Notater

Statistisk sentralbyrå

93/18 Mai, 1993

Nonresponse models for panel surveys

by

Jan F. Bjørnstad and Dag Einar Sommervoll

University of Trondheim and Central Bureau of Statistics of Norway

NONRESPONSE MODELS FOR PANEL SURVEYS

BY JAN F. BJØRNSTAD AND DAG EINAR SOMMERVOLL

UNIVERSITY OF TRONDHEIM AND CENTRAL BUREAU OF STATISTICS OF NORWAY

1. INTRODUCTION

The aim of this paper is to study modelling in panel surveys with nonresponse. We consider population models with a sequential conditional logistic model for the response mechanism. Other types of models for nonresponse in panel surveys are discussed by Fay (1986,1989), Stasny (1987) and Conaway (1993). Two prediction methods utilizing mean imputation and an estimative approach are considered for estimating the population total.

Two applications are considered. The first one is estimation of the population rate of participation in the 1989 Norwegian-Storting election, based on panel data from the 1985 and the 1989 elections. The second problem considered is estimation of car ownership in Norwegian households in 1989 and 1990, with panel data from the Norwegian Consumer Expenditure Survey. In latter case we estimate the proportion of ownership in both years.

2. THE ELECTION PANEL SURVEY

In sections 2 - 6, we shall study modelling in election panel surveys with nonresponse, based on the panel data from the 1985 and 1989 elections. The data can be presented as in the following table:

1985\1989	yes	no	nr	totals
yes	n_{11}	n_{12}	n_{13}	$n_{1\star}$
no	n_{21}	n_{22}	n_{23}	$n_{2\star}$
nr	n_{31}	n_{32}	n_{33}	$n_{3\star}$
totals	$n_{\star 1}$	$n_{\star 2}$	$n_{\star 3}$	n

Here nr is short for nonresponse. Moreover, n_{ij} is the number of persons belonging to the indicated category.

The individuals in the population of eligible voters, V, are labelled $1, \ldots, N$. The panel sample, s, is selected from V. Actually, s is drawn from the subpopulation of eligible voters in 1985 and 1989. Since the true population proportion of voters in 1989 is known, we have a way to evaluate various models. This comparison should hopefully give us some indication on what may be appropriate models for similar problems in the future.

Let us introduce the following random variables for each person i:

$\mathcal{X}_i = 1$	if person nr.i votes in 1985, 0 otherwise
$\mathcal{Y}_i = 1$	if person nr.i votes in 1989, 0 otherwise
$\mathcal{R}_{1i} = 1$	if person nr.i responds in 1985, 0 otherwise
$\mathcal{R}_{2i} = 1$	if person nr.i responds in 1989, 0 otherwise

The panel consists of the following groups:

$$s_{rr} = \{i : \quad i \in s \quad \text{such that} \quad \mathcal{R}_{1i} = 1, \mathcal{R}_{2i} = 1\}$$
$$s_{rm} = \{i : \quad i \in s \quad \text{such that} \quad \mathcal{R}_{1i} = 1, \mathcal{R}_{2i} = 0\}$$
$$s_{mr} = \{i : \quad i \in s \quad \text{such that} \quad \mathcal{R}_{1i} = 0, \mathcal{R}_{2i} = 1\}$$
$$s_{mm} = \{i : \quad i \in s \quad \text{such that} \quad \mathcal{R}_{1i} = 0, \mathcal{R}_{2i} = 0\}$$

The persons that respond in the survey on both occasions are:

$$\#(s_{rr}) = n_{11} + n_{12} + n_{21} + n_{22}$$

We shall estimate the voting proportion P in the population V in 1989, by making use of the known voting proportion in 1985, $p_1 = 0.838$). The rate of voting in s_{rr} is given by:

$$\hat{P}_{rr} = \frac{n_{11} + n_{21}}{n_{11} + n_{21} + n_{12} + n_{22}}$$

Usually in panel surveys, $\hat{P}_{rr}$ will overestimate the true P. In our case we have the following panel data:

1985\1989	yes	no	nr	totals
yes	743	36	188	967
no	42	20	26	88
nr	115	20	162	297
totals	900	76	376	1352

Here :

$$\hat{P}_{rr} = 0.933$$

The true value in 1989 is :

$$P = 0.832$$

It seems likely that part of the bias is due to nonresponse, and we need a model for the response mecanism. The next section presents one approach to modelling panel data.

3. A LOGISTIC MODEL FOR PANEL SURVEYS

Given i, the variables $\mathcal{X}_i, \mathcal{Y}_i, \mathcal{R}_{1i}, \mathcal{R}_{2i}$ defined are clearly dependent. We know for instance, that the probability of voting the second time depends on whether you voted the first time. Let:

$$p_{11} = P(\mathcal{Y}_i = 1|\mathcal{X}_i = 1)$$

$$p_{01} = P(\mathcal{Y}_i = 1|\mathcal{X}_i = 0)$$

So p_{11} is the conditional probability of voting on the second occasion for a person i, given that this person did vote at the first occasion. An equivalent formulation is:

$$log(\frac{P(\mathcal{Y}_i = 1|\mathcal{X}_i)}{P(\mathcal{Y}_i = 0|\mathcal{X}_i)}) = \beta_0 + \beta_1 \mathcal{X}_i \tag{1}$$

Where:

$$\beta_0 = log(\frac{p_{01}}{(1 - p_{01})})$$

$$\beta_1 = log(\frac{p_{11}(1 - p_{01})}{p_{01}(1 - p_{11})})$$

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 eliminated. In our case we expect β_1 to be a positive constant, since $\beta_1 > 0 \iff p_{11} > p_{01}$.

