Discussion Papers No. 311, October 2001
Statistics Norway, Statistical Methods and Standards Division

Li-Chun Zhang

A method of weighting adjustment for survey data subject to nonignorable nonresponse

Abstract:
Weighting adjustment is a standard quasi-randomization approach for survey data subject to nonresponse (Little, 1986). The existing methods are typically based on the assumption that nonresponse is independent of the survey variable conditional to the auxiliary variables used to form the adjustment cells. In this paper we consider nonignorable nonresponse which is independent of certain auxiliary information conditional to the variable of interest. We estimate the size of the sample adjustment cells using a method of moment conditional to the sample. The method relies on only the nonresponse mechanism, and is independent of the sample design. In variance estimation, we evaluate the nonresponse effect on estimation and design, analogously to the concept of design effect. By comparing the nonresponse effects under a nonignorable model against those under an ignorable one, we obtain a means of measuring the effect of nonignorability. We motivate and illustrate our approach for estimation of household composition.

Keywords: weighting adjustment, nonresponse effect, effect of nonignorability, stratified simple random sampling, post-stratification

Acknowledgement: I am very grateful to Jan F. Bjørnstad for many discussions on this work.

Address: Li-Chun Zhang, Statistics Norway, Statistical Methods and Standards Division.
E-mail: li.chun.zhang@ssb.no

Discussion Papers comprise research papers intended for international journals or books. As a preprint a Discussion Paper can be longer and more elaborate than a standard journal article by including intermediate calculation and background material etc.

Abstracts with downloadable PDF files of
Discussion Papers are available on the Internet: http://www.ssb.no

For printed Discussion Papers contact:

Statistics Norway
Sales- and subscription service
N-2225 Kongsvinger

Telephone: +47 62 88 55 00
Telefax: +47 62 88 55 95
E-mail: Salg-abonnement@ssb.no

1 Introduction

For the survey of living conditions (SLC) in 1999, a simple random sample of 4958 persons was selected from all persons of age 16 or over in the population. Household information was obtained from 3758 of them, so that the nonresponse rate was just over 24%. Our objective here is to estimate the number of households by the size of the household in the population. As auxiliary information from the population administrative register, we have the size of the family in which a person is registered. This information can be linked to the sample through a personal identity number. There are important differences between a registered family and a dwelling household. Thus, a household may contain several registered families and generations. While a registered family never involves more than two generations, its members may live in separate households. Exploratory data analysis (Table 1) shows that the nonresponse rate is higher among persons from smaller registered families. This agrees to the fact that smaller households are more difficult to reach than the larger ones. Under-representation of smaller households among the respondents implies that nonoresponse presumably is nonignorable in the sense of Rubin (1976), because it seems unlikely that the probability of nonresponse may be independent of the actual size of the household, given the size of the family in the register.

Table 1: Response rate (%) in the SLC by the registered family size and the person's age

	Number of persons in the registered family									
Age of the person	1		2		3		4		≥ 5	
Under 45	71.4	(625)	76.2	(265)	77.4	(517)	83.8	(722)	81.4	(474)
Between 45 and 64	66.6	(311)	74.7	(581)	78.1	(329)	79.3	(237)	81.9	(116)
Over 64	62.0	(316)	72.4	(410)	80.4	(51)	100	(4)	0	(0)

Note: Numbers in the parentheses indicate how many persons the response rate is based on.

Little and Rubin (1987) distinguish between the modeling and quasi-randomization approach to nonresponse in sample surveys. Apart from the case of missing completely at random (MCAR), a typical assumption of weighting adjustment under the quasi-randomization approach is that nonresponse is independent of the survey variable conditional to the auxiliary variables available. Even when ignorable nonresponse as such is not true, useful adjustments can be obtained due to the correlation between the auxiliary and survey variables (Zhang, 1999). Indeed, once we depart from the MCAR-assumption, the objective of analysis can no longer be to provide a single valid inference, since a nonresponse model, ignorable or not, can never be conclusively established based on the data alone. Nevertheless, contextual evidences and conceptual considerations may suggest that the inference is likely to be less biased under some nonresponse models, possibly nonignorable, than others (e.g. Molenberghs, Goetghebeur, Lipsitz, and Kenward, 1999).

