Discussion Papers No. 297, April 2001
Statistics Norway, Research Department

Jan F. Bjørnstad and Dag Einar Sommervoll

Modeling Binary Panel Data with Nonresponse

Abstract:
This paper studies modeling of nonignorable nonresponse in panel surveys. A class of sequential conditional logistic models for nonresponse is considered. Model-based maximum likelihood estimation and imputation are used for estimating population proportions. Various models are evaluated, and comparisons are made with traditional methods of weighting and direct data imputation. Two cases are considered, (i) the population rate of participation in the 1989 Norwegian Storting election and (ii) estimation of car ownership in Norway in 1989 and 1990.

Keywords: Nonignorable nonresponse, logistic modeling, imputation, election survey, consumer expenditure survey

JEL classification: C42, C13

Acknowledgement: The authors would like to thank Ib Thomsen for numerous discussions. Helpful suggestions by Jørgen Aasness for sections 5 and 6 are gratefully acknowledged.

Address: Jan F. Bjørnstad, Statistics Norway, Division of Statistical Methods and Standards.
E-mail: jan.bjornstad@ssb.no

Dag Einar Sommervoll, Statistics Norway, Research Department.
E-mail: dag.einar.sommervoll@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

The aim of this paper is to study modeling in panel surveys with nonresponse, where the goal is to estimate a population proportion or total. Typically, nonresponse causes biases in the estimates and should not be ignored. The only way to account for nonresponse bias is to model the response process. In this paper we study population models with a sequential logistic model for the response mechanism. Other types of models for nonresponse in panel surveys are discussed by Fay (1986, 1989) and Stasny (1987). Conaway (1993) considers a similar nonresponse model for a different type of panel data. A maximum likelihood estimator, shown to be practically the same as two prediction methods utilizing model-based imputation, is considered for estimating the population proportion. The model-based method, for various models, is compared to traditional methods of weighting and direct data imputation. The traditional methods turn out to be inferior to the model-based procedures, showing that model-driven estimation strategies can work in practice.

Two applications are considered. The first one is the estimation of the population rate of participation in the 1989 Norwegian Storting election, based on panel data from the 1985 and 1989 elections. This example is particularly well-suited for illustrative purposes of the suggested methods and models, since the 1985 and the 1989 population rates of voting are known. The second problem concerns car ownership in Norwegian households in 1989 and 1990, with panel data from the Norwegian Consumer Expenditure Survey. In the latter case we estimate the proportion of ownership in both years.

Section 2 describes the data-structure , the model and the maximum likelihood (ML) method for parameter estimation. Section 3 considers model-based ML estimation of population proportions, the imputation method and imputation-based estimators for population proportions. Section 4 describes the traditional methods for adjusting for nonresponse in panel surveys. Section 5 deals with the election panel survey, and Section 6 deals with the consumer expenditure survey.

2. A logistic model for binary panel surveys

A population of N subjects where N is known is considered. X is a 0/1-variable of interest where $X = 1$ if the subject has a certain attribute A. A panel s is selected from the population in order to observe, for each $i \in s$, X at two different times $t = 1, 2$. We are primarily interested in estimating the true proportion, P, of the attribute A in the population at $t = 2$. For each subject i in the population let

$$X_{ti} = X \text{ at time } t,\ \ t = 1,2, \text{ and } X_i = (X_{1i}, X_{2i}).$$

Then $P = \frac{1}{N}\sum_1^N X_{2i}$. Nonresponse is indicated by $R_i = (R_{1i}, R_{2i})$ where $R_{ti} = 1$ if subject *i* responds at time *t*, and 0 otherwise.

We shall assume a population model for the X_i's. To take nonresponse into account in the statistical analysis, we must model the response mechanism, i.e. the distribution of response R_i conditional on X_i. The sampling mechanism is assumed to be ignorable as is typically the case. In particular, this holds in the two examples considered. The statistical analysis is therefore done *conditional* on the total sample *s*, following the likelihood principle (see Bjørnstad, 1996). Hence, probability considerations based on the sampling design is irrelevant in the statistical analysis. This is the so-called prediction appoach. The data can be represented as in the following table.

Table 2.1. Panel with nonresponse

$t = 1 \backslash t = 2$	$X = 1$	$X = 0$	*mis*	totals
$X = 1$	n_{11}	n_{12}	n_{13}	$n_{1\circ}$
$X = 0$	n_{21}	n_{22}	n_{23}	$n_{2\circ}$
mis	n_{31}	n_{32}	n_{33}	$n_{3\circ}$
totals	$n_{\circ 1}$	$n_{\circ 2}$	$n_{\circ 3}$	n

Here, *mis* is short for missing. Moreover, n_{ij} is the number of subjects in the sample *s* belonging to the indicated category. The panel consists of the following groups, according to the response pattern:

$$\begin{aligned} s_{rr} &= \{i \in s : R_i = (1,1)\} \\ s_{rm} &= \{i \in s : R_i = (1,0)\} \\ s_{mr} &= \{i \in s : R_i = (0,1)\} \\ s_{mm} &= \{i \in s : R_i = (0,0)\}. \end{aligned}$$

2.1. The Model

The population model assumes that $X_1,....,X_N$ are independent, identically distributed. Let $p_1 = P(X_{1i} = 1), p_{11} = P(X_{2i} = 1 | X_{1i} = 1)$ and $p_{01} = P(X_{2i} = 1 | X_{1i} = 0)$. Hence, p_{11} is the conditional probability of attribute A at time $t=2$ given attribute A time $t=1$. Equivalently, we can parametrize p_{11} and p_{01} logistically,

(2.1)
$$\log\left(\frac{P(X_{2i}=1\mid X_{1i}=x_1)}{P(X_{2i}=0\mid X_{1i}=x_1)}\right)=\beta_0+\beta x_1.$$

Then

$$\beta_0=\log\left(\frac{p_{01}}{1-p_{01}}\right) \quad\text{and}\quad \beta_1=\log\left(\frac{p_{11}/(1-p_{11})}{p_{01}/(1-p_{01})}\right).$$

The advantage of the latter formulation is that β_0 and β_1 can take values on the whole real line. Possible boundary problems are therefore omitted.

The model for the response mechanism is developed through parametrizing sequentially conditional probabilities:

$$\begin{aligned}
&P(R_{1i}=r_1, R_{2i}=r_2 \mid X_{1i}=x_1, X_{2i}=x_2)\\
&=P(R_{1i}=r_1\mid X_{1i}=x_1, X_{2i}=x_2)\cdot P(R_{2i}=r_2\mid R_{1i}=r_1, X_{1i}=x_1, X_{2i}=x_2)\\
&=P(R_{1i}=r_1\mid x_1,x_2)\cdot P(R_{2i}=r_2\mid r_1,x_1,x_2).
\end{aligned}$$

Each term is modelled logistically,

(2.2)
$$\log\left(\frac{P(R_{1i}=1\mid x_1,x_2)}{P(R_{1i}=0\mid x_1,x_2)}\right)=\phi_0^{(1)}+\phi_1^{(1)}x_1+\phi_2^{(1)}x_2$$

(2.3)
$$\log\left(\frac{P(R_{2i}=1\mid r_1,x_1,x_2)}{P(R_{2i}=0\mid r_1,x_1,x_2)}\right)=\phi_0^{(2)}+\phi_1^{(2)}r_1+\phi_2^{(2)}x_1+\phi_3^{(2)}x_2$$

Contingency table 2.1 has 8 free cell probabilities. The model (2.1)-(2.3), with p_1, has introduced 10 parameters. For the model to be estimable we need to reduce the number of parameters to a maximum of 8. This can de done in several ways, giving rise to different models as seen in the two applications.

The population model assumes independence between sampled units. The two surveys considered in the examples use a two-step sampling design by first selecting geographical areas (clusters) and then selecting units within each sampled area. An alternative and possibly more appropriate model could have been to assume correlation within clusters. However, the data for two cases were not available on "cluster form". Also for the two *variables* considered here, voting behaviour and car ownership, the independence assumption should work well as a model for analysis. Certainly, when the data are on cluster form, the multi-level modeling approach is an interesting alternative that should be tried.