The model is developed through parametrizing conditional probabilities. Formally we expand the joint probability:

$$\begin{aligned}
& P(\mathcal{X}_i = x_i, \mathcal{Y}_i = y_i, \mathcal{R}_{1i} = r_{1i}, \mathcal{R}_{2i} = r_{2i}) \\
= \; & P(\mathcal{X}_i = x_i) P(\mathcal{Y}_i = y_i, \mathcal{R}_{1i} = r_{1i}, \mathcal{R}_{2i} = r_{2i}|\mathcal{X}_i = x_i) \\
= \; & P(\mathcal{X}_i = x_i) P(\mathcal{Y}_i = y_i|\mathcal{X}_i = x_i) P(\mathcal{R}_{1i} = r_{1i}, \mathcal{R}_{2i} = r_{2i}|\mathcal{X}_i = x_i, \mathcal{Y}_i = y_i) \\
= \; & P(\mathcal{X}_i = x_i) P(\mathcal{Y}_i = y_i|\mathcal{X}_i = x_i) P(\mathcal{R}_{1i} = r_{1i}|\mathcal{X}_i = x_i, \mathcal{Y}_i = y_i) \cdot \\
& P(\mathcal{R}_{2i} = r_{2i}|\mathcal{R}_{1i} = r_{1i}, \mathcal{X}_i = x_i, \mathcal{Y}_i = y_i)
\end{aligned}$$

Following the same lines of thought as above, we assume:

$$log(\frac{P(\mathcal{R}_{1i} = 1|\mathcal{X}_i, \mathcal{Y}_i)}{P(\mathcal{R}_{1i} = 0|\mathcal{X}_i, \mathcal{Y}_i)}) = \phi_0^{(1)} + \phi_1^{(1)} \mathcal{X}_i + \phi_2^{(1)} \mathcal{Y}_i \tag{2}$$

$$log(\frac{P(\mathcal{R}_{2i} = 1|\mathcal{R}_{1i}, \mathcal{X}_i, \mathcal{Y}_i)}{P(\mathcal{R}_{2i} = 0|\mathcal{R}_{1i}, \mathcal{X}_i, \mathcal{Y}_i)}) = \phi_0^{(2)} + \phi_1^{(2)} \mathcal{R}_{1i} + \phi_2^{(2)} \mathcal{X}_i + \phi_3^{(2)} \mathcal{Y}_i \tag{3}$$

In the last two equations we have not included interaction terms.

The model (1)-(3) has introduced 9 parameters, which are not all identifiable. We need to reduce the number of parameters to maximum 8. This can be done in several ways, giving rise to different models.

Model 1

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

This amounts to assuming that the probability of response the first time, does not depend on the voting behavior at the second election.

Model 2

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

In this model we keep the two first equations, and reduce the third one by setting $\phi_2^{(2)} = 0$. This means that the voting behavior in the first election does not affect the probability of response the second time.

Model 3

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

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

This is the intersection of model 1 and model 2.

4. PARAMETER ESTIMATION - THE ELECTON PANEL SURVEY

First we consider estimation of the unknown parameters. The method of maximum likelihood will be applied. The likelihood function is the probability of the data as a function of the parameters. It is given by:

$$
\begin{aligned}
L(\underline{\beta}, \underline{\phi}^{(1)}, \underline{\phi}^{(2)}) = \\
&= \prod_{i \in s_{rr}} P(\mathcal{X}_i = x_i \cap \mathcal{Y}_i = y_i \cap \mathcal{R}_{1i} = 1 \cap \mathcal{R}_{2i} = 1) \\
&\times \prod_{i \in s_{rm}} P(\mathcal{X}_i = x_i \cap \mathcal{R}_{1i} = 1 \cap \mathcal{R}_{2i} = 0) \\
&\times \prod_{i \in s_{mr}} P(\mathcal{Y}_i = y_i \cap \mathcal{R}_{1i} = 0 \cap \mathcal{R}_{2i} = 1) \\
&\times \prod_{i \in s_{mm}} P(\mathcal{R}_{1i} = 0 \cap \mathcal{R}_{2i} = 0)
\end{aligned}
$$

The final expression of the likelihood in terms of $\underline{\beta}, \underline{\phi}^{(1)}, \underline{\phi}^{(2)}$ is given in appendix A.

$log(L)$ is maximized numerically by using a NAG subroutine (E04JAF). To estimate the standard error (s.e.) of the maximum likelihood estimates $\hat{\underline{\theta}} = (\hat{\underline{\beta}}, \hat{\underline{\phi}}^{(1)}, \hat{\underline{\phi}}^{(2)})$, we use parametric bootstrapping by simulating 1000 sets of data assuming $(\underline{\beta}, \underline{\phi}^{(1)}, \underline{\phi}^{(2)}) = (\hat{\underline{\beta}}, \hat{\underline{\phi}}^{(1)}, \hat{\underline{\phi}}^{(2)})$. The estimated standard error of a given estimate, is then the empirical standard deviation for 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 standard error is then given by:

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

with

$$\bar{\hat{\beta}}_0 = \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_{0,i} \qquad \text{and} \qquad N = 1000$$

$\bar{\hat{\beta}}_0$ estimates $E(\hat{\beta}_0)$ at $\theta = \hat{\theta}$. From a simulation study, it seems that the m.l. estimates are approximately unbiased. The m.l. estimates and the corresponding estimated s.e. (in parenthesis) are given in the table below.

