A New Approach to Estimation of Response Probabilities when Missing Data are Not Missing at Random

Michail Sverchkov
BAE Systems IT and Bureau of Labor Statistics, Sverchkov.Michael@bls.gov

Abstract
Often the probability of responding depends directly on the outcome value. This case can be treated by postulating a parametric model for the distribution of the outcomes before nonresponse and a model for the response mechanism. The two models define a parametric model for the joint distribution of the outcomes and response indicators, and therefore the parameters of these models can be estimated by maximization of the likelihood corresponding to this distribution. Modeling the distribution of the outcomes before nonresponse, however, can be problematic since no data is available from this distribution. We propose an alternative approach that allows estimation of the parameters of the response model by first estimating the outcomes distribution of the respondents, and then solving an estimating equation defined by the census likelihood of the response indicators.

Key words: sample distribution, complement-sample distribution, prediction under informative sampling and non-response, estimating equations, missing information principle, non-parametric estimation

1. Introduction

There is almost no survey without nonresponse, but in practice most methods that deal with this problem assume either explicitly or implicitly that the missing data are 'missing at random' (MAR, Rubin, 1976; Little, 1982). However, in many practical situations this assumption is not valid, since the probability of responding often depends directly on the outcome value. In this case, the use of methods that assume that the nonresponse is MAR can lead to large biases of population parameter estimators and large imputation bias.

The case where the missing data are not MAR (NMAR) can be treated by postulating a parametric model for the distribution of the outcomes *before non-response* and a model for the response mechanism. These two models define a parametric model for the joint distribution of the outcomes and response indicators, and therefore the parameters of these models can be estimated by maximization of the likelihood based on the latter joint distribution. See, Greenlees *et al*. (1982), Rubin (1987), Little (1993), Beaumont (2000), Little and Rubin (2002) and Qin *et al*. (2002).

Modeling the distribution of the outcomes *before non-response* can be problematic since it refers to the partly unobserved data. Qin *et al.* (2002) suggests using a non-parametric model for this distribution (empirical likelihood approach). We suggest an alternative approach that allows one to estimate the parameters of the response model by independent parametric or non-parametric estimation of the outcomes distribution *after non-response* (which can be done by use of classic statistical inferences since the latter refers to the observed data) and then by solving estimating equations obtained from the census likelihood function of the response indicators or the equations based on the method of moments (MoM). The derivation of these estimating equations utilizes the relationships between the population, the sample and the sample-complement distributions, as in Pfeffermann and Sverchkov (1999, 2003), Sverchkov and Pfeffermann (2004).

2. Notation

Let Y_i denote the value of an outcome variable Y associated with unit i belonging to a sample $S = \{1,...,n\}$, drawn from a finite population $U = \{1,...,N\}$. Let X_i denote the corresponding values of covariates $X_i = (X_{1i},...,X_{Ki})'$. In what follows we assume that the population outcome values are independent realizations from distributions with unknown probability density functions (*pdf*), $f(Y_i|X_i)$. We use the abbreviation *pdf* for the probability density function when Y_i is continuous and the probability function when Y_i is discrete. Let $R=\{1,...,n_r\}$ define the sample of respondents (the sample with observed outcome values), and $R^c=\{n_r+1,...,n\}$ define the sample of nonrespondents. The response process is assumed to occur stochastically, independently between units. The observed sample of respondents can be viewed therefore as the result of a two-phase sampling process where in the first phase the sample S is selected from U with known inclusion probabilities $\pi_i=\Pr(i\in S)$ and in the second phase the sample R is 'self selected' with unknown response probabilities (Särndal and Swensson, 1987).

Denote by $p(Y_i,X_i)=\Pr(i\in R\,|\,Y_i,X_i,i\in S)$ and let u_i and v_i be any random vectors such that (u_i,v_i) and response indicators, R_i ($R_i=1$ if $i\in R$ and 0 otherwise), are independent given $(Y_i,X_i,i\in S)$. For example, u_i and v_i are functions of (Y_i,X_i), or the responses are completely defined by (Y_i,X_i). In what follows we use the following relationships between population and the sample distribution (Pfeffermann and Sverchkov 1999, 2003 and Sverchkov and Pfeffermann 2004) which can be written in terms of response probabilities as,

$$E(u_i\,|\,v_i,i\in S)=\frac{E(p^{-1}(Y_i,X_i)u_i\,|\,v_i,i\in R)}{E(p^{-1}(Y_i,X_i)\,|\,v_i,i\in R)}, \tag{2.1}$$

$$E(u_i\,|\,v_i,i\in R^c)=\frac{E\{[p^{-1}(Y_i,X_i)-1]u_i\,|\,v_i,i\in R)}{E\{[p^{-1}(Y_i,X_i)-1]\,|\,v_i,i\in R)}. \tag{2.2}$$