2.2. Maximum likelihood parameter estimation

We shall consider estimation of the unknown parameters (no more than 8) in model (2.1)-(2.3). Let us consider the likelihood function, i.e. the probability of the observed data as function of the parameters, given by

$$L(\boldsymbol{\beta},\boldsymbol{\phi}^{(1)},\boldsymbol{\phi}^{(2)}) = L_{rr}\cdot L_{rm}\cdot L_{mr}\cdot L_{mm}$$

where

$$\begin{aligned}
L_{rr} &= \prod_{i\in s_{rr}} P\left(X_{1i}=x_{1i}, X_{2i}=x_{2i}, R_i=(1,1)\right)\\
&= \prod_{i\in s_{rr}} p_1^{x_{1i}}\left(1-p_1\right)^{1-x_{1i}}\left(\frac{1}{1+e^{-(\beta_0+\beta_1 x_{1i})}}\right)^{x_{2i}}\left(\frac{1}{1+e^{\beta_0+\beta_1 x_{1i}}}\right)^{1-x_{2i}}\cdot\frac{1}{1+e^{-\left(\phi_0^{(1)}+\phi_1^{(1)}x_{1i}+\phi_2^{(1)}x_{2i}\right)}}\\
&\quad\cdot\frac{1}{1+e^{-\left(\phi_0^{(2)}+\phi_1^{(2)}+\phi_2^{(2)}x_{1i}+\phi_3^{(2)}x_{2i}\right)}}
\end{aligned}$$

$$\begin{aligned}
L_{rm} &= \prod_{i\in s_{rm}} P\left(X_{1i}=x_{1i}, R_i=(1,0)\right)\\
&= \prod_{i\in s_{rm}}\sum_{x_{2i}=0}^{1}\left\{p_1^{x_{1i}}\left(1-p_1\right)^{1-x_{1i}}\left(\frac{1}{1+e^{-(\beta_0+\beta_1 x_{1i})}}\right)^{x_{2i}}\left(\frac{1}{1+e^{\beta_0+\beta_1 x_{1i}}}\right)^{1-x_{2i}}\cdot\frac{1}{1+e^{-\left(\phi_0^{(1)}+\phi_1^{(1)}x_{1i}+\phi_2^{(1)}x_{2i}\right)}}\right.\\
&\quad\left.\cdot\frac{1}{1+e^{\phi_0^{(2)}+\phi_1^{(2)}+\phi_2^{(2)}x_{1i}+\phi_3^{(2)}x_{2i}}}\right\}
\end{aligned}$$

$$\begin{aligned}
L_{mr} &= \prod_{i\in s_{mr}} P\left(X_{2i}=x_{2i}, R_i=(0,1)\right)\\
&= \prod_{i\in s_{mr}}\sum_{x_{1i}=0}^{1}\left\{p_1^{x_{1i}}\left(1-p_1\right)^{1-x_{1i}}\left(\frac{1}{1+e^{-(\beta_0+\beta_1 x_{1i})}}\right)^{x_{2i}}\left(\frac{1}{1+e^{\beta_0+\beta_1 x_{1i}}}\right)^{1-x_{2i}}\cdot\frac{1}{1+e^{\phi_0^{(1)}+\phi_1^{(1)}x_{1i}+\phi_2^{(1)}x_{2i}}}\right.\\
&\quad\left.\cdot\frac{1}{1+e^{-\left(\phi_0^{(2)}+\phi_2^{(2)}x_{1i}+\phi_3^{(2)}x_{2i}\right)}}\right\}
\end{aligned}$$

$$L_{mm} = \prod_{i \in s_{mm}} P\left(R_i = (0,0)\right)$$

$$= \prod_{i \in s_{mr}} \sum_{x_{1i}=0}^{1} \sum_{x_{2i}=0}^{1} \left\{ p_1^{x_{1i}} (1-p_1)^{1-x_{1i}} \left(\frac{1}{1+e^{-(\beta_0+\beta_1 x_{1i})}} \right)^{x_{2i}} \left(\frac{1}{1+e^{\beta_0+\beta_1 x_{1i}}} \right)^{1-x_{2i}} \cdot \frac{1}{1+e^{\phi_0^{(1)}+\phi_1^{(1)} x_{1i}+\phi_2^{(1)} x_{2i}}} \right.$$

$$\left. \cdot \frac{1}{1+e^{\phi_0^{(2)}+\phi_2^{(2)} x_{1i}+\phi_3^{(2)} x_{2i}}} \right\}.$$

Estimates are found by maximizing log(L) numerically using NAG subroutine E04JAF (described in the NAG Fortran Library Manual March 11, 1984). To estimate the standard error (S.E.) of the maximum likelihood (ML) estimates $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}^{(1)}, \hat{\boldsymbol{\phi}}^{(2)})$, we use parametric bootstrapping (see Efron and Tibshirani (1993, ch.6.5)) by simulating 1000 sets of data assuming $\left(\boldsymbol{\beta}, \boldsymbol{\phi}^{(1)}, \boldsymbol{\phi}^{(2)}\right) = \left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}^{(1)}, \hat{\boldsymbol{\phi}}^{(2)}\right)$. The estimated S.E. of a given estimate is then the empirical standard deviation of this estimate. For example, consider $\hat{\beta}_0$. Let $\hat{\beta}_{0,1}, \ldots, \hat{\beta}_{0,1000}$ be the set of estimated values based on the simulated data. The estimated S.E. is then given by, with $\bar{\hat{\beta}}_0 = \sum_{i=1}^{k} \hat{\beta}_{0,i} / k$ and $k = 1000$,

$$\left[\frac{1}{k-1} \sum_{i=1}^{k} \left(\hat{\beta}_{0,i} - \bar{\hat{\beta}}_0 \right)^2 \right]^{1/2}$$

The simulated mean $\bar{\hat{\beta}}_0$ estimates $E(\hat{\beta}_0)$ at $\theta = \hat{\theta}$. From a simulation study it seems that the ML estimates are approximately unbiased.

3. Estimation of attribute proportion at time *t* = 2

An estimator of P, disregarding the nonresponse groups, is the proportion of A at $t = 2$ among the s_{rr} respondents,

(3.1) $$\hat{P}_{rr} = \frac{n_{11} + n_{21}}{n_{rr}}$$

where n_{rr} is the number of subjects in the survey who respond on both occasions, $n_{rr} = \#(s_{rr}) = n_{11} + n_{21} + n_{12} + n_{22}$. Let π_{ij}, $i = 1,2,3$ and $j = 1,2,3$, be the cell probabilities of table 1. Then, conditionally on n_{rr}, and hence also unconditionally,

$$E\left(\hat{P}_{rr}\right)=\frac{\pi_{11}+\pi_{21}}{\pi_{11}+\pi_{21}+\pi_{12}+\pi_{22}}.$$

We see that $E\left(X_{2i}\right)=P\left(X_{2i}=1\right)=p_1 p_{11}+\left(1-p_1\right)p_{01}$ such that

(3.2) $$E(P)=p_1 p_{11}+\left(1-p_1\right)p_{01}.$$

It follows that $\hat{P}_{rr}$ is unbiased if and only if

(3.3) $$\frac{\pi_{11}+\pi_{21}}{\pi_{11}+\pi_{21}+\pi_{12}+\pi_{22}}=p_1 p_{11}+\left(1-p_1\right)p_{01}.$$

It can be shown that (3.3) is equivalent to

(3.4) $$\phi_1^{(1)}=\phi_2^{(1)}=\phi_2^{(2)}=\phi_3^{(2)}=0$$

i.e., that $P\left(R_i=\left(r_1,r_2\right)\mid X_i=x_i\right)$ is independent of x_i. This means that the response mechanism is ignorable, which is rarely the case. Hence, typically $\hat{P}_{rr}$ will be a biased estimator of *P*. In our first application on voting participation it turns out that $\hat{P}_{rr}$ overestimates *P* by a wide margin.

