(B_{NR|s}^2\right) - \left[E_p E_q\left(\hat{Y}_w - \hat{Y}_s|s\right)\right]^2 = E_p\left(B_{NR|s}^2\right) - B_{pq}^2\left(\hat{Y}_w\right) \end{aligned} \tag{B.3}$$

Using notation introduced in Proposition 3.1, we now get

$$V_{pq}\left(\hat{Y}_w\right) = V_{SAM} + V_{NR} + 2Cov_p\left(\hat{Y}_s, B_{NR|s}\right) + E_p\left(B_{NR|s}^2\right) - B_{pq}^2\left(\hat{Y}_w\right) \tag{B.4}$$

The desired result (3.4) now follows from (B.1) and (B.4).

C. Proof of Proposition 3.2.

It is easily seen that the two expressions (1.1) and (2.4) are identical when $v_{sk} = g_{sk\theta}/\theta_k$. This means that we seek the response probabilities θ_k that satisfy the equation

$$\theta_k = \frac{1 + q_k(\Sigma_s d_k \mathbf{x}_k - \Sigma_r d_k \mathbf{x}_k/\theta_k)'(\Sigma_r d_k q_k \mathbf{x}_k \mathbf{x}_k'/\theta_k)^{-1}\mathbf{x}_k}{1 + q_k(\Sigma_s d_k \mathbf{x}_k - \Sigma_r d_k \mathbf{x}_k)'(\Sigma_r d_k q_k \mathbf{x}_k \mathbf{x}_k')^{-1}\mathbf{x}_k} \text{ for } k \in r \tag{C.1}$$

This equation usually has many solutions. Among all these solutions we choose the one that satisfies

$$\Sigma_r \frac{d_k \mathbf{x}_k}{\theta_k} = \Sigma_s d_k \mathbf{x}_k \tag{C.2}$$

This choice is reasonable because

$$E_q\left(\Sigma_r \frac{d_k \mathbf{x}_k}{\theta_k} | s\right) = \Sigma_s d_k \mathbf{x}_k.$$

Condition (C.2) inserted in (C.1) provides a solution that we denote $\hat{\theta}_k$, namely,

$$\hat{\theta}_k = \frac{1}{1 + q_k(\Sigma_s d_k \mathbf{x}_k - \Sigma_r d_k \mathbf{x}_k)'(\Sigma_r d_k q_k \mathbf{x}_k \mathbf{x}_k')^{-1}\mathbf{x}_k} \text{ for } k \in r \tag{C.3}$$

It is easy to see that (C.3) satisfies the condition (C.2). We have

$$\begin{aligned}\Sigma_r d_k \mathbf{x}_k'/\hat{\theta}_k &= \Sigma_r d_k \mathbf{x}_k' + (\Sigma_s d_k \mathbf{x}_k - \Sigma_r d_k \mathbf{x}_k)'(\Sigma_r d_k q_k \mathbf{x}_k \mathbf{x}_k')^{-1}\Sigma_r d_k q_k \mathbf{x}_k \mathbf{x}_k' \\ &= \Sigma_r d_k \mathbf{x}_k' + (\Sigma_s d_k \mathbf{x}_k - \Sigma_r d_k \mathbf{x}_k)' = \Sigma_s d_k \mathbf{x}_k'\end{aligned}$$

D. Proof of Propositions 4.1 to 4.3.

To prove Proposition 4.1, we note from (2.2) that

$$\hat{Y}_{wU} - Y = \Sigma_r d_k y_k + (\Sigma_U \mathbf{x}_k - \Sigma_r d_k \mathbf{x}_k)'\mathbf{B}_r - \Sigma_U y_k \tag{D.1}$$

where $\mathbf{B}_r = (\Sigma_r d_k q_k \mathbf{x}_k \mathbf{x}_k')^{-1}\Sigma_r d_k q_k \mathbf{x}_k y_k$