Note that (2.1) implies

$$E[p^{-1}(Y_i,X_i)\,|\,i\in R]=1/E[p(Y_i,X_i)\,|\,i\in S]. \tag{2.3}$$

Remark 2.1 In the following sections we concentrate on estimation of the response probabilities $p(Y_i,X_i)$. Note that if the response probabilities or their estimates are known then the sample respondents can be considered as a sample from the finite population with known or estimated selection probabilities $\tilde{\pi}_i=\pi_i\hat{p}(Y_i,X_i)$. Then population model parameters (or finite population parameters) can be estimated as if there were no non-response with these new inclusion probabilities, see Särndal and Swensson (1987). One can use these probabilities for imputation also using the relationship between the sample and sample-complement distributions derived in Sverchkov and Pfeffermann (2004),

$$f(u_i\,|\,v_i,i\in R^c)=\frac{[p^{-1}(Y_i,X_i)-1]f(u_i\,|\,v_i,i\in R)}{E\{[p^{-1}(Y_i,X_i)-1]\,|\,v_i,i\in R\}}. \tag{2.4}$$

3. Estimation of the Response Probabilities when Non-Response is NMAR based on Estimating Equations obtained from the Likelihood function of the Response Indicators (EEL)

In this section we suggest a new approach that enables us to estimate the probabilities $\Pr(i\in R\,|\,Y_i,X_i,i\in S)$ based only on knowledge of the respondents' *pdf*, $f[Y_i\,|\,X_i,i\in R]$. Let $p(Y_i,X_i;\gamma)=\Pr(i\in R\,|\,Y_i,X_i,i\in S;\gamma)$ and suppose that $p(Y_i,X_i;\gamma)$ is differentiable with respect to (vector) parameter γ.

In this and in the first part of the next section we consider the following scenario (***Scenario A***): The covariates are observed for all non-respondents, i.e. Observed Data= $\{Y_i,i\in R,X_k,k\in S\}$.

Under this scenario, if the missing data were later observed, γ could be estimated by solving the likelihood equations,

$$\sum_{i\in R}\frac{\partial\log p(Y_i,X_i;\gamma)}{\partial\gamma}+\sum_{i\in R^c}\frac{\partial\log[1-p(Y_i,X_i;\gamma)]}{\partial\gamma}=0. \tag{3.1}$$

Similarly to the Missing Information Principle (Cipillini *et al*, 1955, Orchard and Woodbury 1972), since the outcome values are missing for $j \in R^c$, we propose to solve instead,

$$0 = E[\sum_{i \in R} \frac{\partial \log p(Y_i, X_i; \gamma)}{\partial \gamma} + \sum_{i \in R^c} \frac{\partial \log[1 - p(Y_i, X_i; \gamma)]}{\partial \gamma} \mid \text{Observed Data}] \text{, i.e.,}$$

$$0 = E[\sum_{i \in R} \frac{\partial \log p(Y_i, X_i; \gamma)}{\partial \gamma} + \sum_{i \in R^c} \frac{\partial \log[1 - p(Y_i, X_i; \gamma)]}{\partial \gamma} \mid Y_i, i \in R, X_k, k \in S]$$

$$= \sum_{i \in R} \frac{\partial \log p(Y_i, X_i; \gamma)}{\partial \gamma} + \sum_{i \in R^c} E[\frac{\partial \log[1 - p(Y_i, X_i; \gamma)]}{\partial \gamma} \mid i \in R^c, X_k, k \in S]$$

$$= \sum_{i \in R} \frac{\partial \log p(Y_i, X_i; \gamma)}{\partial \gamma} + \sum_{j \in R^c} \frac{E\{(p^{-1}(Y_j, X_j; \gamma) - 1) \frac{\partial \log[1 - p(Y_j, X_j; \gamma)]}{\partial \gamma} \mid X_j, j \in R\}}{E\{(p^{-1}(Y_j, X_j; \gamma) - 1) \mid X_j, j \in R\}} \qquad (3.2a)$$

$$= \sum_{i \in R} \frac{\partial p(Y_i, X_i; \gamma)}{\partial \gamma} p^{-1}(Y_i, X_i; \gamma) - \sum_{j \in R^c} \frac{\int p^{-1}(Y_j, X_j; \gamma) \frac{\partial p(Y_j, X_j; \gamma)}{\partial \gamma} f(Y_j \mid X_j, j \in R) dY_j}{\int p^{-1}(Y_j, X_j; \gamma) f(Y_j \mid X_j, j \in R) dY_j - 1} \, . \qquad (3.2b)$$