Little (1986) discusses adjustment methods under the assumption of ignorable nonresponse. The household composition being categorical variables, it is natural in the present case to form adjustment cells by response propensity stratification according to the nonresponse probability of each unit. Motivated by the nonresponse situation in the SLC, we begin by defining a number of *nonresponse classes* in the sample which, among other things, depend on the size of the household (Table 2). The sizes of the nonresponse classes are therefore unknown among the nonrespondents. We assume that, within each nonresponse class, the probability of nonresponse is independent of the size of the family in the register. Any identifiable subgroup of a nonresponse class can now be used as an adjustment cell. With the simple multinomial sampling, our model of conditional independence is formally a decomposable graphical model (Lauritzen, 1996). which again is a subclass of the log-linear models (Forster and Smith, 1998). To estimate the sizes of the adjustment cells among the nonrespondents, we apply a method of moment conditional to the sample, which depends on only the nonresponse mechanism. The method is thus valid regardless of the underlying sampling distribution of the selected units. The details of the weighting adjustment will be explained in Section 2.1 and 2.2.

Table 2: Definition of nonresponse classes in the SLC

No.	Nonresponse class	No.	Nonresponse class
I	1-person household, person's age under 45	VII	3-person household
II	1-person household, person's age between 45 and 64	VIII	4-person household
III	1-person household, person's age over 64	IX	Others
IV	2-person household, person's age under 45		
V	2-person household, person's age between 45 and 64		
VI	2-person household, person's age over 64		

From the quasi-randomization perspective, both the sampling error and the nonresponse contribute to the total variance of an estimator. Variance calculation is more informative if it is able to describe to us the various effects of nonresponse. Denote by E_θ and Var_θ expectation and variance with respect to the nonresponse mechanism, and E_π and Var_π that with respect to the sample design. To facilitate the derivation of the total variance of an estimator, denoted by $\hat{T}$, it is often helpful to employ either of the following two decompositions, i.e.

$$Var(\hat{T}) = E_\pi[Var_\theta(\hat{T})] + Var_\pi(E_\theta[\hat{T}]) = E_\theta[Var_\pi(\hat{T})] + Var_\theta(E_\pi[\hat{T}]),$$

where the inner expectation and variance are treated as conditional ones. For instance, Rao and Sitter (1995) apply the former approach, whereas Fay (1991) and Shao and Steel (1999) make use of the latter. However, while both $E_\pi[Var_\theta(\hat{T})]$ and $Var_\theta(E_\pi[\hat{T}])$ are mainly due to nonresponse, neither of them summarizes in itself *all* the effects of nonresponse.

In Section 2.3 we define the nonresponse effect (neff) on respectively estimation and sampling, in analogy to the well-known concept of design effect (deff). Described in words, the neff on estimation is the ratio between the total variance of an estimator, and the sampling variance of the same estimator in the absence of nonresponse, under the same sample design. Typically, the latter can be estimated using standard methods by treating the imputed data as if they had been observed. The neff on estimation, however, does not contain all the nonresponse effect. Nonresponse could also affect the sample design because, in general, the respondents may differ systematically from the nonrespondents. Had the nonresponse status been known for the whole population at the design stage, we could have considered a stratified design, in which the actual sample design was separately applied within the subpopulation of the respondents and that of the nonrespondents. This would have led to a variance reduction except when there in fact is no systematic difference between the two subpopulations. The neff on design is thus defined as the ratio between the unstratified and the stratified sampling variance, both in the absence of nonresponse. The overall neff is now given by the product of the neff on estimation and the neff on design, which measures the total variance inflation due to nonresponse.

It is clear that the nonresponse effects can only be evaluated under an assumed nonresponse model. By comparing the neff's across different models, we are able to measure the alternative nonresponse assumptions against each other. Of special interest are measures of a nonignorable model against an ignorable one. We define the effect of nonignorability (eff_n) for estimation as the ratio between the neff on estimation under a nonignorable and an ignorable model. Whereas the eff_n for design is similarly defined between the neff on design under the two models. The overall effect of nonignorability is given by the product of the eff_n's on estimation and design. In cases where we have a set of nonignorable models for consideration, we may prefer to fix one ignorable model for base-line comparison. Together, deff and eff_n measure the various effects of missing data in terms of variance. Section 2.3 provides the details in the case of stratified simple random sampling. Empirical results based on the SLC are discussed in Section 3.