We obtain an approximation of the bias $E_{pq}(\hat{Y}_{wU}) - Y$ by replacing each of the random terms in (D.1) by its expected value. We have $E_{pq}(\Sigma_r d_k y_k) = \Sigma_U \theta_k y_k$ and, analogously, $E_{pq}(\Sigma_r d_k x_k) = \Sigma_U \theta_k x_k$. Further, for large response sets, $E_{pq}(\mathbf{B}_r) \approx \mathbf{B}_U^{\theta}$, where $\mathbf{B}_U^{\theta}$ is as defined in Proposition 4.1. Therefore, $E_{pq}(\hat{Y}_{wU} - Y) \approx -\Sigma_U(1-\theta_k)y_k + \Sigma_U(1-\theta_k)\mathbf{x}_k'\mathbf{B}_U^{\theta} = -\Sigma_U(1-\theta_k)E_k^{\theta}$, with $E_k^{\theta} = y_k - \mathbf{x}_k'\mathbf{B}_U^{\theta}$. Proposition 4.1 is thereby verified.

It is easy to follow the proof just given for the bias of $\hat{Y}_{wU}$ and see that the same expression is obtained for the estimator $\hat{Y}_{ws}$.

To prove Proposition 4.2, note that the nonresponse bias expression (4.1) can be written

$$\begin{aligned}B_{pq}(\hat{Y}_w) \approx -\Sigma_U(1-\theta_k)E_k^{\theta} &= \Sigma_U(1-\theta_k)\mathbf{x}_k'\mathbf{B}_U^{\theta} - \Sigma_U(1-\theta_k)y_k \\ &= \Sigma_U \mathbf{x}_k'\mathbf{B}_U^{\theta} - \Sigma_U \theta_k \mathbf{x}_k'\mathbf{B}_U^{\theta} - \Sigma_U(1-\theta_k)y_k\end{aligned} \tag{D.2}$$

Now, by assumption, $q_k \mu' \mathbf{x}_k = 1$ for all k, so the second component of (D.2) becomes $\Sigma_U \theta_k \mathbf{x}_k'\mathbf{B}_U^{\theta} = \mu'\Sigma_U \theta_k q_k \mathbf{x}_k \mathbf{x}_k'\mathbf{B}_U^{\theta} = \Sigma_U \theta_k q_k \mu' \mathbf{x}_k y_k = \Sigma_U \theta_k y_k$. Thus, $B_{pq}(\hat{Y}_w) \approx \Sigma_U \mathbf{x}_k'\mathbf{B}_U^{\theta} - \Sigma_U y_k = -\Sigma_U E_k^{\theta}$

Finally, it is easily seen that $\mathbf{B}_U^{\theta} = \mathbf{B}_U + \mathbf{B}_{UE}^{\theta}$, where $\mathbf{B}_U$ and $\mathbf{B}_{UE}^{\theta}$ are as defined in Proposition 4.2. Therefore,

$$B_{pq}(\hat{Y}_w) \approx \Sigma_U \mathbf{x}_k'\mathbf{B}_U^{\theta} - \Sigma_U y_k = \Sigma_U \mathbf{x}_k'\mathbf{B}_U + \Sigma_U \mathbf{x}_k'\mathbf{B}_{UE}^{\theta} - \Sigma_U y_k = \Sigma_U \mathbf{x}_k'\mathbf{B}_{UE}^{\theta} - \Sigma_U E_k$$

However, when $q_k = 1/\mu'\mathbf{x}_k$ for all $k \in U$, $\Sigma_U E_k = 0$ and therefore $B_{pq}(\hat{Y}_w) \approx \Sigma_U \mathbf{x}_k'\mathbf{B}_{UE}^{\theta}$. Proposition 4.2 is thereby verified.

To prove Proposition 4.3, multiply element k in the first component of (D.2) by $\theta_k(1 + q_k \lambda' \mathbf{x}_k)$, which by assumption equals 1 for all k. We get

$$\Sigma_U \mathbf{x}_k'\mathbf{B}_U^{\theta} = \Sigma_U \theta_k \mathbf{x}_k'\mathbf{B}_U^{\theta} + \lambda'\Sigma_U \theta_k q_k \mathbf{x}_k \mathbf{x}_k'\mathbf{B}_U^{\theta}$$

$$= \Sigma_U \theta_k \mathbf{x}_k' \mathbf{B}_U^\theta + \lambda' \Sigma_U \theta_k q_k \mathbf{x}_k \mathbf{x}_k' (\Sigma_U \theta_k q_k \mathbf{x}_k \mathbf{x}_k')^{-1} \Sigma_U \theta_k q_k \mathbf{x}_k y_k$$

$$= \sum_U \theta_k \mathbf{x}_k' \mathbf{B}_U^\theta + \sum_U \theta_k q_k \lambda' \mathbf{x}_k y_k$$