The third equation follows from (2.2) where we assume for simplicity that $p(Y_j, X_j; \gamma)$ and $(X_k, k \in S)$ are independent given X_j. Note that the second sum in (3.2a) and (3.2b) predicts the unobserved second sum in (3.1). Note also that if $p(Y_j, X_j; \gamma)$ is a function of X_j and γ only (missing data are MAR) then (3.2b) reduces to a common system of log-likelihood equations, $\sum_{i \in R} \frac{\partial \log p(X_i; \gamma)}{\partial \gamma} - \sum_{i \in R^c} \frac{\partial \log[1 - p(X_i; \gamma)]}{\partial \gamma} = 0$.

Estimating functions (3.2b) suggest the following two-step estimation procedure:
Step 1. Fit the model $f_r(Y_i \mid X_i) = f[Y_i \mid X_i, i \in R]$. Note that this *pdf* refers to the respondents' sample and therefore can be identified and estimated from the observed data using classic statistical inferences.
Step 2. Approximate (3.2b) by replacing $f_r(Y_i \mid X_i)$ by its estimate, $\hat{f}_r(Y_i \mid X_i)$, and solve (3.2b) for γ .

Note that instead of estimation of f_r in (3.2b) one can estimate respective expectations in (3.2a) non-parametrically, and after substituting the estimates in (3.2a) solve them for γ . For example, for discrete X -s and an arbitrary function g , $E[g(Y_j, X_j, \gamma) \mid X_j = x, j \in R]$ can be estimated by the respective mean, $\left(\sum_{j \in R: X_j = x} 1 \right)^{-1} \sum_{j \in R: X_j = x} g(Y_j, X_j, \gamma)$. For continuous X -s let $m(x, \gamma)$ be an estimator of $E(g(Y_j, X_j, \gamma) \mid X_j = x, j \in R)$, for example the Nadaraya-Watson estimator, $m(x, \gamma) = \frac{\sum_{j \in R} K[(x - X_j)/h] g(Y_j, X_j, \gamma)}{\sum_{j \in R} K[(x - X_j)/h]}$, where h and K are a scale-factor and a kernel. Estimating the respective conditional expectations in the second sum of (3.2a) by $m(x, \gamma)$ one obtains the following *estimating equations*,

$$\sum_{i \in R} \frac{\partial p(Y_i, X_i; \gamma)}{\partial \gamma} p^{-1}(Y_i, X_i; \gamma) - \sum_{j \in R^c} \frac{\sum_{k \in R} K[(X_k - X_j)/h] p^{-1}(Y_k, X_k; \gamma) \frac{\partial p(Y_k, X_k; \gamma)}{\partial \gamma}}{\sum_{k \in R} K[(X_k - X_j)/h][p^{-1}(Y_k, X_k; \gamma) - 1]} = 0 \, , \qquad (3.3)$$

which defines an estimator of γ .

Estimating equations (3.3) do not require any knowledge of the model for the respondents. On the other hand one can expect that the estimates obtained by solving (3.3) will be less stable than the estimates obtained from (3.2b) by the above two step estimation procedure when the model for the respondents can be fitted well.

4. Estimation of the Response Probabilities when Non-Response is NMAR based on Estimating Equations obtained by the Method of Moments (MoM)

The response probabilities can also be estimated by solving estimating equations obtained by application of MoM and the relationship (2.1).

Under scenario A, by (2.1), for any function G (for example $G(X_i) = (X_i^1,...,X_i^K)'$, so that $E[G(X_i) \mid i \in S]$ is a vector of the first K moments of X over the sample distribution),

$$E[G(X_i) \mid i \in S] = E\{E[G(X_i) \mid Y_i, i \in S] \mid i \in S\}$$

$$= \frac{E\{p^{-1}(Y_i, X_i, \gamma)\dfrac{E[p^{-1}(Y_i, X_i, \gamma)G(X_i) \mid Y_i, R_i = 1]}{E[p^{-1}(Y_i, X_i, \gamma) \mid Y_i, R_i = 1]} \mid R_i = 1\}}{E\{p^{-1}(Y_i, X_i, \gamma) \mid R_i = 1\}}, \tag{4.1}$$

and $E[p^{-1}(Y_i, X_i, \gamma) \mid R_i = 1] = n / E(n_r)$. (4.2)

Eq. 4.2 follows from (2.3) and $\sum_{i \in S} p(Y_i, X_i, \gamma) = E(n_r)$.