2 Method

2.1 A conditional independence nonresponse model

Denote by s the sample. Let y_i, for $y_i = 1, ..., J$, be the nonresponse class indicator of unit $i \in s$. In particular, the definition of the nonresponse class may depend on the survey variables (such as in Table 2), which are unknown for the nonrespondent units. Let x_i, for $x_i = 1, ..., K$, be some auxiliary variable which is available for all $i \in s$. Let $R_i = 1$ if response, and $R_i = 0$ if nonresponse. The conditional independence nonresponse model is given by

$$P[R_i = 1|x_i = x, y_i = y] = P[R_i = 1|y_i = y]. \tag{1}$$

Let n_{xy} be the number of respondent units with $(x_i, y_i) = (x, y)$. Define m_{xy} similarly for the nonrespondents, which is unknown except from the marginal total $m_x = \sum_y m_{xy}$. We have

	Response					Nonresponse (Unobserved)			
	$Y=1$	$Y=2$	$\cdots$	$Y=J$	Nonresponse	$Y=1$	$Y=2$	$\cdots$	$Y=J$
$X=1$	n_{11}	n_{12}	$\cdots$	n_{1J}	m_1	m_{11}	m_{12}	$\cdots$	m_{1J}
$X=2$	n_{21}	n_{22}	$\cdots$	n_{2J}	m_2	m_{21}	m_{22}	$\cdots$	m_{2J}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$X=K$	n_{K1}	n_{K2}	$\cdots$	n_{KJ}	m_K	m_{K1}	m_{K2}	$\cdots$	m_{KJ}

Under the nonresponse model (1), we notice that, at the current $\{n_{xy}, \hat{m}_{xy}\}$, we have

$$\hat{P}[R_i = 0|y_i = y] = (\sum_x n_{xy} + \sum_x \hat{m}_{xy})^{-1}(\sum_x \hat{m}_{xy})$$

and

$$\hat{E}[m_{xy}|n_{xy} + \hat{m}_{xy}] = (n_{xy} + \hat{m}_{xy})\hat{P}[R_i = 0|y_i = y].$$

Conditional to the observed $m_x = \sum_y \hat{m}_{xy}$, we update $\hat{m}_{xy}$ by

$$\hat{m}_{xy} = m_x \hat{E}[m_{xy}|n_{xy} + \hat{m}_{xy}](\sum_{j=1}^{J} \hat{E}[m_{xj}|n_{xj} + \hat{m}_{xj}])^{-1},$$

and iterate. Notice that this is the EM algorithm for data arising from the simple multinomial sampling. Convergence is usually not a problem. However, it is good practice to choose moderate sizes of J and K, so as to avoid setting up tables with many small or empty cells. See Smith, Skinner, and Clarke (1999) for more detailed discussions on this issue. Due to the restriction of $m_x = \sum_y \hat{m}_{xy}$, the obtained $\{\hat{m}_{xy}\}$ do not always exactly satisfy, for $y = 1, ..., J$,

$$\frac{\hat{m}_{1y}}{n_{1y} + \hat{m}_{1y}} = \frac{\hat{m}_{2y}}{n_{2y} + \hat{m}_{2y}} = \cdots = \frac{\hat{m}_{Ky}}{n_{Ky} + \hat{m}_{Ky}}. \tag{2}$$

We may consider the algorithm above as a method of conditional moment regardless of the sampling distribution of the (x, y)-cells. Any selected sample contains a certain number of units with $(x_i, y_i) = (x, y)$, denoted by c_{xy} where $\sum_y c_{xy} = \sum_y n_{xy} + m_x$. The nonresponse mechanism which generates n_{xy} and m_{xy} has a Binomial distribution given c_{xy}. At each iteration we take expectation with respective to the nonresponse mechanism alone, conditional to the current value of $\hat{c}_{xy} = n_{xy} + \hat{m}_{xy}$. In this way the estimates $\{\hat{m}_{xy}\}$ are independently derived of the sampling distribution. It follows that we generally do not use $\sum_x (n_{xy} + \hat{m}_{xy})/(\sum_{x,y} n_{xy} + \sum_x m_x)$ as an estimate of the proportion of $y_i = y$ in the population. To infer from the imputed sample to the population, we still need to apply some weighting method appropriate for the sample design.