Including the response mechanism into the analysis, we shall use the maximum likelihood estimator under the model (2.1)-(2.3), assuming $p_1=P\left(X_{1i}=1\right)$ is known. It is shown that this estimator is identical to an imputation-based estimator under a saturated model of 8 unknown parameters. We also present a second imputation-based estimator that differs from the ML estimator by no more than *n*/*N*.

Since, from (3.2), $E(P)=p_1 p_{11}+\left(1-p_1\right)p_{01}$, the ML estimator is given by

(3.5) $$\hat{P}_{ML}=p_1\hat{p}_{11}+\left(1-p_1\right)\hat{p}_{01}$$

where $\hat{p}_{11},\hat{p}_{01}$ are ML estimates.

A common approach to correct for nonresponse is by imputation of the missing values in the sample. The method of imputation is to assign the estimated expected value conditional on nonresponse. Others who have used this method include Greenlees et al. (1982) and Bjørnstad & Walsøe (1991). We can express $P=t/N$ where $t=\sum_{i=1}^{N}X_{2i}$. In the case of complete data, i.e., $s_{rr}=s$, the optimal unbiased estimator of *t* is, from Thomsen (1981), given by

(3.6) $$\hat{t} = N\left(p_1 \hat{p}_{11}^{(c)} + (1 - p_1)\, \hat{p}_{01}^{(c)} \right)$$

where $\hat{p}_{11}^{(c)}, \hat{p}_{01}^{(c)}$ are the ML estimates, i.e.,

(3.7) $$\hat{p}_{11}^{(c)} = \frac{\sum_s X_{1i} X_{2i}}{\sum_s X_{1i}}$$

(3.8) $$\hat{p}_{01}^{(c)} = \frac{\sum_s (1 - X_{1i}) X_{2i}}{\sum_s (1 - X_{1i})}.$$

When we have nonresponse, the missing values in *s* are imputed and an imputation-based estimator is then $\hat{t}$ and the corresponding *P*-estimator computed for the "imputed" completed sample. I.e., we impute the unkown values in $\hat{p}_{11}^{(c)}$ and $\hat{p}_{01}^{(c)}$. Let $\hat{P}$ denote probability under the estimates $\hat{\boldsymbol{\theta}}$, and let $\hat{p}_{11,I}^{(c)}, \hat{p}_{01,I}^{(c)}$ be the imputation-based versions of $\hat{p}_{11}^{(c)}$ and $\hat{p}_{01}^{(c)}$. Then the imputation-based estimators of *P* and *t* become

$$\hat{P}_I = p_1 \hat{p}_{11,I}^{(c)} + (1 - p_1)\, \hat{p}_{01,I}^{(c)} \quad \text{and} \quad \hat{t}_I = N p_1 \hat{p}_{11,I}^{(c)} + N(1 - p_1)\, \hat{p}_{01,I}^{(c)}.$$

Using model (2.1)-(2.3) we obtain the imputed values: For $i \in s_{rm}$: $X_{2i}^* = \hat{P}\left(X_{2i} = 1 \mid X_{1i}, R_i = (1,0)\right)$, for $i \in s_{mr}$: $X_{1i}^* = \hat{P}\left(X_{1i} = 1 \mid X_{2i}, R_i = (0,1)\right)$, and for $i \in s_{mm}$: $X_{2i}^* = \hat{P}\left(X_{2i} = 1 \mid R_i = (0,0)\right)$, $X_{1i}^* = \hat{P}\left(X_{1i} = 1 \mid R_i = (0,0)\right)$ and $\left(X_{1i} X_{2i}\right)^* = \hat{P}\left(X_{1i} = 1, X_{2i} = 1 \mid R_i = (0,0)\right)$. With a saturated model of 8 unknown parameters, the fit of the data (by taking estimated expected values of the n_{ij}'s) is perfect. Then $\hat{P}_{ML} = \hat{P}_I$ (shown in the appendix).

An alternative to (3.6) as a basic estimator in the case of complete data is achieved by noting that (with $\bar{s} = \{i : i \notin s\}$) $t = \sum_s X_{2i} + \sum_{\bar{s}} X_{2i}$, $\sum_s X_{2i}$ is observed and $z = \sum_{\bar{s}} X_{2i}$ can be estimated by estimating $E\left(\sum_{\bar{s}} X_{2i}\right) = (N - n) P\left(X_{2i} = 1\right) = (N - n)\left(p_1 p_{11} + (1 - p_1)\, p_{01}\right)$.

Hence, a complete data estimator is given by

(3.9) $$\hat{t}^{(c)} = \sum_s X_{2i} + (N - n)\left(p_1 \hat{p}_{11}^{(c)} + (1 - p_1)\, \hat{p}_{01}^{(c)} \right).$$

When we have nonresponse we can represent t as

$$t = \sum_{s_{rr}} X_{2i} + \sum_{s_{mr}} X_{2i} + \sum_{s_{rm}} X_{2i} + \sum_{s_{mm}} X_{2i} + z\,.$$

$z = \sum_{\bar{s}} X_{2i}$ is estimated by $\hat{z} = (N-n)\left(p_1 \hat{p}_{11} + (1-p_1)\,\hat{p}_{01}\right)$. That is, we replace $\hat{p}_{11}^{(c)}, \hat{p}_{01}^{(c)}$ by the current ML estimates $\hat{p}_{11}, \hat{p}_{01}$. The missing X_{2i} are imputed as before giving us the imputation-based estimator

$$\hat{t}_I^{(c)} = \sum_{s_{rr}} X_{2i} + \sum_{s_{mr}} X_{2i} + \sum_{s_{rm}} X_{2i}^* + \sum_{s_{mm}} X_{2i}^* + (N-n)\left(p_1 \hat{p}_{11} + (1-p_1)\,\hat{p}_{01}\right) \text{ and } \hat{P}_I^{(c)} = \hat{t}_I^{(c)} / N\,.$$

$\hat{P}_I^{(c)}$ and $\hat{P}_{ML}$ will give approximately the same results. In fact, we always have the bound $\left|\hat{P}_I^{(c)} - \hat{P}_{ML}\right| \le n/N$ (shown in the appendix). In our cases , the maximal difference is less than 10^{-3}.

In addition to being based on different complete data estimators (3.6) and (3.9), the imputation is also done differently in $\hat{t}_I$ and $\hat{t}_I^{(c)}$. In $\hat{t}_I^{(c)}$ we impute only in $\sum_s X_{2i}$, while for $\hat{t}_I$ all missing values in $\hat{t}$ are imputed. Typically, however, $\hat{P}_I^{(c)}$ and $\hat{P}_I$ give approximately the same results as indicated by the comparisons to $\hat{P}_{ML}$.

4. Traditional methods based on weighting and direct data imputation

We shall compare the modeling approach with traditional weighting and imputation methods that do not require a specific model for the response mechanism. Reviews of weighting and direct data imputation in panel surveys can be found in Kalton (1986) and Lepkowski (1989). We consider one imputation method and four weighting-based methods. Each method is equivalent to constructing a certain adjusted 2×2-table; either for s or s-s_{mm} as shown in table 4.1.