Let $\dim(\gamma) = K+1$. If for some G, $\dim(G) = K$, the system of estimating functions (4.1) – (4.2) has an unique solution then it defines the parameter γ. Therefore one can apply MoM to (4.1) – (4.2). For example, if Y_i is finite and discrete then all expectations in (4.1) – (4.2) can be approximated by the respective means,

$$n^{-1}\sum_{i \in S} G(X_i) = \frac{\sum_{i \in R} p^{-1}(Y_i, X_i, \gamma)\dfrac{\sum_{j \in R} p^{-1}(Y_j, X_j, \gamma)G(X_j)\mathbf{1}_{(Y_j = Y_i)}}{\sum_{j \in R} p^{-1}(Y_j, X_j, \gamma)\mathbf{1}_{(Y_j = Y_i)}}}{\sum_{i \in R} p^{-1}(Y_i, X_i, \gamma)}, \text{ where } \mathbf{1}_{(Y_j = Y_i)} = \begin{cases} 1 \text{ if } Y_j = Y_i \\ 0 \text{ otherwise} \end{cases} \tag{4.3}$$

and $\sum_{i \in R} p^{-1}(Y_i, X_i, \gamma) = n$, (4.4)

and γ can be estimated by solving *estimating equations* (4.3) – (4.4). If Y_i is continuous then one can estimate the respective expectations given Y_i in (4.1) by an appropriate non-parametric estimator, for example by the Nadaraya-Watson estimator as in (3.3).

Consider another scenario (***Scenario B***): Let Z_i be a vector which is independent of the response indicators, R_i, given $(Y_i, X_i, i \in S)$. The covariates and the values of Z_i are observed only for the respondents and the finite population total $Z = \sum_{k \in U} Z_i$ is known, i.e. Observed Data $= \{Y_i, X_i, Z_i, \pi_i, i \in R, Z\}$. In practice Z_i could be a sub-vector of the covariates, X_i, or some other variable believed to be dependent on the response only through Y_i and X_i.

Under scenario B (assuming $\dim\gamma = \dim Z_i + 1$) one can replace (4.1) by (4.5),

$$E[Z_i] = \frac{E\{\pi_i^{-1}p^{-1}(Y_i, X_i, \gamma)\dfrac{E[\pi_i^{-1}p^{-1}(Y_i, X_i, \gamma)Z_i \mid Y_i, R_i = 1]}{E[\pi_i^{-1}p^{-1}(Y_i, X_i, \gamma) \mid Y_i, R_i = 1]} \mid R_i = 1\}}{E\{\pi_i^{-1}p^{-1}(Y_i, X_i, \gamma) \mid R_i = 1\}}, \tag{4.5}$$

and estimate $E[Z_i]$ by Z / N.

Remark 4.1 Note that (4.1) – (4.2) does not necessarily have a unique solution. For example, if X_i takes only two values, say 0 and 1, then (4.1) – (4.2) can not be uniquely solved if $\dim(\gamma) > 2$ for any function *G*.

Remark 4.2 In practical situations $\dim \gamma$ can be less than $\dim Z_i + 1$. In this case, if it is preferred to use all auxiliary information, then γ can be estimated (for example) by solving $\hat{\gamma} = \arg\min_{\gamma} \Delta'(\gamma)\Delta(\gamma)$ where

$$\Delta(\gamma) = \hat{E}[Z_i] - \frac{\hat{E}\{\pi_i^{-1} p^{-1}(Y_i, X_i, \gamma) \dfrac{\hat{E}[\pi_i^{-1} p^{-1}(Y_i, X_i, \gamma) Z_i \mid Y_i, R_i = 1]}{\hat{E}[\pi_i^{-1} p^{-1}(Y_i, X_i, \gamma) \mid Y_i, R_i = 1]} \mid R_i = 1\}}{\hat{E}\{\pi_i^{-1} p^{-1}(Y_i, X_i, \gamma) \mid R_i = 1\}}.$$

5. Empirical illustration

In order to illustrate the performance of the estimators described in the previous sections we designed a small Monte Carlo study. For simplicity assume that the finite population and the sample coincide, $U = S$. The study consists of the following four steps.