TABLE I

	$M1$	$M2$	$M3$
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)
$\phi_0^{(1)}$	−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 ()	0.443 (0.475)	0 ()

	$M1$	$M2$	$M3$
$\phi_0^{(2)}$	−0.445	−1.21	−1.01
	(2.264)	(1.03)	(0.357)
$\phi_1^{(2)}$	1.369	1.36	1.45
	(0.188)	(0.197)	(0.149)
$\phi_2^{(2)}$	0.574	0	0
	(0.512)	()	()
$\phi_3^{(2)}$	−0.080	1.40	1.05
	(2.495)	(1.17)	(0.446)
p_{11}	0.954	0.926	0.937
	(0.021)	(0.027)	(0.014)
p_{01}	0.678	0.5125	0.572
	(0.104)	(0.092)	(0.068)

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

$$\hat{p}_{11}^{(r)} = \frac{743}{779} = 0.954 \qquad \text{and} \qquad \hat{p}_{01}^{(r)} = \frac{42}{62} = 0.677$$

5. ESTIMATION OF VOTING PARTICIPATION

We shall consider three methods, the estimative approach and two imputation estimators.

5.1 Estimative approach

We see that

$$E(\mathcal{Y}_i) = P(\mathcal{Y}_i = 1) = p_1 p_{11} + (1 - p_1) p_{01}$$

This implies that, with $P = \frac{1}{N} \sum_{i=1}^{N} \mathcal{Y}_i$,

$$E(P) = \frac{1}{N} \sum_{i=1}^{N} E(\mathcal{Y}_i) = P(\mathcal{Y}_i = 1) = p_1 p_{11} + (1 - p_1) p_{01}$$

Here, $p_1 = 0.838$.

One simple estimator is therefore given by: $\hat{P}_e = p_1 \hat{p}_{11} + (1 - p_1) \hat{p}_{01}$.

5.2 Imputation

Our method of imputation is natural under a population model. Others that have used this method include Greenlees et al. (1982) and Bjørnstad & Walsøe (1991).

Let $t = \sum_{i=1}^{N} \mathcal{Y}_i$, and $P = t/N$.

A general and common method for estimating t, when we have nonresponse, is by imputation of the missing values in s.

Now, (with $\bar{s} = \{i : \quad i \notin s\}$):

$$t = \sum_{s_{rr}} \mathcal{Y}_i + \sum_{s_{mr}} \mathcal{Y}_i + \sum_{s_{rm}} \mathcal{Y}_i + \sum_{s_{mm}} \mathcal{Y}_i + \sum_{\bar{s}} \mathcal{Y}_i$$

$z = \sum_{\bar{s}} \mathcal{Y}_i$ is estimated by estimating

$$E(\sum_{\bar{s}} \mathcal{Y}_i) = (N-n)P(\mathcal{Y}_i = 1) = (N-n)(p_1 p_{11} + (1-p_1)p_{01})$$

Giving:

$$\hat{z} = (N-n)\hat{P}(\mathcal{Y}_i = 1) = (N-n)(p_1 \hat{p}_{11} + (1-p_1)\hat{p}_{01})$$

For $i \in s - (s_{rr} \cup s_{mr})$, the imputed values of $\mathcal{Y}_i$ are:

$$i \in s_{rm} \qquad \mathcal{Y}_i^\star = \quad \text{estimated} \quad P(\mathcal{Y}_i = 1 | \mathcal{X}_i, \mathcal{R}_{1i} = 1, \mathcal{R}_{2i} = 0)$$

and

$$i \in s_{mm} \qquad \mathcal{Y}_i^\star = \quad \text{estimated} \quad P(\mathcal{Y}_i = 1 | \mathcal{R}_{1i} = 0, \mathcal{R}_{2i} = 0)$$

The imputation estimator of t is then:

$$\hat{t}_I = \sum_{s_{rr}} \mathcal{Y}_i + \sum_{s_{mr}} \mathcal{Y}_i + \sum_{s_{rm}} \mathcal{Y}_i^\star + \sum_{s_{mm}} \mathcal{Y}_i^\star + (N-n)(p_1 \hat{p}_{11} + (1-p_1)\hat{p}_{01})$$

and the corresponding

$$\hat{P}_I = \frac{\hat{t}_I}{N}$$

$\hat{P}_I$ and $\hat{P}_e$ will give approximately the same results. In fact we always have the bound, $|\hat{P}_I - \hat{P}_e| \leq \frac{n}{N}$ (see the appendix). In our case the maximal difference is less than 10^{-3}.