Table 4.1. Adjusted panel without nonresponse

t = 1\ t = 2	X = 1	X = 0	totals
X = 1	n_{11}^*	n_{12}^*	$n_{1\circ}^*$
X = 0	n_{21}^*	n_{22}^*	$n_{2\circ}^*$
totals	$n_{\circ 1}^*$	$n_{\circ 2}^*$	n^*

Here, $n^* = |s| = n$ or $n^* = |s - s_{mm}| = n - n_{33}$. Table 4.1 is then used in (3.7) and (3.8) to produce estimates of p_{11} and p_{01}, $\hat{p}_{11}^* = n_{11}^* / n_{1\circ}^*$, $\hat{p}_{01}^* = n_{21}^* / n_{2\circ}^*$. From (3.6) it follows that in the case of known p_1, the *P*-estimate is given by

(4.1) $$\hat{P}_e^* = p_1 \hat{p}_{11}^* + (1 - p_1)\, \hat{p}_{01}^* .$$

When p_1 is unknown it is estimated by $\hat{p}_1 = n_{1\circ}^* / n^*$. Then (4.1) is modified to

(4.2) $$\hat{P}^* = \hat{p}_1 \hat{p}_{11}^* + (1 - \hat{p}_1)\, \hat{p}_{01}^* = n_{\circ 1}^* / n^*$$

which corresponds to $\hat{P}_{rr}$ based on s_{rr} (see (3.1)). Of course, $\hat{P}^*$ is an estimator of *P* also when p_1 is known, but $\hat{P}_e^*$ is a theoretically better estimator. Also, for the case considered in this paper $\hat{P}_e^*$ actually works better.

4.1. Direct data imputation

The imputation method discards s_{mm} and employs mean stratified imputation in the other nonresponse groups. Missing values of $X_{2i}, i \in s_{rm}$, are imputed as mean of observed X_{2i}-values given X_{1i}:

Given $$X_{1i} = 1: X_{2i}^* = \frac{n_{11}}{n_{11} + n_{12}} .$$

Given $$X_{1i} = 0: X_{2i}^* = \frac{n_{21}}{n_{21} + n_{22}} .$$

Similarly, missing values for $X_{1i}, i \in s_{mr}$, are imputed as the mean of observed X_{1i}-values given X_{2i}. Let a_1, a_2, a_3 be the inverses of the response rates for the rows in table 2.1 corresponding to X_{1i} = 1,0, *mis*. Similarly b_1, b_2, b_3 are the inverse response rates for the columns corresponding to X_{2i}-values.

$$a_i = \frac{n_{i\circ}}{n_{i1} + n_{i2}}$$

$$b_j = \frac{n_{\circ j}}{n_{1j} + n_{2j}} .$$

The constructed imputed 2×2-table is given below.

Table 4.2. Imputed table, without s_{mm}

	$X_2 = 1$	$X_2 = 0$	Totals
$X_1 = 1$	$(a_1 + b_1 - 1)n_{11}$	$(a_1 + b_2 - 1)n_{12}$	$b_1 n_{11} + b_2 n_{12} + n_{13}$
$X_1 = 0$	$(a_2 + b_1 - 1)n_{21}$	$(a_2 + b_2 - 1)n_{22}$	$b_1 n_{21} + b_2 n_{22} + n_{23}$
Totals	$a_1 n_{11} + a_2 n_{21} + n_{31}$	$a_1 n_{12} + a_2 n_{22} + n_{32}$	$n - n_{33}$

We note that mean imputation for 0/1-variables is equivalent to assigning value 1 to a proportion equal to the mean in a given stratum. E.g., given $X_{1i} = 1$, $\frac{n_{11}}{n_{11}+n_{12}} \cdot n_{13}$ of the X^*_{2i}-values in s_{rm} are equal to 1, the rest is 0. We see that the imputation-based estimates $\hat{p}^*_{11}$ and $\hat{p}^*_{01}$ are as follows.

$$\hat{p}^*_{11} = \frac{(a_1 + b_1 - 1)n_{11}}{b_1 n_{11} + b_2 n_{12} + n_{13}}, \quad \hat{p}^*_{01} = \frac{(a_2 + b_1 - 1)n_{21}}{b_1 n_{21} + b_2 n_{22} + n_{23}}.$$

Let $\hat{P}^*_{e,I}$ and $\hat{P}^*_I$ denote the *P*-estimates given by (4.1) and (4.2) for this imputation method.

4.2. Weighting

The methods of weighting are all based on weighing observed responses to account for the nonresponse groups. The weights are equal to inverses of response rates in certain adjustment cells. One traditional weighing scheme is to weigh s_{rr} - data to account for the nonresponse groups s_{rm} , s_{mr} and s_{mm}. This can be done in two different ways. One way is to first account for s_{rm} and s_{mm} by weighing s_{rr} - data using X_1 as auxiliary variable, and then weigh the adjusted 3×2 - table to account for s_{mr} , using X_2 as auxiliary variable. Hence we have adjustment cells according to X_1 = (1,0, *mis*) with the weights:

$$\text{Row } i \ (n_{i1}, n_{i2}) \text{ gets the weights } a_i, \text{ for } i = 1,2,3.$$

The row-weighting to account for s_{rm} and s_{mm} produces the following table.

Table 4.3. Row-weighted table

	X_2 = 1	X_2 = 0	Totals
X_1 = 1	$a_1 n_{11}$	$a_1 n_{12}$	$n_{1\circ}$
X_1 = 0	$a_2 n_{21}$	$a_2 n_{22}$	$n_{2\circ}$
X_1 = *mis*	$a_3\, n_{31}$	$a_3\, n_{32}$	$n_{3\circ}$
Totals	$a_1 n_{11} + a_2 n_{21} + a_3 n_{31}$	$a_1 n_{12} + a_2 n_{22} + a_3 n_{32}$	n

The weights on the second step to account for $X_1 = mis$ are then :

$$\text{first column weight} = \frac{a_1 n_{11} + a_2 n_{21} + a_3 n_{31}}{a_1 n_{11} + a_2 n_{21}}$$

$$\text{second column weight} = \frac{a_1 n_{12} + a_2 n_{22} + a_3 n_{32}}{a_1 n_{12} + a_2 n_{22}}.$$

The final weighted-adjusted 2×2-table, called the W_1-method, is given below:

Table 4.4. Weighted table, row-column

	X_2 = 1	X_2 = 0	Totals
X_1 = 1	$(1+f(1))a_1 n_{11}$	$(1+f(2))a_1 n_{12}$	$n_{1\circ} + a_1(n_{11}f(1) + n_{12}f(2))$
X_1 = 0	$(1+f(1))a_2 n_{21}$	$(1+f(2))a_2 n_{22}$	$n_{2\circ} + a_2(n_{21}f(1) + n_{22}f(2))$
Totals	$a_1 n_{11} + a_2 n_{21} + a_3 n_{31}$	$a_1 n_{12} + a_2 n_{22} + a_3 n_{32}$	n

Here, $f(j) = a_3 n_{3j} \big/ \left(a_1 n_{1j} + a_2 n_{2j}\right)$. The corresponding P-estimates given by (4.1) and (4.2) are denoted by $\hat{P}^*_{e,W_1}$ and $\hat{P}^*_{W_1}$ respectively.

Instead of weighing the rows first we can reverse the order and first weigh s_{rr} to account for s_{mr} and s_{mm} by giving the columns the weights b_1,b_2,b_3 and then weighing the rows of the adjusted table. This column-row scheme is called the W2-method and the corresponding P-estimates given by (4.1) and (4.2) are denoted by $\hat{P}^*_{e,W_2}$ and $\hat{P}^*_{W_2}$ respectively.

Two other weighting methods are similar to W1 and W2, the difference being that they disregard s_{mm} and adjust s - s_{mm} in the same way as W1 and W2 adjust the whole sample s. In the two cases we consider they give practically the same results as the mean imputation method in Section 4.1, and we shall not consider these any further.

5. The election panel survey

For illustrative purposes we shall now consider a panel survey where the population totals of *A* are known at both times. This case concerns the rate of participation in the 1989 Norwegian Storting election, based on panel data from the 1985 and 1989 elections. Table 5.1 below gives the data.

Table 5.1. Panel data for election survey

1985\1989	voted	did note vote	*mis*	totals
voted	743	36	188	967
did not vote	42	20	26	88
mis	115	20	162	297
totals	900	76	376	1352

We shall estimate the voting proportion *P* in 1989 by making use of the known voting proportion in 1985, p_1 = 0.838. From the actual 1989 election we know the true value of *P*, 0.832. It is of interest to see how the maximum likelihood estimator $\hat{P}_{ML}$, based on different models, behave in this particular case. This gives us a way to evaluate various models, and gives us some indication on what may be appropriate models for similar problems in the future. We shall also see how this estimator compares to the traditional methods of accounting for nonresponse in Section 4 as well as the estimator $\hat{P}_{rr}$ and a poststratified estimator based solely on the response sample s_{rr}. It turns out that we do need to include a nonignorable model for the response mechansim (RM).