Step A: Generate independently $M = 1000$ finite populations $\{Y_i^{(m)}, X_i^{(m)}; i = 1,...,1000\}$, where $X_i^{(m)} \sim$ Uniform(0,1), $P(Y_i^{(m)} = 1 \mid X_i) = (\exp\{\theta_0 + \theta_1 X_i^{(m)}\} + 1)^{-1}$, $\theta_0 = -0.1$, $\theta_1 = -1$, and $P(Y_i^{(m)} = 0 \mid X_i^{(m)}) = 1 - P(Y_i^{(m)} = 1 \mid X_i^{(m)})$. For each finite population calculate the mean of the outcome variable, $\overline{Y}^{(m)} = \frac{1}{1,000} \sum_{i=1}^{1,000} Y_i^{(m)}$ = the proportion of the positive outcomes.

Step B: For each generated population generate independently response indicators from the logistic model, $P(R_i^{(m)} = 1 \mid Y_i^{(m)}, X_i^{(m)}) = \dfrac{1}{\exp(\gamma_0 + \gamma_1 X_i^{(m)} + \gamma_2 Y_i^{(m)}) + 1}$.

We repeat the study for different values of $\gamma = (\gamma_0, \gamma_1, \gamma_2)$, see Figures 1 - 4.

Step C: For each sample of respondents, $\{Y_i^{(m)}, X_i^{(m)}; i \in (k : R_k^{(m)} = 1)\}$, estimate response probabilities using three estimators:

(1) $\hat{P}_i^{(m)MAR} = Q(X_i^{(m)}; \hat{\gamma}_0^{(m)MAR}, \hat{\gamma}_1^{(m)MAR})$ where $Q(X_i^{(m)}; \gamma_0, \gamma_1) = \dfrac{1}{\exp(\gamma_0 + \gamma_1 X_i^{(m)}) + 1}$ and $(\hat{\gamma}_0^{(m)MAR}, \hat{\gamma}_1^{(m)MAR})$ is a solution of the likelihood equations $\sum_{i \in R} \dfrac{\partial \log Q(X_i^{(m)}; \gamma_0, \gamma_1)}{\partial \gamma_d} + \sum_{i \in R^c} \dfrac{\partial \log[1 - Q(X_i^{(m)}; \gamma_0, \gamma_1)]}{\partial \gamma_d} = 0$, $d = 0,1$ (MLE estimate assuming that the response is MAR and response model is logistic). These estimates were derived using Proc Logistic of SAS.

(2) $\hat{P}_i^{(m)MoM} = p(Y_i^{(m)}, X_i^{(m)}; \hat{\gamma}_0^{(m)MoM}, \hat{\gamma}_1^{(m)MoM}, \hat{\gamma}_2^{(m)MoM})$ where $p(Y_i^{(m)}, X_i^{(m)}; \gamma_0, \gamma_1, \gamma_2) = \dfrac{1}{\exp(\gamma_0 + \gamma_1 X_i^{(m)} + \gamma_2 Y_i^{(m)}) + 1}$ and $(\hat{\gamma}_0^{(m)MoM}, \hat{\gamma}_1^{(m)MoM}, \hat{\gamma}_2^{(m)MoM})$ is a solution of (4.3) - (4.4). Proc IML SAS optimization functions were used for solving (4.3) – (4.4).

(3) $\hat{P}_i^{(m)EEL} = p(Y_i^{(m)}, X_i^{(m)}; \hat{\gamma}_0^{(m)EEL}, \hat{\gamma}_1^{(m)EEL}, \hat{\gamma}_2^{(m)EEL})$ where $(\hat{\gamma}_0^{(m)EEL}, \hat{\gamma}_1^{(m)EEL}, \hat{\gamma}_2^{(m)EEL})$ is a solution of (3.2b). Recall that solving (3.2b) requires estimating $f_r(Y_i \mid X_i)$. We used $P(Y_i = 1 \mid X_i, R_i = 1) = (\exp\{a_0 + a_1 X_i\} + 1)^{-1}$ as the working model for $f_r(Y_i \mid X_i)$ and estimated the parameters (a_0, a_1) by Proc Logistic SAS. Note that the true model for $f_r(Y_i \mid X_i)$ is not necessarily linear logistic (unless $\gamma_1 = 0$), although the later can be a good approximation. Proc IML SAS optimization functions were used for solving (3.2b).