There is a third estimator, that utilizes the information present in the panel explicitly. It is constructed the following way:

Assume that we have no nonresponse, i.e. $s_{rr} = s$. Then, from Thomsen (1981),the optimal estimator is given by :

$$\check{P} = p_1 \check{p}_{11} + (1-p_1)\check{p}_{01}$$

where

$$\check{p}_{11} = \frac{\sum_s \mathcal{X}_i \mathcal{Y}_i}{\sum_s \mathcal{X}_i}$$

and

$$\check{p}_{01} = \frac{\sum_s (1-\mathcal{X}_i)\mathcal{Y}_i}{\sum_s (1-\mathcal{X}_i)}.$$

With nonresponse we impute the unkown values in $\check{p}_{11}$ and $\check{p}_{01}$. More precisely: Impute the following values:

$$\begin{aligned}
i \in s_{rm}: &\quad \mathcal{Y}_i^* = \text{ estimated } \quad P(\mathcal{Y}_i = 1|\mathcal{X}_i, \mathcal{R}_{1i} = 0, \mathcal{R}_{2i} = 1) \\
i \in s_{mr}: &\quad \mathcal{X}_i^* = \text{ estimated } \quad P(\mathcal{X}_i = 1|\mathcal{Y}_i, \mathcal{R}_{1i} = 1, \mathcal{R}_{2i} = 0) \\
i \in s_{mm}: &\quad \mathcal{X}_i^* = \text{ estimated } \quad P(\mathcal{X}_i = 1|\mathcal{R}_{1i} = 0, \mathcal{R}_{2i} = 0) \\
i \in s_{mm}: &\quad (\mathcal{X}_i\mathcal{Y}_i)^* = \text{ estimated } \quad P(\mathcal{X}_i = 1, \mathcal{Y}_i = 1|\mathcal{R}_{1i} = 0, \mathcal{R}_{2i} = 0)
\end{aligned}$$

Let $\hat{P}(\cdot|\cdot)$ indicate the estimated $P(\cdot|\cdot)$. The imputation estimator based on $\check{P}$ is then

$$\check{P}_I = p_1\check{p}_{11,I} + (1-p_1)\check{p}_{01,I}$$

Where

$$\check{p}_{11,I} = \frac{A}{B}$$

with

$$\begin{aligned}
A = & n_{11} + \\
& n_{13}\hat{P}(\mathcal{Y}_i = 1|\mathcal{X}_i = 1, \mathcal{R}_{1i} = 1, \mathcal{R}_{2i} = 0) + \\
& n_{31}\hat{P}(\mathcal{X}_i = 1|\mathcal{Y}_i = 1, \mathcal{R}_{1i} = 0, \mathcal{R}_{2i} = 1) + \\
& n_{33}\hat{P}(\mathcal{X}_i = 1, \mathcal{Y}_i = 1|\mathcal{R}_{1i} = 0, \mathcal{R}_{2i} = 0)
\end{aligned}$$

and

$$\begin{aligned}
B = & (n_{11} + n_{12}) + n_{13} + \\
& n_{31}\hat{P}(\mathcal{X}_i = 1|\mathcal{Y}_i = 1, \mathcal{R}_{1i} = 0, \mathcal{R}_{2i} = 1) + \\
& n_{32}\hat{P}(\mathcal{X}_i = 1|\mathcal{Y}_i = 0, \mathcal{R}_{1i} = 0, \mathcal{R}_{2i} = 1) + \\
& n_{33}\hat{P}(\mathcal{X}_i = 1|\mathcal{R}_{1i} = 0, \mathcal{R}_{2i} = 0)
\end{aligned}$$

and

$$\check{p}_{01,I} = \frac{C}{D}$$

where

$$D = n - B$$

and

$$\begin{aligned}
C = n_{\star 1} & + n_{13}\hat{P}(\mathcal{Y}_i = 1|\mathcal{X}_i = 1, \mathcal{R}_{1i} = 1, \mathcal{R}_{2i} = 0) \\
& + n_{23}\hat{P}(\mathcal{Y}_i = 1|\mathcal{X}_i = 0, \mathcal{R}_{1i} = 1, \mathcal{R}_{2i} = 0) \\
& + n_{33}\hat{P}(\mathcal{Y}_i = 1|\mathcal{R}_{1i} = 0, \mathcal{R}_{2i} = 0) - A
\end{aligned}$$

The strength of this estimator is that we only impute for the missing values. The first estimator uses only the estimated parameters to estimate P and implicitly also impute for observed values.

Given a set of estimated parameters one can construct a panel A, by taking the expectation values of the cells. If panel A equals the original panel, the fit is perfect, and we say that the parameters reproduce the panel. When the parameters reproduce the panel, then estimator 1 and estimator 3 are exactly equal. This is established in the appendix.

In models 1 and 2 we reproduce the panel, and model 3 nearly reproduces the panel. As a consequence the three estimators $\hat{P}_e, \hat{P}_I$ and $\check{P}_I$ will give estimates that are approximately equal. Only the value of $\hat{P}_e$ is given below, for the different models. The estimator $\check{P}$ computed for s_{rr} is denoted by $\tilde{P}$.

5.3 Estimates (P=0.832)

In parentheses, the estimated s.e. are given.

	$M1$	$M2$	$M3$
$\hat{P}_e$	0.911	0.858	0.880
	(0.034)	(0.034)	(0.019)

$$\tilde{P}_e = 0.909$$

6. DISCUSSION - THE ELECTION PANEL SURVEY

Comparing $\hat{P}_e$ to $\tilde{P}$, we see that model 1 does not work. It does not correct for the bias due to nonresponse, that we know is present. Let us consider the modeling aspects for the distribution of $\mathcal{R}_{1i}$ given $\mathcal{X}_i$ and $\mathcal{Y}_i$. Using the m.l. estimates from Table I, we find the following estimates (s.e):

	$M1$		$M2$		$M3$	
$\hat{P}(\mathcal{R}_{1i} = 1 \mid \mathcal{X}_i = 1, \mathcal{Y}_i = 1)$	0.854	(0.015)	0.858	(0.015)	0.856	(0.015)
$\hat{P}(\mathcal{R}_{1i} = 1 \mid \mathcal{X}_i = 1, \mathcal{Y}_i = 0)$	0.854	(0.015)	0.795	(0.071)	0.856	(0.015)
$\hat{P}(\mathcal{R}_{1i} = 1 \mid \mathcal{X}_i = 0, \mathcal{Y}_i = 1)$	0.402	(0.041)	0.453	(0.080)	0.396	(0.041)
$\hat{P}(\mathcal{R}_{1i} = 1 \mid \mathcal{X}_i = 0, \mathcal{Y}_i = 0)$	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 to model 2 indicates that $\mathcal{R}_{1i}$ may depend slightly on $\mathcal{Y}_i$, even when $\mathcal{X}_i$ is known. It has the effect of lowering the response probability for those who did not participate in the 1989 election.