5.1. Traditional methods and poststratification

In addition to the traditional methods from Section 4 and the rate $\hat{P}_{rr}$ of voting in s_{rr}, we shall consider the *s*-optimal estimator $\hat{P}^{(c)}$, given by (3.6), based on the data in s_{rr}. It is given by

$$\hat{P}^{(r)} = p_1 \hat{p}_{11}^{(r)} + (1 - p_1)\, \hat{p}_{01}^{(r)}$$

where $\hat{p}_{11}^{(r)} = n_{11} / (n_{11} + n_{12})$ and $\hat{p}_{01}^{(r)} = n_{21} / (n_{21} + n_{22})$. We see that $\hat{P}^{(r)}$ is the poststratified estimator using X_1 as the stratifying variable. Both $\hat{P}_{rr}$ and $\hat{P}^{(r)}$ assume implicitly ignorable response mechanism (RM). These two estimators together with the methods described in Section 4, to adjust for nonresponse, give the following estimates.

Table 5.2. Traditional estimates of attribute proportion

Method	p_{11} - estimate	p_{01} - estimate	P- estimate
$\hat{P}_{rr}$	-	-	0.933
$\hat{P}^{(r)}$	0.954	0.677	0.909
Mean imputation	0.9471	0.6493	0.899
W1	0.9419	0.6224	0.890
W2	0.9458	0.6395	0.896

Clearly, all these estimators overestimate *P*. Comparing $\hat{P}^{(r)}$ and $\hat{P}_{rr}$, it seems that poststratification corrects for some of the bias, while at the same time indicating that part of the bias is due to nonignorable nonresponse. The traditional methods of adjusting for nonresponse improve only slightly on the purely s_{rr}-based methods. It seems clear that the RM cannot be ignored and that we do need to include a nonignorable model for RM in the analysis. In the next section we shall look at the model-based estimator $\hat{P}_{ML}$, given by (3.5), for three different models.

5.2. Maximum likelihood estimation under nonignorable response models

The model (2.1)-(2.3) has 9 unknown parameters and we need to reduce the number of parameters to no more than 8. This can be done in several ways giving rise to different models.

Model 1 $\phi_2^{(1)} = 0$.

This amounts to the reasonable assumption that the probability of response the first time does not depend on the voting behaviour at the second election. Note, however, that this is equivalent with assuming that voting behaviour in 1989 is not related to the response behaviour in 1985, conditional on voting behaviour in 1985.

Model 2 $\phi_2^{(2)} = 0$

In this model we keep (2.1) and (2.2) and reduce (2.3). Voting behaviour in the first election does not affect the probability of response the second time. We do, however, assume that voting behaviour in the second election and response in the first may be related.

Model 3 $\phi_2^{(1)} = 0$, $\phi_2^{(2)} = 0$

Here, response at either time depends only on the voting behaviour at that time.

The ML parameter estimates and the corresponding estimated SE (in parentheses) are given in the following table.

Table 5.3. Maximum likelihood estimates in election models

Parameter	Model 1	Model 2	Model 3
β_0	0.766 (0.484)	0.049 (0.387)	0.292 (0.286)
β_1	2.27 (0.346)	2.48 (0.298)	2.42 (0.286)
p_{11}	0.954 (0.021)	0.926 (0.027)	0.937 (0.014)
p_{01}	0.678 (0.104)	0.5125 (0.092)	0.572 (0.068)
$\phi_0^{(1)}$	-0.377 (0.169)	-0.630 (0.281)	-0.403 (0.172)
$\phi_1^{(1)}$	2.12 (0.243)	1.99 (0.352)	2.17 (0.247)
$\phi_2^{(1)}$	–	0.443 (0.475)	–
$\phi_0^{(2)}$	-0.445 (2.264)	-1.21 (1.03)	-1.01 (0.357)
$\phi_1^{(2)}$	1.369 (0.188)	1.36 (0.197)	1.45 (0.149)
$\phi_2^{(2)}$	0.574 (0.512)	–	–
$\phi_3^{(2)}$	-0.080 (2.495)	1.40 (1.17)	1.05 (0.446)

We note that $\phi_1^{(1)}$ is significantly different from 0 under all three models. This indicates that response behaviour in 1985 depends on the voting behaviour in the same year. Also, clearly $\phi_1^{(2)} \neq 0$ and the response behaviour in 1985 and 1989 are correlated. The main difference between the models regarding how $\phi^{(1)}$ and $\phi^{(2)}$ are estimated concerns $\phi_3^{(2)}$. Under Model 1 it seems that voting behaviour in 1989 does not affect the response behaviour. This does not seem reasonable from earlier experiences regarding voting behaviour (see, e.g., Thomsen and Siring, 1983). The parameters for estimating P are p_{11} and p_{01}. Recall that the s_{rr}-estimates are $\hat{p}_{11}^{(r)}$ = 0.954 and $\hat{p}_{01}^{(r)}$ = 0.677 (with $\hat{P}^{(r)}$ = 0.909). Under the ignorable RM-model (3.4), the ML estimates of p_{11} and p_{01} are 0.950 and 0.635 respectively, with P-estimate equal to 0.899. We note that Model 2 and Model 3 estimate p_{01}

significantly lower than $\hat{p}_{01}^{(r)}$, while Model 1 does not. This affects the *P*-estimates significantly as we see below.

Models 1 and 2 give perfect fits, and Model 3 gives a nearly perfect fit. We know then from Section 3, that as a consequence, the three estimators $\hat{P}_{ML}, \hat{P}_I$ and $\hat{P}_I^{(c)}$ will give approximately equal estimates and only $\hat{P}_{ML}$ is given below for the different models. The estimated SE are given in parentheses.

Estimate of *P* (=0.832)	Model 1	Model 2	Model 3
$\hat{P}_{ML}$	0.909 (0.034)	0.859 (0.034)	0.878 (0.019)

5.3. Model comparisons

The saturated Models 1 and 2 give perfect fit of the data to the models. Model 3 gives a nearly perfect fit. Therefore, we cannot evaluate and compare the models by traditional goodness-of-fit criteria. Note that goodness-of-fit testing in contingency tables is concerned with estimating the cell probabilities $\boldsymbol{\pi} = (\pi_{ij}; i, j = 1,2,3)$. Models 1,2 will give the ML estimates $\hat{\pi}_{ij} = n_{ij}/n$, while Model 3 has $\hat{\pi}_{ij} \approx n_{ij}/n$. Our goal for these models is, however, not to estimate $\boldsymbol{\pi}$, but rather *P* or equivalently $E(P) = P(X_{2i} = 1)$. Hence, we should evaluate the models with this in mind. Now,

(5.1) $$P(X_{2i} = 1) = P(R_{2i} = 1)P(X_{2i} = 1 | R_{2i} = 1) + P(R_{2i} = 0)P(X_{2i} = 1 | R_{2i} = 0).$$