Step D: For each sample calculate three Hajek-type estimates of the population mean of the outcome variable, $\overline{Y}^{(m)}$:

"MAR": $\hat{\overline{Y}}^{(m)MAR} = \dfrac{\sum_{i:R_i=1} Y_i^{(m)} / \hat{P}_i^{(m)MAR}}{\sum_{i:R_i=1} 1/\hat{P}_i^{(m)MAR}}$, "MoM": $\hat{\overline{Y}}^{(m)MoM} = \dfrac{\sum_{i:R_i=1} Y_i^{(m)} / \hat{P}_i^{(m)MoM}}{\sum_{i:R_i=1} 1/\hat{P}_i^{(m)MoM}}$,

"EEL": $\hat{\overline{Y}}^{(m)EEL} = \dfrac{\sum_{i:R_i=1} Y_i^{(m)} / \hat{P}_i^{(m)EEL}}{\sum_{i:R_i=1} 1/\hat{P}_i^{(m)EEL}}$.

For each of these three estimators calculate Empirical Bias and Empirical Root Mean Square Error over 1,000 simulations, $Bias(\hat{\overline{Y}}) = \frac{1}{1,000}\sum_{m=1}^{1,000}(\hat{\overline{Y}}^{(m)} - \overline{Y}^{(m)})$ and $RMSE(\hat{\overline{Y}}) = \sqrt{\frac{1}{1,000}\sum_{m=1}^{1,000}(\hat{\overline{Y}}^{(m)} - \overline{Y}^{(m)})^2}$.

We repeat the study for different values of the parameter γ. The results are summarized in Figures 1 – 4.

Figure 1. Empirical Bias of three estimators (Gamma_0=−1, Gamma_1=−1)

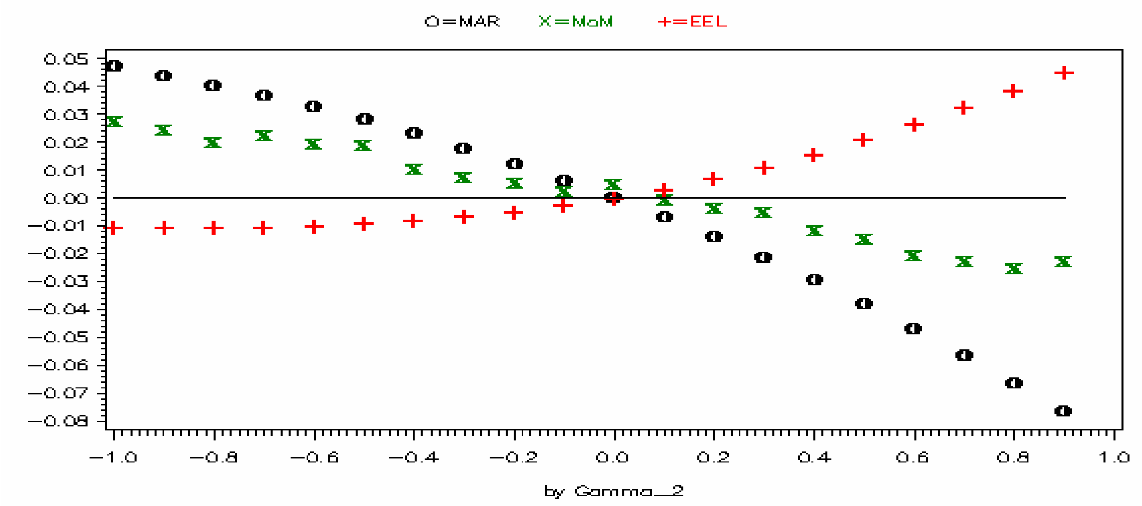
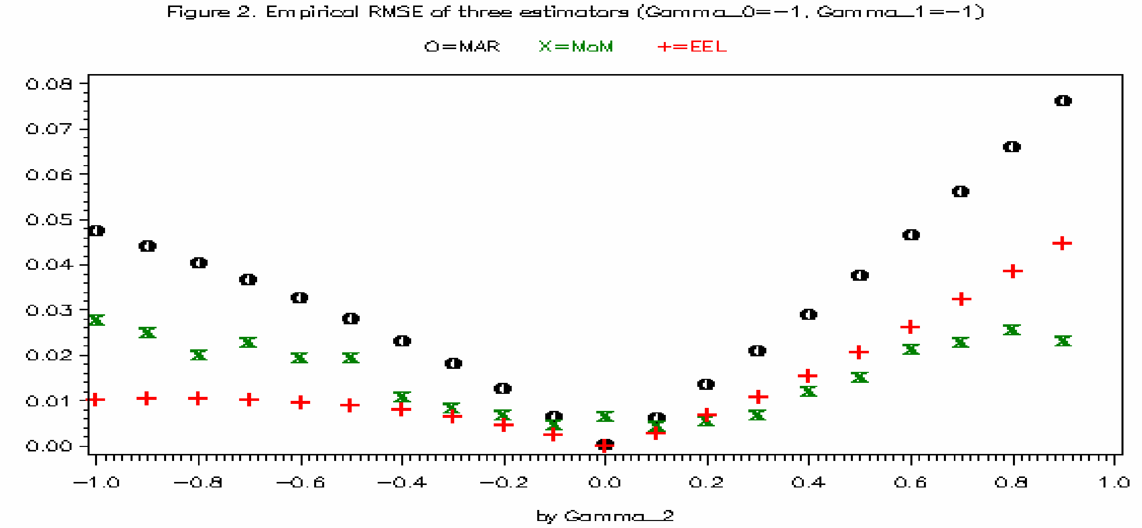