2.2 Weighting adjustment

Let $s_y = \{i \in s; y_i = y\}$ be an adjustment cell in the sample by response propensity stratification. The *adjustment weight* of any respondent unit $i \in s_y$ is given by

$$a_i = (\sum_x n_{xy})^{-1}(\sum_x n_{xy} + \sum_x \hat{m}_{xy}). \tag{3}$$

Let $s_{xy} = \{i \in s; (x_i, y_i) = (x, y)\}$. Since all $i \in s_{xy}$ have the same response probability under model (1), we could also use s_{xy} as an adjustment cell, i.e. for any respondent $i \in s_{xy}$,

$$a_i = n_{xy}^{-1}(n_{xy} + \hat{m}_{xy}). \tag{4}$$

There will be no difference between (3) and (4) provided $\{\hat{m}_{xy}\}$ exactly satisfy (2). Otherwise, a_i by (3) is more stable than that by (4), and leads to estimators with smaller variances. Whereas a_i by (4) may have better control over the bias, especially for domain estimates. Notice that the sum of the adjustment weights over the respondent units is by definition the size of the sample, which entails adjustment for nonresponse under model (1).

The adjustments (3) and (4) differ somewhat from the standard weighting class adjustment. In cases where the adjustment cells are formed using the auxiliary variables alone, we always know which adjustment cell a nonrespondent unit belongs to. The design weight of a respondent unit is then adjusted by a factor estimated at the population level. For instance, let s_c be such an adjustment cell in the sample. For any respondent unit $i \in s_c$, we would adjust its design weight by the factor $\sum_{i \in s_c} \pi_i^{-1} / \sum_{i \in s_c; r_i = 1} \pi_i^{-1}$, where π_i is the inclusion probability of unit i. In contrast, the adjustment weight a_i under the nonignorable model (1) is derived from estimates at the sample level. That is, we estimate the nonresponse sample at the (x, y)-cell level, i.e. $\{\hat{m}_{xy}\}$, without specifying to which adjustment cell a nonrespondent unit belongs.

For any respondent unit $i \in s$, we define its weight as

$$w_i = N(\pi_i^{-1} a_i)(\sum_{i \in s; r_i = 1} \pi_i^{-1} a_i)^{-1},$$

where $N = \sum_{i \in s} \pi_i^{-1} = \sum_{i \in s; r_i = 1} w_i$ is the size of the population. In the case of $r_i = 1$ for all $i \in s$, this reduces to the weighted sample mean estimator since $a_i = 1$. The post-stratified weights are similarly given within each post-stratum. Let N_h be the size of the population in post-stratum h, and s_h the corresponding sample post-stratum. For any respondent unit $i \in s_h$, we let

$$w_i = N_h(\pi_i^{-1} a_i)(\sum_{i \in s_h; r_i = 1} \pi_i^{-1} a_i)^{-1}. \tag{5}$$

Let z_i be a survey variable of interest. We estimate its population total by

$$\hat{T} = \sum_{i \in s; r_i = 1} w_i z_i = \sum_{i \in s} r_i w_i z_i, \tag{6}$$

where we set $r_i w_i z_i = 0$ in the case of $r_i = 0$, without assigning any explicit values to w_i or z_i.