Using the m.l. estimates, we can do the same for $\mathcal{R}_{2i}$ given $\mathcal{X}_i$, $\mathcal{Y}_i$, and $\mathcal{R}_{1i}$.

	$M1$		$M2$		$M3$	
$\hat{P}(\mathcal{R}_{2i} = 1 \vert \mathcal{R}_{1i} = 1, \mathcal{X}_i = 1, \mathcal{Y}_i = 1)$	0.806	(0.021)	0.825	(0.023)	0.817	(0.017)
$\hat{P}(\mathcal{R}_{2i} = 1 \vert \mathcal{R}_{1i} = 1, \mathcal{X}_i = 0, \mathcal{Y}_i = 1)$	0.704	(0.096)	0.825	(0.023)	0.817	(0.017)
$\hat{P}(\mathcal{R}_{2i} = 1 \vert \mathcal{R}_{1i} = 1, \mathcal{X}_i = 1, \mathcal{Y}_i = 0)$	0.807	(0.176)	0.539	(0.115)	0.603	(0.084)
$\hat{P}(\mathcal{R}_{2i} = 1 \vert \mathcal{R}_{1i} = 1, \mathcal{X}_i = 0, \mathcal{Y}_i = 0)$	0.706	(0.131)	0.539	(0.115)	0.603	(0.084)
$\hat{P}(\mathcal{R}_{2i} = 1 \vert \mathcal{R}_{1i} = 0, \mathcal{X}_i = 1, \mathcal{Y}_i = 1)$	0.514	(0.043)	0.547	(0.060)	0.553	(0.042)
$\hat{P}(\mathcal{R}_{2i} = 1 \vert \mathcal{R}_{1i} = 0, \mathcal{X}_i = 0, \mathcal{Y}_i = 1)$	0.378	(0.123)	0.547	(0.060)	0.553	(0.042)
$\hat{P}(\mathcal{R}_{2i} = 1 \vert \mathcal{R}_{1i} = 0, \mathcal{X}_i = 1, \mathcal{Y}_i = 0)$	0.517	(0.238)	0.230	(0.095)	0.264	(0.072)
$\hat{P}(\mathcal{R}_{2i} = 1 \vert \mathcal{R}_{1i} = 0, \mathcal{X}_i = 0, \mathcal{Y}_i = 0)$	0.380	(0.184)	0.230	(0.095)	0.264	(0.072)

From model 1 it seems clear that the behaviour in the 1985 election influences the response behaviour in 1989 election (when we have controlled for 1985 response/nonresponse) more than the voting behaviour in the 1989 elections. This is rather surprising.

One important aspect when comparing $\hat{P}_e$ under different models, is that the subpopulation of new voters is not sampled in the panel survey. It is well known that the 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)).

Hence, we cannot expect $\hat{P}_e$ to adjust fully for the bias in the sample.

It seems that $\hat{P}_e$ under model 2 does as well as could be expected.

This can be amplified by looking at the estimated voting rates in the subpopulations of respondents and nonrespondents for the two elections. We find:

	$M1$	$M2$	$M3$
$\hat{P}(\mathcal{X}_i = 1 \vert \mathcal{R}_{1i} = 1)$	0.917	0.917	0.918
$\hat{P}(\mathcal{X}_i = 1 \vert \mathcal{R}_{1i} = 0)$	0.568	0.568	0.561
$\hat{P}(\mathcal{Y}_i = 1 \vert \mathcal{R}_{2i} = 1)$	0.922	0.922	0.922
$\hat{P}(\mathcal{Y}_i = 1 \vert \mathcal{R}_{2i} = 0)$	0.882	0.695	0.770

(Note that the rate of voting of the nonrespondents seems to increase with time. This is what one would expect since the persons in the survey are four years older the second time.)

The only substantial diffence in the estimates lies in the estimates of $P(\mathcal{Y}_i = 1 \vert \mathcal{R}_{2i} = 0)$. It seems clear that model 3 and especially model 1 overestimate the voting participation in the 89 election among nonrespondents. This can be seen by comparing $\hat{P}(\mathcal{Y}_i = 1|\mathcal{R}_{2i} = 0)$ to $\hat{P}(\mathcal{X}_i = 1|\mathcal{R}_{1i} = 0)$.

It is rather surprising that model 2 is seemingly much more appropriate than model 1.

Model 1 seems at first glance more intuitive, since in model 2 the response behaviour in 1985 is assumed to depend on the voting behaviour four years later.

Clearly, however, we must include the combined voting behaviour for (1985,1989), when we are modelling the response behaviour in 1985. This does not seem necessary for the response behaviour in 1989.

7. THE CONSUMER EXPENDITURE PANEL SURVEY

In the last two paragraphs we are going to study modelling in panels with nonresponse from the Norwegian Consumer Expenditure Survey. The units are here households instead of persons as in the election example. We shall estimate the percentage of households that own a car. A household is said to own a car if at least one of the persons in the household owns a car.