In terms of $\boldsymbol{\pi}$, $P(R_{2i} = 1) = \pi_{\circ 1} + \pi_{\circ 2}$, where $\pi_{\circ j} = \pi_{1j} + \pi_{2j} + \pi_{3j}$. Furthermore, $P(X_{2i} = 1 | R_{2i} = 1) = \pi_{\circ 1}/(\pi_{\circ 1} + \pi_{\circ 2})$. Saturated models all have the same ML estimate of $\pi_{\circ j}$, $\hat{\pi}_{\circ j} = n_{\circ j}/n$. It follows from (5.1) that saturated models estimate $P(X_{2i} = 1)$ by:

$$\frac{n_{\circ 1}}{n} + \frac{n_{\circ 3}}{n}\hat{P}(X_{2i} = 1 | R_{2i} = 0)$$

where $\hat{P}(X_{2i} = 1 | R_{2i} = 0)$ is the ML estimate. Since Model 3 is approximately saturated, it follows that, for estimating *P*, the three models differ only in how $P(X_{2i} = 1 | R_{2i} = 0)$ is estimated. We would expect that $P(X_{2i} = 1 | R_{2i} = 0)$ is not too different from $P(X_{1i} = 1 | R_{1i} = 0)$. The rate of voting among the nonrespondents may, however, increase slightly with time, since the panel is aging. It is well known that voting participation among young voters is smaller than the population rate. Furthermore, among the young voters there is a lower rate of voting in the nonresponse group (see Thomsen and Siring, 1983). The point now is that, when p_1 is known, the ML estimate of $P(X_{1i}=1 \mid R_{1i}=0)$ is identically the same for all saturated models and is given by

$$(5.2) \qquad \hat{P}(X_{1i}=1 \mid R_{1i}=0) = \frac{np_1 - n_{1\circ}}{n_{3\circ}}.$$

This is seen as follows. Obviously (5.1) holds for (X_{1i}, R_{1i}) and the ML estimates of $P(R_{1i}=1)$ and $P(X_{1i}=1 \mid R_{1i}=1)$ are $\hat{\pi}_{1\circ}+\hat{\pi}_{2\circ}=(n_{1\circ}+n_{2\circ})/n$ and $\hat{\pi}_{1\circ}/(\hat{\pi}_{1\circ}+\hat{\pi}_{2\circ})=n_{1\circ}/(n_{1\circ}+n_{2\circ})$ respectively. Hence, from (5.1),

$$p_1 = \frac{n_{1\circ}}{n} + \frac{n_{3\circ}}{n}\hat{P}(X_{1i}=1 \mid R_{1i}=0)$$

and (5.2) follows. We conclude that a criterion for evaluating and comparing (nearly) saturated models aimed at estimating *P* is given by

$$(5.3) \qquad \left|\hat{P}(X_{2i}=1 \mid R_{2i}=0) - \hat{P}(X_{1i}=1 \mid R_{1i}=0)\right| = \left|\hat{P}(X_{2i}=1 \mid R_{2i}=0) - \frac{np_1 - n_{1\circ}}{n_{3\circ}}\right|$$

In the election panel survey, the estimated voting rates in the subpopulations of respondents for the two elections are 0.917 for 1985 and 0.922 in 1989, in all three models. For the nonrespondents the estimated voting rates are given below.

Table 5.4. Estimated voting rates for nonrespondents

	Model 1	Model 2	Model 3
$\hat{P}(X_{1i}=1 \mid R_{1i}=0)$	0.559	0.559	0.554
$\hat{P}(X_{2i}=1 \mid R_{2i}=0)$	0.882	0.695	0.770

Based on (5.3), Model 2 is clearly to be preferred among the three models. Of course, knowing *P* in this case makes it easy to confirm that Model 2 works best, but even if *P* was not known we would make the same evaluation based on (5.3). It seems clear that the voting participation in the 89 election among nonrespondents is overestimated by Model 3 and especially Model 1. The rate of voting of the nonrespondents does, however, seem to increase with time as expected.

Comparing $\hat{P}_{ML}$ to $\hat{P}^{(r)}$ and the traditional nonresponse-adjustment methods $\hat{P}^*_{e,I}$, $\hat{P}^*_{e,W_1}$, $\hat{P}^*_{e,W_2}$ gives additional support to the contention that Model 1 does not work. It does not correct for the bias due to nonresponse that we know is present.

Let us now consider the modeling aspects for the distribution of R_{1i} given X_{1i} and X_{2i}. Using the ML estimates from table 5.3, we find the following estimates (with estimated SE in parentheses)[1]:

Table 5.5. Estimated conditional probabilities of response

	Model 1	Model 2	Model 3
$\hat{P}(R_{1i}=1 \mid X_{1i}=1, X_{2i}=1)$	0.854 (0.015)	0.858 (0.015)	0.856 (0.15)
$\hat{P}(R_{1i}=1 \mid X_{1i}=1, X_{2i}=0)$	0.854 (0.015)	0.795 (0.071)	0.856 (0.15)
$\hat{P}(R_{1i}=1 \mid X_{1i}=0, X_{2i}=1)$	0.402 (0.041)	0.453 (0.080)	0.396 (0.041)
$\hat{P}(R_{1i}=1 \mid X_{1i}=0, X_{2i}=0)$	0.402 (0.041)	0.347 (0.065)	0.396 (0.041)

We see, by comparing Model 1 and Model 3, that assuming $\phi_2^{(2)}=0$ in addition to $\phi_2^{(1)}=0$ has little effect on these conditional response probabilities. Comparing these models to Model 2 indicates that R_{1i} may depend slightly on X_{2i} even when X_{1i} is known. This dependence has the effect of lowering the response probability for those who did not participate in the 1989 election.

Looking at the estimated distribution of R_{2i} given X_{1i}, X_{2i}, R_{1i} for Model 1 we find that $\hat{P}(R_{2i}=1 \mid R_{1i}=1, X_{1i}, X_{2i})$ varies from 0.71 to 0.81 and $\hat{P}(R_{2i}=1 \mid R_{1i}=0, X_{1i}, X_{2i})$ lies between 0.38 and 0.51, while $\hat{P}(R_{2i}=1 \mid R_{1i}, X_{1i}, X_{2i}=0)$ and $\hat{P}(R_{2i}=1 \mid R_{1i}, X_{1i}, X_{2i}=1)$ differs by no more than 0.003. Hence, Model 1 seems to imply that the behaviour in the 1985 election influences the response behaviour in the 1989 election (when we have controlled for 1985 response/nonresponse) more than the voting behaviour in the 1989 election. This further indicates that Model 1 is unsuitable.

One important aspect when comparing $\hat{P}_{ML}$ under the different models, is that the subpopulation of new voters is not sampled in the panel survey. Since the voting participation among young voters is

[1] At the end of Section 2.2 we described how SE of parameter estimates were computed using parametric bootstrapping by simulating 1000 sets of data. Each estimation based on the simulated data gives estimates of the various conditional probabilities. These are used to give the estimates of the SE's.

smaller than the population rate, we cannot and should not expect $\hat{P}_{ML}$ to adjust fully for the bias in the sample. It seems that $\hat{P}_{ML}$ under Model 2 does as well as could be expected.

Model 1 may seem at first glance quite intuitive, assuming that the response behaviour in 1985 is independent of the voting behaviour four years later. Note, however, that an equivalent fomulation is that X_{2i} does not depend on R_{1i}, given X_{1i}. We have shown that this is not a reasonable assumption. Model 2 assumes instead that R_{2i} does not depend on X_{1i}, given X_{2i} and R_{1i}. Our evaluation shows that this is a much more reasonable assumption. So clearly, we must include the combined voting behaviour for (1985, 1989) when modelling the response behaviour in 1985, while this does not seem necessary for the response behaviour in 1989.

If we disregard the knowledge of p_1 and assume it to be unknown, then only Model 3 is estimable. It turns out that the estimated rate of participation is about 0.91 both years. Clearly, this model does not work. We note that for the traditional nonresponse-adjustment methods from Section 4, the estimator $\hat{P}^* = n^*_{\circ 1}/n^*$ gives values between 0.913 and 0.922 and $p^*_1 = n^*_{1\circ}/n^*$ gives values between 0.911 and 0.915. There is simply not enough information in the data to correct for the nonresponse bias when estimating p_1. Evidently, when p_1 is unknown one needs auxiliary information known for the total sample s for poststratification purposes. To get rid of the nonresponse bias completely one should also include callback information, if available. In the next case p_1 is unknown, but we shall still be able to estimate it because of the special nature of the data.