Figure 2. Empirical RMSE of three estimators (Gamma_0=−1, Gamma_1=−1)



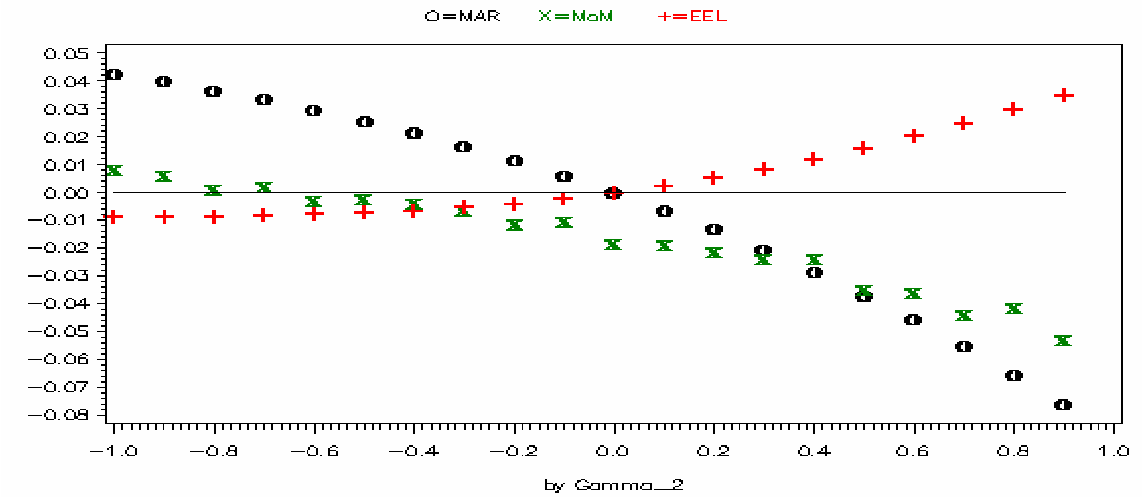


Figure 3. Empirical Bias of three estimators (Gamma_0=−1, Gamma_1=0)