As in section 2 the panel data can be represented by the following table:

1989\1990	yes	no	nr	totals
yes	n_{11}	n_{12}	n_{13}	$n_{1\star}$
no	n_{21}	n_{22}	n_{23}	$n_{2\star}$
nr	n_{31}	n_{32}	n_{33}	$n_{3\star}$
totals	$n_{\star 1}$	$n_{\star 2}$	$n_{\star 3}$	n

Here, nr is short for nonresponse and n_{ij} is the number of households belonging to the indicated category. The sample s is drawn from the subset of households that are registered as households in 1989 and 1990. Let us introduce the following random variables for each household i:

$\mathcal{X}_i = 1$ if household nr.i owns a car in 1989, 0 otherwise

$\mathcal{Y}_i = 1$ if household nr.i owns a car in 1990, 0 otherwise

$\mathcal{R}_{1i} = 1$ if household nr.i responds in 1989, 0 otherwise

$\mathcal{R}_{2i} = 1$ if household nr.i responds in 1990, 0 otherwise

The data are:

1989\1990	yes	no	nr	totals
yes	133	1	62	196
no	3	30	16	49
nr	28	10	142	180
totals	164	41	220	425

We apply the same notation for the different strata of respondents and nonrespondents as in section 2. The proportion of ownership in 1990 in s_{rr} is given by:

$$\hat{P}_{rr} = \frac{n_{11} + n_{21}}{n_{11} + n_{21} + n_{12} + n_{22}} = 0.814$$

Looking at the marginals give 0.80 both years. This seems to indicate that there might be a bias due to nonresponse. We apply the same kind of model as used for the election panel. Unlike our first example, the true percentage the first year is not known. One solution to this problem is to estimate p_1 as well. Looking at the three models M1 to M3, we see that letting p_1 be a parameter as well, gives us two 9 parameter models (M1 and M2) and one 8 parameter model (M3). As noted before the maximal number of identifiable parameter is 8. We will therefore only use model M3 in the following. In order to separate this new model, with p_1 free, from the earlier ones, we designate it by M4.

The likelihood function is of the same form, though the estimation is different since p_1 is a parameter as well.

The m.l. estimates and the corresponding estimated s.e. (in parenthesis) are given in the table below.

β_0	-2.319 (0.496)	$\phi_1^{(2)}$	2.075 (0.188)
β_1	7.19 (0.346)	$\phi_2^{(2)}$	0 ()
$\phi_0^{(1)}$	-0.0699 (0.169)	$\phi_3^{(2)}$	5.178 (2.495)
$\phi_1^{(1)}$	0.500 (0.243)		
$\phi_2^{(1)}$	0 ()	p_{11}	0.992 (< 0.001)
$\phi_0^{(2)}$	-1.356 (2.264)	p_{01}	0.090 (0.061)

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

$$\hat{p}_{11} = \frac{133}{134} = 0.993 \qquad \text{and} \qquad \hat{p}_{01} = \frac{3}{33} = 0.09$$

In section 4 we introduced three estimators, the first one was: $\hat{P}_e = \hat{p}_1\hat{p}_{11} + (1-\hat{p}_1)\hat{p}_{01}$ Note that we have indicated that in this case we estimate p_1. Estimator 2 and estimator 1 differ by less than 10^{-3}. (The number of households are approximately $1.9 \cdot 10^6$.) It turns out that the estimated parameters nearly reproduces the panel, giving that estimator 3 and estimator 1 are approximately equal.

The estimates of p_1 and P are:

$$\begin{aligned} \hat{p}_1 &= 0.761 \quad (0.021) \\ \hat{P}_e &= 0.777 \quad (0.051) \end{aligned}$$

8. DISCUSSION

In this case we have no true values to compare with. The only evaluation we can do is to judge if the various estimates are plausible or not. We see that model 4 reduces the repondent percentage by 3.9 the first year and 2.3 the second year. That the true percentages are significantly less than the percentages among respondents seem likely. The estimated parameters give rise to estimates of a number of different conditional probabilities. Consider:

$$\begin{array}{lll} \hat{P}(\mathcal{R}_{1i} = 1|\mathcal{X}_i = 1) & 0.603 & (0.053) \\ \hat{P}(\mathcal{R}_{1i} = 1|\mathcal{X}_i = 0) & 0.490 & (0.097) \end{array}$$

Using the m.l. estimates, we can do the same thing for $\mathcal{R}_{2i}$ given $\mathcal{X}_i$, $\mathcal{Y}_i$, and $\mathcal{R}_{1i}$.

$$\begin{array}{lll} \hat{P}(\mathcal{R}_{2i} = 1|\mathcal{R}_{1i} = 1, \mathcal{Y}_i = 1) & 0.684 & (0.035) \\ \hat{P}(\mathcal{R}_{2i} = 1|\mathcal{R}_{1i} = 1, \mathcal{Y}_i = 0) & 0.672 & (0.089) \\ \hat{P}(\mathcal{R}_{2i} = 1|\mathcal{R}_{1i} = 0, \mathcal{Y}_i = 1) & 0.213 & (0.045) \\ \hat{P}(\mathcal{R}_{2i} = 1|\mathcal{R}_{1i} = 0, \mathcal{Y}_i = 0) & 0.205 & (0.035) \end{array}$$

Note that the response behavior the first year strongly influences the response probability the second year, while the state of ownership has little effect. $\hat{P}_e$ is 1.6% units higher than

$\hat{p}_1$. 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. The probability of owning a car is likely to increase with age.