6. The consumer expenditure panel survey

In this example we estimate the proportion of car ownership in Norwegian households in 1989 and 1990 with panel data from the Norwegian Consumer Expenditure Survey. The units are now households and $X_{1i} = 1$ if household i owns a car in 1989, and similarly for X_{2i} in 1990. In this case, $p_1 = P(X_{1i} = 1)$ is unknown and we estimate the proportion of ownership in both years. The data is given in table 6.1.

Table 6.1. Panel data for car ownership

1989\1990	$X_2 = 1$	$X_2 = 0$	*mis*	totals
$X_1 = 1$	133	1	62	196
$X_1 = 0$	3	30	16	49
mis	28	10	142	180
totals	164	41	220	425

6.1. Traditional methods

We consider the traditional mean imputation and weighting approaches from Section 4, with $\hat{P}^* = n^*_{\circ 1}/n^*$ and $p^*_1 = n^*_{1\circ}/n^*$, and the proportions of ownership in the response sample s_{rr}, $\hat{p}_{1,rr}$ and $\hat{P}_{rr}$.

Table 6.2. Traditional estimates of proportions of car ownership

Method	$\hat{p}_1^*$	$\hat{P}^*$
Mean imputation	0.791	0.802
W1	0.770	0.780
W2	0.792	0.803
s_{rr} *-based*	$\hat{p}_{1,rr}$ *= 0.802*	$\hat{P}_{rr}$ *= 0.814*

The marginal rates of ownership in 1989 and 1990, 196/245 and 164/205, gives 0.80 in both years. Compared to W1-method it seems that there might be a small bias due to nonresponse. In the next section we shall look at the model-based maximum likelihood estimators for p_1 and P for two different models as well as studying closer the weighting procedures.

6.2. Maximum likelihood estimation

Looking at the three models in Section 5, we see that letting p_1 be an unknown parameter, we can only use Model 3 since the maximal number of identifiable parameters is 8. In order to separate this new model, with p_1 to be estimated, from the earlier ones, it is called Model 3*. We shall also consider the following model:

Model 4 $\phi_2^{(2)} = \phi_3^{(2)} = 0$, and unkown p_1

In this model we have that, conditional on R_{1i}, R_{2i} is independent of $\left(X_{1i}, X_{2i}\right)$.

The likelihood function is of the same form as in Section 2.2. The ML estimates of p_{11} and p_{01} for Models 3* and 4 are as follows.

Parameter	Model 3*	Model 4
p_{11}	0.9924	0.9925
p_{01}	0.0896	0.0909

Based on s_{rr}, the estimates of p_{11} and p_{01} are given by:

$$\hat{p}_{11}^{(r)} = 133/134 = 0.9925 \quad \text{and} \quad \hat{p}_{01}^{(r)} = 3/33 = 0.0909 .$$

Comparing these estimates to the ML estimates there seems to be no bias due to nonresponse regarding *change* in ownership category. In Section 3 we presented three model-based estimators of P when p_1 is known. Using ML estimate $\hat{p}_1$ in place of p_1 we obtain three modified estimators. E.g., $\hat{P}_{ML} = \hat{p}_1\hat{p}_{11} + (1-\hat{p}_1)\,\hat{p}_{01}$. $\hat{P}_I^{(c)}$ and $\hat{P}_{ML}$ differ by less than 0.001 (the total number of households is approximately $1.9 \cdot 10^6$). We have nearly perfect fit to the data, so that $\hat{P}_I$ and $\hat{P}_{ML}$ are approximately equal here. Hence, only $\hat{P}_{ML}$ is given in the following analysis. The estimates of p_1 and P turn out to be, for the two models

	Model 3*	Model 4
$\hat{p}_1$	0.761	0.765
$\hat{P}_{ML}$	0.777	0.780

The SE's are about 0.02 for $\hat{p}_1$ and 0.05 for $\hat{P}_{ML}$. Under the model (3.4) of ignorable RM, the ML estimates of p_{11}, p_{01} and p_1 are 0.9918, 0.0834 and 0.791 respectively, with $\hat{P}_{ML}$ = 0.802. Hence, a nonignorable RM-model seems reasonable. We see that W1 produces estimates of p_1 and P that are very close to the model-based ML estimates. The other standard methods gives results that are practically identical to the ML estimates under ignorable RM. We shall now try to find the reasons for these results.

Mean imputation assume implicitly the ignorable model (3.4), $\phi_1^{(1)} = \phi_2^{(1)} = \phi_2^{(2)} = \phi_3^{(2)} = 0$, for RM. So it is reasonable that the estimates from this approach are similar to the ML estimates under the same model. W1 and W2 can adjust for certain nonresponse biases. If we look carefully at W1 we see that this weighting scheme implicitly requires the following independent structure:

(a) X_{2i} and R_{2i} are independent, conditional on X_{1i} and $R_{1i} = 1$

(b) X_{2i} and R_{2i} are independent, conditional on $R_{1i} = 0$

(c) X_{1i} and R_{1i} are independent, conditional on X_{2i} and $R_{2i} = 1$.

In terms of the model (2.1) - (2.3),

(a) $\Leftrightarrow$ $\phi_3^{(2)} = 0$

(b) $\Leftrightarrow$ $\phi_2^{(2)} = 0$, when $\phi_3^{(2)} = 0$

(c) $\Leftrightarrow$ $\phi_1^{(1)} = 0$, when $\phi_2^{(2)} = \phi_3^{(2)} = 0$.

Hence, W1 assumes that $\phi_1^{(1)} = \phi_2^{(2)} = \phi_3^{(2)} = 0$, but do allow for $\phi_2^{(1)} \neq 0$, i.e., R_{1i} and X_{2i} can be dependent. W1 will therefore account for different X_2-distributions in the strata $R_{1i} = 0$ and $R_{1i} = 1$. In this particular case, the proportions of ($X_{2i} = 1$) are 28/38 = 0.737 and 136/167 = 0.814 for these strata. W1 will therefore lead to lower p_1- and P-estimates than the methods based on ignorable RM.

The scheme W2 requires similarily $\phi_1^{(1)} = \phi_2^{(1)} = \phi_3^{(2)} = 0$ but allows for $\phi_2^{(2)} \neq 0$, i.e., W2 will account for different X_1-distributions in the strata $R_{2i} = 0$ and $R_{2i} = 1$. For these panel data, however, we have no nonresponse bias here. The proportions of ($X_{1i} = 1$) are 62/78 = 0.795 and 134/167 = 0.802 in these strata, explaining why W2 performs similar to the methods based on ignorable RM.

6.3. Model comparisons

In this case we have no true values to compare with, but we can judge if the various estimates are plausible or not. We see that Model 3*, in estimating ownership for the whole population, reduces the respondent ownership percentage by 3.9 the first year and 2.3 the second year. It seems likely that the true percentages are less than the percentages among respondents. This is supported by the estimates provided by the weighting scheme W1. The estimated response probabilities the second year, under Model 3*, are (with SE in parentheses):

$$\hat{P}(R_{2i} = 1 \mid R_{1i} = 1, X_{2i} = 1) \quad = \quad 0.684 \ \ (0.035)$$

$$\hat{P}(R_{2i} = 1 \mid R_{1i} = 1, X_{2i} = 0) \quad = \quad 0.672 \ \ (0.089)$$

$$\hat{P}(R_{2i} = 1 \mid R_{1i} = 0, X_{2i} = 1) \quad = \quad 0.213 \ \ (0.045)$$

$$\hat{P}(R_{2i} = 1 \mid R_{1i} = 0, X_{2i} = 0) \quad = \quad 0.205 \ \ (0.035).$$

Note that the response behaviour the first year strongly influences the response probability the second year, while state of ownership has little effect indicating that $\phi_3^{(2)} \approx 0$. The best weighting scheme W1 allowed for $\phi_2^{(1)} \neq 0$. Model 4 is therefore an alternative to Model 3*, letting $\phi_2^{(2)} = \phi_3^{(2)} = 0$, with no assumption about $\phi_2^{(1)}$. We have three different RM-nonignorable models producing very similar estimates of p_1 and P. These estimates differs from the estimates based on the RM-ignorable model, indicating that there is some bias due to nonresponse. We can compute estimates of the conditional probabilities that a household owns a car, given response and nonresponse. Model 3* gives:

$$\begin{aligned}
\hat{P}\left(X_{1i}=1 \mid R_{1i}=1\right) &= 0.800 \\
\hat{P}\left(X_{1i}=1 \mid R_{1i}=0\right) &= 0.708 \\
\hat{P}\left(X_{2i}=1 \mid R_{2i}=1\right) &= 0.800 \\
\hat{P}\left(X_{2i}=1 \mid R_{2i}=0\right) &= 0.755
\end{aligned}$$

We observe that the model reproduces the observed marginals, and estimates the ownership rates in the nonresponse groups to be significantly less. Note also that the probability of owning a car increases in the subpopulation of nonrespondents. $\hat{P}_{ML}$-percentage seems to be about 1-1.5 higher than $\hat{p}_1$-percentage. This could be a trend, though it is probably not. More likely, it is a panel effect. The persons in the household are one year older the second year and the probability of owning a car is likely to increase with age.