2.3 Variance estimation and nonresponse effects

Take first the case of simple random sampling without replacement. We evaluate the conditional variance of the post-stratified estimator given by (5) and (6) with $h = x$ (Holt and Smith, 1979). Shao and Sitter (1996) discusses Bootstrap variance estimation for imputed survey data. Under condition (i) the sample size is not small, and (ii) the sampling fraction is negligible, the various proposed Bootstrap methods all agree closely with the infinite-population nonparametric Bootstrap for missing data (Efron, 1994). Let $s_x = \{i \in s; x_i = x\}$ and $n_x = \sum_y n_{xy}$. We form a Bootstrap sample by stratified resampling of $n_x + m_x$ units from each s_x, with all the associated (y_i, z_i, r_i) values, randomly and with replacement. We group the Bootstrap sample into $\{n^*_{xy}, m^*_x\}$ as defined in Section 2.1, based on which we obtain $\hat{T}^*$ by the weighting adjustment method described in Section 2.1 and 2.2. Independent repetitions give us $\hat{T}^*_{(1)}, ..., \hat{T}^*_{(B)}$, and

$$v = V\hat{a}r(\hat{T} | \{n_x + m_x\}) = (B-1)^{-1} \sum_{b=1}^{B} (\hat{T}^*_{(b)} - B^{-1} \sum_{d=1}^{B} \hat{T}^*_{(d)})^2. \tag{7}$$

Consider now the case of $z_i = I_{y_i = y}$, where $I_{y_i = y} = 1$ if $y_i = y$, and 0 otherwise. Let N_x be the size of the subpopulation with $x_i = x$, and $\hat{p}_{xy} = (n_{xy} + \hat{m}_{xy})/(n_x + m_x)$, such that

$$v_0 = \sum_x N_x^2 (n_x + m_x)^{-1} \hat{p}_{xy}(1 - \hat{p}_{xy}) \qquad \text{and} \qquad \hat{T} = \sum_x N_x \hat{p}_{xy}. \tag{8}$$

Had $\hat{m}_{xy}$ been observed, $\hat{T}$ would have been the simple post-stratified estimator of the population total of z_i, whereas v_0 would have been an estimate of its conditional sampling variance assuming negligible $(n_x + m_x)/N_x$. Typically, we have $v > v_0$, where the increment is entirely caused by the fact that y_i is missing from the nonrespondents. Since both v and v_0 are derived under the same sample design, we may define the *nonresponse effect (neff) on estimation* as

$$\text{neff}_{\text{est}} = v_0^{-1} v.$$

Nonresponse can also affect the sample design because, in general, the respondents may differ systematically from the nonrespondents. Had r_i been known throughout the population, therefore, we could have considered a stratified design according to r_i. Let $n_{1,x} = n_x$ and $n_{0,x} = m_x$. Let $\hat{N}_{r,x} = N_x n_{r,x}/(n_x + m_x)$ for $r = 0, 1$. Let $\hat{p}_{1,xy} = n_{xy}/n_x$, and $\hat{p}_{0,xy} = \hat{m}_{xy}/m_x$, such that

$$v_1 = \sum_r \sum_x \hat{N}_{r,x}^2 n_{r,x}^{-1} \hat{p}_{r,xy}(1 - \hat{p}_{r,xy}) \qquad \text{and} \qquad \hat{T} = \sum_r \sum_x \hat{N}_{r,x} \hat{p}_{r,xy}. \tag{9}$$

Notice that $\hat{T}$ is now the sum of two within-stratum post-stratified estimates, whereas v_1 would have been an estimate of its conditional sampling variance, had $(\hat{N}_{1,x}, \hat{N}_{0,x})$ been known to us in the first place. We may therefore define the *nonresponse effect (neff) on design* as

$$\text{neff}_{\text{dsg}} = v_1^{-1} v_0.$$

The *(overall) nonresponse effect* is conveniently given by the product of neff_{est} and neff_{dsg}, i.e.

$$\text{neff} = \text{neff}_{\text{est}} \cdot \text{neff}_{\text{dsg}} = v_1^{-1} v.$$

The neff can only be defined under an assumed nonresponse model. By comparing the neff's obtained under alternative nonresponse models, we are able to measure different assumptions against each other. In particular, we are interested in comparing a nonignorable model against an ignorable one. Under the present setting, we define the ignorable model as