However, since $\theta_k q_k \lambda' \mathbf{x}_k = 1 - \theta_k$, $\Sigma_U \mathbf{x}_k' \mathbf{B}_U^\theta = \Sigma_U \theta_k \mathbf{x}_k' \mathbf{B}_U^\theta + \Sigma_U (1 - \theta_k) y_k$. Thus, it is seen that (D.2) equals zero and Proposition 4.3 is verified.

8. References

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

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

Binder, D.A., Michaud, S., and Poirier, C. (1994). Model-based Reweighting for Nonresponse Adjustment. Proceedings of Survey Research Methods Section, American Statistical Association, 989–994.

Deville, J.C. and Särndal, C.E. (1992). Calibration Estimators in Survey Sampling. Journal of the American Statistical Association, 87, 376–382.

Dupont, F. (1994). Calibration Used as a Nonresponse Adjustment. New Approaches in Classification and Data Analysis. (Data on publishing details unknown.)

Ekholm, A. and Laaksonen, S. (1991). Weighting via Response Modeling in the Finnish Household Budget Survey. Journal of Official Statistics, 3, 325–337.

Fuller, W.A., Loughin, M.M., and Baker, H.D. (1994). Regression Weighting in the Presence of Nonresponse with Application to the 1987–1988 Nationwide Food Consumption Survey. Survey Methodology, 20, 75–85.

Holt, D. and Elliott, D. (1991). Methods of Weighting for Unit-nonresponse. The Statistician, 40, 333–342.

Jagers, P. (1986). Post-Stratification Against Bias in Sampling. International Statistical Review, 54, 159–167.

Kalton, G. and Kasprzyk, D. (1986). The Treatment of Missing Data. Survey Methodology, 12, 1–16.

Little, R.J.A. (1986). Survey Nonresponse Adjustments for Estimates of Means. International Statistical Review, 54, 139–157.

Lundström, S. (1997). Calibration as a Standard Method for Treatment of Nonresponse. Doctoral dissertation, Stockholm University.

Nascimento Silva, P.L.D. and Skinner, C.J. (1997). Variable Selection for Regression Estimation in Finite Populations. Survey Methodology, 23, 23–32.

Oh, H.L. and Scheuren, F.J. (1983). Weighting Adjustment for Unit Nonresponse. In W.G. Madow, I. Olkin and D.B. Rubin (eds.), Incomplete Data in Sample Surveys, Vol. 2. New York: Academic Press, 143–184.

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.

Särndal, C.E., Swensson, B., and Wretman, J.H. (1992). Model Assisted Survey Sampling. New York: Springer-Verlag.

Singh, A.C., Wu, S., and Boyer, R. (1995). Longitudinal Survey Nonresponse Adjustment by Weight Calibration for Estimation of Gross Flows. Proceedings of the Survey Research Methods Section, American Statistical Association, 396–401.

Statistics Sweden (1980). Räkna med bortfall. Statistics Sweden. (In Swedish.)

Thomsen, I. (1973). A Note on the Efficiency of Weighting Subclass Means to Reduce the Effects of Nonresponse when Analysing Survey Data. Statistisk tidskrift, 11, 278–285.

Thomsen, I. (1978). A Second Note on the Efficiency of Weighting Subclass Means to Reduce the Effects of Nonresponse when Analyzing Survey Data. Statistisk tidskrift, 16, 191–196.

Thomsen, I. and Siring, E. (1983). On the Causes and Effects of Nonresponse: Norwegian Experiences. In W.G. Madow, I. Olkin, and D.B. Rubin (eds.), Incomplete Data in Sample Surveys, Vol. 3. New York: Academic Press, 25–59.

Received July 1998
Revised February 1999