As mentioned in Section 6, using Model 3* for the election panel survey data leads to estimated rates of participation of around 0.91 both years. Evidently, Model 3* does not work in this case. One important difference in two cases is that the last panel involves a nearly absorbing state, ownership of cars, whereas the election panel lacks a state with this feature. Obviously, a nearly absorbing state gives more information about the conditional probabilities involved. This is probably the reason for the seemingly better results with Model 3* in the case of car ownership.

7. Conclusions

We have considered a model-driven approach to panel surveys with nonresponse present as an alternative to methods of weighting and direct data imputation that are currently in use. For the two illustrations it is found that, on the whole, the traditional methods are inferior to model-based procedures. This is not surprising since the traditional methods implicitly assume that the response mechanism is essentially ignorable which is rarely the case.

Various models have been evaluated, especially in the election panel survey. Among the alternatives considered in this case, a clear "winner" is the model assuming independence between voting behaviour in the first election and response behaviour in the second election, when we have controlled for response in the first and voting in the second election.

References

Bjørnstad, J.F. (1996): On the generalization of the likelihood function and the likelihood principle, *Journal of the American Statistical Association* **91**, 791-806.

Bjørnstad, J.F. and Walsøe, H.K. (1991): Predictive likelihood in nonresponse problems, *American Statistical Association 1991 Proceedings of the Section on Survey Research Methods*, 152-156.

Conaway, M.R. (1993): Non-ignorable non-response models for timeordered categorical variables, *Applied Statistics* **42**, 105-115.

Efron, B. and Tibshirani, R.J. (1993): *An Introduction to Bootstrap*, Chapman & Hall.

Fay, R.E. (1986): Casual models for patterns of nonresponse. *Journal of the American Statistical Association* **81**, 354-365.

Fay, R.E. (1989): "Estimated nonignorable nonresponse in longitudinal surveys through causal modeling", in Kasprzyk, D., Duncan, G., Kalton, G. And Singh, M.P. (eds.): *Panel Surveys* , Academic Press, 375-399.

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

Kalton, G. (1986): Handling wave nonresponse in panel surveys, *Journal of Official Statistics* **2**, 303-314.

Lepkowski, J.M. (1989): "Treatment of wave nonresponse in panel surveys", in Kasprzyk, D., Duncan, G., Kalton, G. And Singh, M.P. (eds.): *Panel Surveys*, Academic Press, 348-374.

NAG Fortran Library Manual Mark 11 **3**.

Stasny, E. (1987): Some Markov chain models for nonresponse in estimating gross labor force flows, *Journal of Official Statistics* **3**, 359-373.

Thomsen, I. (1981): The use of Markov chain models in sampling from finite populations, *Scandinavian Journal of Statistics* **8**, 1-9.

Thomsen, I. and Siring, E. (1983): On the causes and effects of non-response. Norwegian Experiences, *Incomplete Data in Sample Surveys* Vol.3, Academic Press.

Appendix

Lemma 1. $|\hat{P}_I^{(c)} - \hat{P}_{ML}| \le \frac{n}{N}$.

Proof. Let $A = \sum_{s_{rr}} X_{2i} + \sum_{s_{mr}} X_{2i} + \sum_{s_{rm}} X_{2i}^{*} + \sum_{s_{mm}} X_{2i}^{*}$. Since

$$\hat{t}_I^{(c)} = A + (N-n)\left(p_1 \hat{p}_{11} + (1-p_1)\right)\hat{p}_{01} \quad \text{and} \quad \hat{P}_I^{(c)} = \hat{t}_I^{(c)} / N$$

we get:
$$\hat{P}_I^{(c)} = \frac{1}{N} A + \frac{N-n}{N} \hat{P}_{ML}$$

Rearranging:
$$\left|\hat{P}_I^{(c)} - \hat{P}_{ML}\right| = \frac{n}{N}\left|\frac{1}{n} A - \hat{P}_{ML}\right| \le \frac{n}{N}$$ ♣

Lemma 2. *Assume that the fit of the data to the model is perfect. Then* $\hat{P}_{ML} = \hat{P}_I$.

Proof. For convenience we introduce the following notation:

$$P\left(X_{1i} = a, X_{2i} = b, R_{1i} = c, R_{2i} = d\right) = P(a,b,c,d)$$

$$\text{and} \quad P(a,b,c,-) = P(a,b,c,0) + P(a,b,c,1) \quad \text{etc.}$$

Let $\hat{P}(a,b,c,d)$ be the estimated $P(a,b,c,d)$. Similarly for $\hat{P}(a,b,c,-),..$ etc. Since $\hat{P}_{ML}$ and $\hat{P}_I$ are different only in the way the transition probabilities p_{11} and p_{01} are estimated, it is sufficient to show that the estimates of p_{11} and p_{01} are equal. Due to symmetry it is enough to show that $\hat{p}_{11,I}^{(c)} = \hat{p}_{11}$. We have

$$\hat{p}_{11} = \frac{\hat{P}(1,1,-,-)}{\hat{P}(1,-,-,-)}$$

Furthermore, using the imputed values from Section 3,

$$\hat{p}_{11,I}^{(c)} = \frac{A}{B}$$

where

$$A = n_{11} + n_{13} \frac{\hat{P}(1,1,1,0)}{\hat{P}(1,-,1,0)} + n_{31} \frac{\hat{P}(1,1,0,1)}{\hat{P}(-,1,0,1)} + n_{33} \frac{\hat{P}(1,1,0,0)}{\hat{P}(-,-,0,0)}$$

$$B = (n_{11} + n_{12}) + n_{13} + n_{31} \frac{\hat{P}(1,1,0,1)}{\hat{P}(-,1,0,1)} + n_{32} \frac{\hat{P}(1,0,0,1)}{\hat{P}(-,0,0,1)} + n_{33} \frac{\hat{P}(1,-,0,0)}{\hat{P}(-,-,0,0)}$$

Since the fit is perfect we have:

$$n\hat{P}(1,1,1,1) = n_{11} \qquad n\hat{P}(1,0,1,1) = n_{12} \qquad n\hat{P}(1,-,1,0) = n_{13}$$

$$n\hat{P}(0,1,1,1) = n_{21} \qquad n\hat{P}(0,0,1,1) = n_{22} \qquad n\hat{P}(0,-,1,0) = n_{23}$$

$$n\hat{P}(-,1,0,1) = n_{31} \qquad n\hat{P}(-,0,0,1) = n_{32} \qquad n\hat{P}(-,-,0,0) = n_{33}.$$

Replacing the n_{ij}'s in A and B with the corresponding $\hat{P}$'s gives us immediately that

$$\hat{p}_{11,I}^{(c)} = \frac{\hat{P}(1,1,-,-)}{\hat{P}(1,-,-,-)} = \hat{p}_{11}. \qquad ♣$$