We can compute estimates of the conditional probabilities of belonging to a household that owns a car, given response and nonresponse:

	$M4$
$\hat{P}(\mathcal{X}_i = 1 \mid \mathcal{R}_{1i} = 1)$	0.800
$\hat{P}(\mathcal{X}_i = 1 \mid \mathcal{R}_{1i} = 0)$	0.708
$\hat{P}(\mathcal{Y}_i = 1 \mid \mathcal{R}_{2i} = 1)$	0.800
$\hat{P}(\mathcal{Y}_i = 1 \mid \mathcal{R}_{2i} = 0)$	0.755

We observe that the model reproduces the observed marginals, and estimate the unobserved probabilities to be significantly less. Note also that the probability of owning a car increases in the subpopulation of nonrespondents.

We could use M4 in the election panel survey data. The estimated rate of participation is then around 0.91 both years. Evidently M4 does not work in this case. One important difference in the two panels is that the last panel involves an approximately absorbing state, ownership of cars, whereas the election panel lack a state with this feature. Obviously, a nearly absorbing state gives more information about the different conditional probabilities involved. This is probably the reason for the better results with M4 in the case of car ownership.

9. ACKNOWLEDGMENTS

The authors would like to thank Ib Thomsen for numerous discussions. Helpful suggestions by Jørgen Aasness for sections 7 and 8 are gratefully acknowledged.

Appendix

LEMMA 1.

$$|\hat{P}_I - \hat{P}_e| \leq \frac{n}{N}$$

PROOF: Let $A = \sum_{s_{rr}} \mathcal{Y}_i + \sum_{s_{mr}} \mathcal{Y}_i + \sum_{s_{rm}} \mathcal{Y}_i^\star + \sum_{s_{mm}} \mathcal{Y}_i^\star$.
Since

$$\hat{t}_I = A + (N-n)(p_1 \hat{p}_{11} + (1-p_1)\hat{p}_{01}$$

and

$$\hat{P}_I = \frac{\hat{t}_I}{N}$$

we get :

$$\hat{P}_I = \frac{1}{N}A + \frac{N-n}{N}\hat{P}_e$$

Rearranging

$$|\hat{P}_I - \hat{P}_e| = \frac{n}{N}|\frac{1}{n}A - \hat{P}_e| \leq \frac{n}{N}$$

♣

LEMMA 2. *Assume that the parameters estimated reproduce the panel. Then estimator 1 is equal to estimator 3.*

PROOF: For convenience we introduce the following notation:

$$P(\mathcal{X}_i = a, \mathcal{Y}_i = b, \mathcal{R}_{1i} = c, \mathcal{R}_{2i} = d) = P(a,b,c,d)$$

and

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

Let $\hat{P}(a,b,c,d)$ be the estimated $P(a,b,c,d)$. Similar for $P(a,b,c,-)$,.. etc.

Since estimator 1 and estimator 3 are different only in the way the transition probabilities p_{11} and p_{01} are estimated, it would be sufficient to show that they are estimated equal. Due to symmetry it is enough to show $\check{p}_{11,I} = \hat{p}_{11}$. In our notation:

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

Furthermore:

$$\check{p}_{11,I} = \frac{A}{B}$$

where

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

and

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

Since the parameters reproduce the panel, we have:

$$\begin{array}{lll} n\hat{P}(1,1,1,1) = n_{11} & n\hat{P}(1,0,1,1) = n_{12} & n\hat{P}(1,-,1,0) = n_{13} \\ n\hat{P}(0,1,1,1) = n_{21} & n\hat{P}(0,0,1,1) = n_{22} & n\hat{P}(0,-,1,0) = n_{23} \\ n\hat{P}(-,1,0,1) = n_{31} & n\hat{P}(-,0,0,1) = n_{32} & n\hat{P}(-,-,0,0) = n_{33} \end{array}$$

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

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

♣

The logarithm of the likelihood function is given on the following page:

$$log(L) =$$

$$n_{11} * log(p_1 * \frac{exp(\beta_1+\beta_2)}{(1+exp(\beta_1+\beta_2))} * \frac{exp(\phi_1^{(1)}+\phi_2^{(1)})}{(1+exp(\phi_1^{(1)}+\phi_2^{(1)}))} * \frac{exp(\phi_1^{(2)}+\phi_2^{(2)}+\phi_3^{(2)}+\phi_4^{(2)})}{(1+exp(\phi_1^{(2)}+\phi_2^{(2)}+\phi_3^{(2)}+\phi_4^{(2)}))})$$

$$+n_{12} * log(p_1 * \frac{1}{(1+exp(\beta_1+\beta_2))} * \frac{exp(\phi_1^{(1)}+\phi_2^{(1)})}{(1+exp(\phi_1^{(1)}+\phi_2^{(1)}))} * \frac{exp(\phi_1^{(2)}+\phi_2^{(2)}+\phi_3^{(2)})}{(1+exp(\phi_1^{(2)}+\phi_2^{(2)}+\phi_3^{(2)}))})$$

$$+n_{21} * log((1-p_1) * \frac{exp(\beta_1)}{(1+exp(\beta_1))} * \frac{exp(\phi_1^{(1)})}{(1+exp(\phi_1^{(1)}))} * \frac{exp(\phi_1^{(2)}+\phi_2^{(2)}+\phi_4^{(2)})}{(1+exp(\phi_1^{(2)}+\phi_2^{(2)}+\phi_4^{(2)}))})$$