$$P[R_i = 1 | x_i = x, y_i = y] = P[R_i = 1 | x_i = x]. \tag{10}$$

The method of conditional moment gives us $\hat{m}_{xy} = m_x n_{xy}/n_x$. The post-stratified estimator of T is the same with or without imputing $\{\hat{m}_{xy}\}$. Let $\text{neff}_{\text{est}}^{(pst)}$ and $\text{neff}_{\text{dsg}}^{(pst)}$ be respectively the neff on estimation and design. We have $\text{neff}_{\text{dsg}}^{(pst)} = 1$ by definition, i.e. stratification with respect to r_i has no effect at all. Recall that in (9), v_1 is calculated assuming proportional allocation in the two population strata. Let $\text{neff}_{\text{est}}^{(imp)}$ and $\text{neff}_{\text{dsg}}^{(imp)}$ be respectively the neff on estimation and design under the nonignorable model (1). We define the *effect of nonignorability* (*eff*$_n$) *for estimation* of model (1) against model (10) as

$$\text{eff}_{\text{n,est}}(imp, pst) = \text{neff}_{\text{est}}^{(imp)} / \text{neff}_{\text{est}}^{(pst)}.$$

We define the *effect of nonignorability* (eff_n) *for design* of the same pair of models as

$$\text{eff}_{\text{n,dsg}}(imp, pst) = \text{neff}_{\text{dsg}}^{(imp)}/\text{neff}_{\text{dsg}}^{(pst)} = \text{neff}_{\text{dsg}}^{(imp)}.$$

The *(overall) effect of nonignorability* of model (1) against model (10) is given by

$$\text{eff}_\text{n}(imp, pst) = \text{neff}(imp)/\text{neff}(pst) = \text{eff}_{\text{n,est}}(imp, pst) \cdot \text{eff}_{\text{n,dsg}}(imp, pst).$$

Together, neff and eff_n measure the various aspects of the effect of missing data. We may generalize formulae (7) - (9) to stratified simple random sampling, where the strata cut across the division of the sample by x under model (1) and (10). Let $g = 1, ..., G$ be the stratum-index. Bootstrap for v is the same as before, except that the stratified resampling is carried out within each s_g. The formulae (8) and (9) can easily be rewritten given $\{n_{gxy}\}$ and $\{\hat{m}_{gxy}\}$, i.e. the number of respondent and nonrespondent units from s_g with $(x_i, y_i) = (x, y)$. We estimate $\hat{m}_{xy}$ as before since the methods of conditional moment are valid for arbitrary design. We obtain $\hat{m}_{gxy}$ by the raking such that $\sum_g \hat{m}_{gxy} = \hat{m}_{xy}$ and $\sum_y \hat{m}_{gxy} = m_{gx}$. As starting values we set

$$\hat{m}_{gxy} = \hat{m}_{xy} n_{gxy} n_{xy}^{-1}.$$

So far, we have considered the case of $z_i = I_{y_i=y}$. The Bootstrap v is the same for arbitrary z_i. To obtain v_0 and v_1 in general, we impute z_i^* as follows. Conditional to (g, x), we let exactly $\hat{m}_{gxy}$ units have value y, where $\hat{m}_{gxy}$ is obtained as above. For each $i \in s$, with $(g_i, x_i, y_i^*, r_i) = (g, x, y, 0)$ where y_i^* denotes the imputed value of y_i, we draw z_i^* from $\{z_i; (g_i, x_i, y_i, r_i) = (g, x, y, 1)\}$, randomly and *with* replacement. We now estimate the sampling variance v_0^* and v_1^* based on $\{(g_i, x_i, z_i^*); i \in s\}$, where $z_i^* = z_i$ if $r_i = 1$. Repetitions give us v_0 and v_1 as the averaged values of v_0^* and v_1^*. Notice that we only use the hot-deck imputation for the analysis of neff and eff_n. Finally, for surveys with nonnegligible sampling fractions, we need to employ the finite-population correction in v_0 and v_1. Whereas for v, we must apply Bootstrap methods appropriate for the finite-population, such as those described in Shao and Sitter (1996).