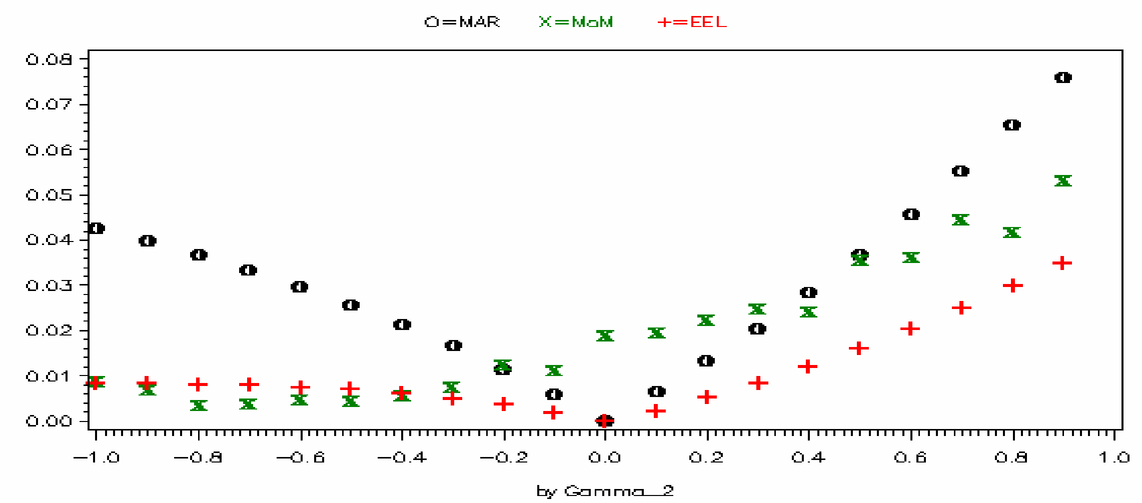


Figure 4. Empirical RMSE of three estimators (Gamma_0=−1, Gamma_1=0)

Conclusions

1) "MAR" estimates of the population proportions of positive outcomes, $\bar{Y}^{(m)}$, are significantly biased except for the case of $\gamma_2 = 0$ which corresponds to "missing data are MAR".
2) Although both suggested estimators, the "MoM" estimator and the "EEL" estimator are also significantly biased for $\gamma_2 \neq 0$ their bias and RMSE are much smaller than those from the "MAR" estimator.
3) In the case $\gamma_2 = 0$, RMSE is larger for the "MoM" estimator than for the two other estimators. Note that Method of Moment type estimators are usually less effective than estimators based on MLE principle.
4) In our limited study the "EEL" estimator performs not significantly worse in the sense of bias and RMSE than the "MAR" estimator even when "missing data are MAR", $\gamma_2 = 0$. Note also that in this case the estimates $\hat{\gamma}_2^{(m)EECL}$ differ insignificantly from zero (the results are not shown in the paper). Therefore testing the hypothesis $H_0 : \hat{\gamma}_2^{(m)EECL} = 0$ is equivalent to testing whether the response is MAR or NMAR (if the model for response mechanism is specified correctly). Methods for testing H_0, in particular estimating the variance of $\hat{\gamma}_2^{(m)EECL}$, is a topic for future research.

Another important topic for future research is connected with "identifiabilty" of the parameter γ: for example, even if the solution of (3.1) is unique it does not necessarily imply the unique solution for (3.2b). See also Remark 4.1.

Acknowledgements

Some of this research is supported by a grant from the United Sates-Israel Binational Science Foundation (BSF). The author thanks Alan Dorfman and Danny Pfeffermann for useful discussions.

References

Beaumont, J.F. (2000). "An estimation method for nonignorable nonresponse", *Survey Methodology*, **26**, 131-136.

Cepillini, R., Siniscialco, M., and Smith, C.A.B. (1955). "The estimation of gene frequencies in a random mating population", *Annals of Human Genetics*, **20**, 97-115.

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

Little, R.J.A. (1982). "Models for nonresponse in sample surveys", *Journal of the American Statistical Association* **77** , 237-250.

Little, R.J.A. (1993) "Pattern-mixture models for multivariate incomplete data", *Journal of the American Statistical Association*, **88** , 125-134.

Little, R.J.A. and Rubin, D.B. (2002). "*Statistical analysis with missing data",* New York: Wiley.

Rubin, D.B. (1976). "Inference and missing data", *Biometrika* **63**, 581-590.

Rubin, D.B. (1987). "*Multiple imputation for nonresponse in surveys",* New York: Wiley

Qin, J., Leung, D. and Shao, J. (2002). "Estimation with Survey data under nonignorable nonresponse or informative sampling", *Journal of the American Statistical Association* **97**, 193-200.

Orchard, T., and Woodbury, M.A. (1972). "A missing information principle: theory and application", *Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 697-715.

Pfeffermann, D., and Sverchkov, M. (1999). "Parametric and semi-parametric estimation of regression models fitted to survey data", *Sankhya* **61**, 166-186.

Pfeffermann, D., and Sverchkov, M. (2003). "Fitting generalized linear models under informative probability sampling", In: *Analysis of Survey Data*, eds. C. J. Skinner and R. L. Chambers. New York: John Wiley & Sons. pp 175-195.

Sarndal C.E., and Swensson B. (1987). "A general view of estimation for two fases of selection with applications to two-phase sampling and nonresponse", I*nternational Statistical Review* **55**, 279-294.

Sverchkov, M., and Pfeffermann, D. (2004). "Prediction of finite population totals based on the sample distribution" *Survey Methodology* **30**, 79-92.

The opinions expressed in this paper are those of the author and do not necessarily represent the policies of the Bureau of Labor Statistics