$$+n_{22} * log((1-p_1) * \frac{1}{(1+exp(\beta_1))} * \frac{exp(\phi_1^{(1)})}{(1+exp(\phi_1^{(1)}))} * \frac{exp(\phi_1^{(2)}+\phi_2^{(2)})}{(1+exp(\phi_1^{(2)}+\phi_2^{(2)}))})$$

$$+n_{13} * log(p_1 * \frac{exp(\beta_1+\beta_2)}{(1+exp(\beta_1+\beta_2))} * \frac{exp(\phi_1^{(1)}+\phi_2^{(1)})}{(1+exp(\phi_1^{(1)}+\phi_2^{(1)}))} * \frac{1}{(1+exp(\phi_1^{(2)}+\phi_2^{(2)}+\phi_3^{(2)}+\phi_4^{(2)}))} +$$

$$p_1 * \frac{1}{(1+exp(\beta_1+\beta_2))} * \frac{exp(\phi_1^{(1)}+\phi_2^{(1)})}{(1+exp(\phi_1^{(1)}+\phi_2^{(1)}))} * \frac{1}{(1+exp(\phi_1^{(2)}+\phi_2^{(2)}+\phi_3^{(2)}))})$$

$$+n_{23} * log((1-p_1) * \frac{exp(\beta_1)}{(1+exp(\beta_1))} * \frac{exp(\phi_1^{(1)})}{(1+exp(\phi_1^{(1)}))} * \frac{1}{(1+exp(\phi_1^{(2)}+\phi_2^{(2)}+\phi_4^{(2)}))} +$$

$$(1-p_1) * \frac{1}{(1+exp(\beta_1))} * \frac{exp(\phi_1^{(1)})}{(1+exp(\phi_1^{(1)}))} * \frac{1}{(1+exp(\phi_1^{(2)}+\phi_2^{(2)}))})$$

$$+n_{31} * log(p_1 * \frac{exp(\beta_1+\beta_2)}{(1+exp(\beta_1+\beta_2))} * \frac{1}{(1+exp(\phi_1^{(1)}+\phi_2^{(1)}))} * \frac{exp(\phi_1^{(2)}+\phi_3^{(2)}+\phi_4^{(2)})}{(1+exp(\phi_1^{(2)}+\phi_3^{(2)}+\phi_4^{(2)}))} +$$

$$(1-p_1) * \frac{exp(\beta_1)}{(1+exp(\beta_1))} * \frac{1}{(1+exp(\phi_1^{(1)}))} * \frac{exp(\phi_1^{(2)}+\phi_4^{(2)})}{(1+exp(\phi_1^{(2)}+\phi_4^{(2)}))})$$

$$+n_{32} * log(p_1 * \frac{1}{(1+exp(\beta_1+\beta_2))} * \frac{1}{(1+exp(\phi_1^{(1)}+\phi_2^{(1)}))} * \frac{exp(\phi_1^{(2)}+\phi_3^{(2)})}{(1+exp(\phi_1^{(2)}+\phi_3^{(2)}))} +$$

$$(1-p_1) * \frac{1}{(1+exp(\beta_1))} * \frac{1}{(1+exp(\phi_1^{(1)}))} * \frac{exp(\phi_1^{(2)})}{(1+exp(\phi_1^{(2)}))})$$

$$+n_{33} * log(p_1 * \frac{exp(\beta_1+\beta_2)}{(1+exp(\beta_1+\beta_2))} * \frac{1}{(1+exp(\phi_1^{(1)}+\phi_2^{(1)}))} * \frac{1}{(1+exp(\phi_1^{(2)}+\phi_3^{(2)}+\phi_4^{(2)}))} +$$

$$p_1 * \frac{1}{(1+exp(\beta_1+\beta_2))} * \frac{1}{(1+exp(\phi_1^{(1)}+\phi_2^{(1)}))} * \frac{1}{(1+exp(\phi_1^{(2)}+\phi_3^{(2)}))} +$$

$$(1-p_1) * \frac{exp(\beta_1)}{(1+exp(\beta_1))} * \frac{1}{(1+exp(\phi_1^{(1)}))} * \frac{1}{(1+exp(\phi_1^{(2)}+\phi_4^{(2)}))} +$$

$$(1-p_1) * \frac{1}{(1+exp(\beta_1))} * \frac{1}{(1+exp(\phi_1^{(1)}))} * \frac{1}{(1+exp(\phi_1^{(2)}))})$$

REFERENCES

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," pp. 152-156.

Conaway, M.R. (1993), *Non-ignorable non-response models for timeordered categorical Variables*, Appl. Statist. **42**, 105-115.

Fay, R.E. (1986), *Casual models for patterns of nonresponse*, J. Am. Statist. Ass. **81**, 354-365.

Fay, R.E. (1989), *Estimated nonignorable nonresponse in longitudinal surveys through causal modeling*, "Panel surveys (eds. Kasprzyk,D., Duncan, G., Kalton, G. and Singh, M.P.)," pp. 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 being imputed*, J. Am. Statist. Ass. **77**, 251-261.

Stasny, E. (1987), *Some Markov chain models for nonresponse in estimating gross labor force flows*, J. Off. Statist. **3**, 359-373.

Thomsen, I. (1981), *The use of Markov chain models in sampling from finite populations*, Scand. J. Statist. **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, New York.