3 Application

The basic idea for estimation of household composition in the absence of nonresponse can be described as follows. Let $z_i = 1, ..., Q$ be the classification of households. The sample can be grouped into $\{c_{xz}\}$, where c_{xz} is the number of persons with $(x_i, z_i) = (x, z)$. Conditional to $x_i = x$, i.e. among the subpopulation of registered families of the size x, all the persons have the same inclusion probability under the sample design of the SLC. It follows that

$$c_x^{-1} c_{xz} \qquad \text{where} \quad c_x = \sum_{q=1}^{Q} c_{xq}$$

is an estimate of the probability that a person, taken randomly from the subpopulation where $x_i = x$, lives in a household with $z_i = z$. Let N_x be the number of persons within the subpopulation with $x_i = x$. Let $I_{z_i=z} = 1$ if $z_i = z$ and $I_{z_i=z} = 0$ otherwise. We obtain

$$\hat{T}_z = \sum_x \sum_{i \in s_x} w_i I_{z_i=z} \qquad \text{where} \quad w_i = c_x^{-1} N_x \quad \text{for} \quad i \in s_x.$$

as an estimate of the number of *persons* who live in households with $z_i = z$. In case that z is the size of the household, $z^{-1}\hat{T}_z$ is an estimate of the number of households of the size z. Given nonresponse, $c_{xz} = n_{xz} + m_{xz}$, where m_{xz} is missing and needs to be estimated.

We apply the method developed in Section 2.1 - 2.3 to the data of SLC 1999. Both the observed and imputed data under model (1) are given in Table 3. Notice that the distribution of households by the household size is shifted towards the lower end among the nonrespondents, which would not have happened under the ignorable model (10). The adjustment weights are almost identical

Table 3: Sample of the SLC by the size of the family and the size of the household

Number of persons in the family	Number of persons in the household				
Respondents	1	2	3	4	≥ 5
1	565	236	30	12	6
2	37	830	49	12	5
3	57	148	460	24	9
4	54	47	100	578	18
≥ 5	26	13	19	57	366
Nonrespondents					
1	299	93	8	2	1
2	19	289	12	2	1
3	26	52	115	4	2
4	24	17	25	96	4
≥ 5	12	5	5	9	78

either by (3) or (4). Table 4 gives the estimates by (4) and (5) with $h = x$, which are equivalent to the simple post-stratified estimates based on the estimated $\{\hat{c}_{xz}\}$. The nonignorable model (1) and the ignorable model (10) differ most strongly for 1-person households, where the nonignorable model gives higher estimates both in terms of total and proportion. This is expected given the nonignorability of nonresponse. Belsby and Bjørnstad (1997) study several methods for estimation of household composition, based on the data of the Consumer Expenditure Survey 1992 with 32% nonresponse. They find that the ignorable nonresponse model (10) leads to under-estimation of 1-person households, compared to the results of the Census 1990. The bias there was about -6% for the proportion of 1-person households. In light of this it seems plausible that the estimates under the nonignorable model here are less biased.

Table 4: Estimation of the number of households by the size of the household

	Number of persons in household					
Ignorable nonresponse	1	2	3	4	≥ 5	Total
Proportion (%)	40.5	31.7	12.0	10.6	5.3	100
Total ($\times 1000$)	857	672	254	224	112	2118
Standard error ($\times 1000$)	22	12	7	5	3	14
neff_{est}	1.36	1.37	1.23	1.22	1.18	1.26
neff_{dsg}	1	1	1	1	1	1
Nonignorable nonresponse						
Proportion (%)	42.4	31.2	11.5	9.9	5.1	100
Total ($\times 1000$)	916	674	248	214	110	2163
Standard error ($\times 1000$)	25	14	9	6	3	16
neff_{est}	1.64	1.73	1.83	1.47	1.48	1.62
neff_{dsg}	1.007	1.002	1.003	1.010	1.001	1.010
$\text{eff}_{\text{n,est}}$ for estimation	1.21	1.26	1.50	1.21	1.26	1.28

Also given in Table 4 are the corresponding Bootstrap total standard errors of the estimates, as well as the neff's under both models and the effect of nonignorability for estimation. The $\text{eff}_{\text{n,dsg}}$ equals to the neff_{dsg} under the nonignorable model in this case because $\text{neff}_{\text{dsg}} = 1$ under the ignorable model. Under both models, the neff on estimation completely dominates the neff on design. Take e.g. the estimate of the total number of households under the nonignorable model, the variance increment is 62% due to neff_{est}, whereas it is only 1% due to neff_{dsg}. The systematc difference between respondents and nonrespondents (Table 3) is thus not large enough to make an impact under a stratified design. The corresponding neff under the ignorable model is 1.26, which seems to agree with the nonresponse rate of 24%. The nonignorable model leads to larger standard errors of the estimates compared to the ignorable model. Since $\text{eff}_{\text{n,dsg}} \doteq 1$ for all the estimates, the inflation of variance is almost entirely due to estimation, i.e. the difference in the imputation methods. The effect of nonignorability varies for different estimates, where the $\text{eff}_{\text{n,est}}$ is especially large for the number of 3-person households. Finally, the estimated standard errors of the total of 1-person households suggest that, the difference between the ignorable and nonignorable models is significant in this respect.

4 Summary

Standard weighting class techniques are useful estimation methods for sample surveys subject to nonresponse. However, the existing methods may not be quite effective for correcting the bias caused by nonignorable nonresponse. Less biased estimates may be obtained using the method developed in this article. It is possible to define the nonresponse model in a robust manner, even when we are unable to link all the appropriate auxiliary information to the survey. For instance, under the stratified simple random sampling, it may be plausible to simply use the stratum-index g as x under model (1). Such a model is not meant to explain all the nonresponse. It is an instrument by which we may achieve better adjustment of the bias caused by nonresponse. Contextual evidences and conceptual considerations, however, are important for judging whether the estimates are less biased under the nonignorable model than the ignorable one. Like the weighting class approach in general, our method is feasible in large-scale surveys. The neff on estimation and design have been defined in analogy to the well-known concept of deff, and are much more informative than a single nonresponse rate. Moreover, they provide a means for describing the effect of a nonignorable nonresponse assumption compared to an ignorable one. Estimation of the total variance under the stratified simple random sampling can be accomplished using the Bootstrap. For future applications it is helpful to have available practical methods of variance estimation under more complicated sample designs.

References

Belsby, L. and Bjørnstad, J.F. (1997). Modeling and estimation methods for household size in the presence of nonresponse. Technical report, Statistics Norway (Discussion Papers 206).

Efron, B. (1994). Missing data, imputation, and the Bootstrap (with discussion). *J. Amer. Statist. Assoc.*, **89**, 463–479.

Fay, R.E. (1991). A design-based perspective on missing data variance. In *Proceedings of the 1992 Annual Research Conference, U.S. Bureau of the Census*, pp. 900–905.

Forster, J.J. and Smith, P.W.F. (1998). Model-based inference for categorical survey data subject to non-ignorable non-response (with discussion). *J. Roy. Statist. Soc. B*, **60**, 57–70.

Holt, D. and Smith, T.M.F. (1979). Post stratification. *J. Roy. Statist. Soc. A*, **142**, 33–46.

Lauritzen, S.L. (1996). *Graphical Models.* Clarendon Press, Oxford.

Little, R.J.A. (1986). Survey nonresponse adjustments for estimates of means. *Int. Statist. Rev.*, **54**, 139–157.

Little, R.J.A. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data.* New York: Wiley.

Molenberghs, G., Goetghebeur, E.J.T., Lipsitz, S.R., and Kenward, M. (1999). Nonrandom missingness in categorical data: Strengths and limitations. *The American Statistician*, **53**, 110–118.

Rao, J.N.K. and Sitter, R.R. (1995). Variance estimation under two-phase sampling with application to imputation for missing data. *Biometrika*, **82**, 453–460.

Rubin, D.B. (1976). Inference and missing data. *Biometrika*, **63**(3), 581–592.

Shao, J. and Sitter, R.R. (1996). Bootstrap for imputed survey data. *J. Amer. Statist. Assoc.*, **91**, 1278–1288.

Shao, J. and Steel, P. (1999). Variance estimation for survey data with composite imputation and nonnegligible sampling fractions. *J. Amer. Statist. Assoc.*, **94**, 254–265.

Smith, P.W.F., Skinner, C.J., and Clarke, P.S. (1999). Allowing for non-ignorable non-response in the analysis of voting intention data. *Appl. Statist.*, **48**, 563–577.

Zhang, L.-C. (1999). A note on post-stratification when analyzing binary survey data subject to nonresponse. *J. Off. Statist.*, **15**, 329–334.