Survey Methodology, December 2000
Vol. 26, No. 2, pp. 131-136
Statistics Canada, Catalogue No. 12-001

An Estimation Method for Nonignorable Nonresponse

Jean-François Beaumont [1]

Abstract

When a survey response mechanism depends on a variable of interest measured within the same survey and observed for only part of the sample, the situation is one of nonignorable nonresponse. In such a situation, ignoring the nonresponse can generate significant bias in the estimation of a mean or of a total. To solve this problem, one option is the joint modelling of the response mechanism and the variable of interest, followed by estimation using the maximum likelihood method. The main criticism leveled at this method is that estimation using the maximum likelihood method is based on the hypothesis of error normality for the model involving the variable of interest, and this hypothesis is difficult to verify. In this paper, the author proposes an estimation method that is robust to the hypothesis of normality, so constructed that there is no need to specify the distribution of errors. The method is evaluated using Monte Carlo simulations. The author also proposes a simple method of verifying the validity of the hypothesis of error normality whenever nonresponse is not ignorable.

Key Words: Nonignorable nonresponse; Maximum likelihood; Estimation equations; Regression imputation; Reweighting.

1. Introduction

When a survey response mechanism depends on a variable of interest measured in the same survey and observed for only part of the sample, the situation is one of nonignorable nonresponse. In measuring income, for example, it may be realistic to assume that low income earners will exhibit a lower tendency to respond than high income earners, or vice versa. Readers will find in Little (1982) a formal definition of the concept of nonignorable nonresponse. In such a situation, ignoring the nonresponse can generate significant bias in the estimation of a mean or of a total. To solve this problem, one option is the joint modelling of the response mechanism and the variable of interest, followed by estimation using the method of maximum likelihood, used for example in Greenlees, Reece and Zieschang (1982), and imputation of the missing values. The main criticism leveled at this method is that estimation using the method of maximum likelihood is based on the hypothesis of error normality for the model involving the variable of interest, and this hypothesis is difficult to verify.

Rancourt, Lee and Särndal (1994) described simple correction factors aimed at reducing the bias generated by nonresponse that is not ignorable without reference to the hypothesis of normality and in the absence of a response mechanism model. These correction factors, however, are only available for ratio imputation.

In this paper, the author proposes an estimation method that is robust with respect to the hypothesis of normality, so constructed that there is no need to specify the distribution of errors. The author also proposes a simple method of verifying the validity of the hypothesis of error normality whenever nonresponse is not ignorable.

In section 2, the problem is defined and some notation is introduced. In section 3, various estimators of the mean of a population are introduced for a variety of hypotheses concerning the response mechanism and the distribution of data. In section 4, an estimation method is proposed for nonignorable nonresponse. In section 5, the author describes the results of a simulation study used to compare the estimators described in the two preceding sections. Finally, the last section contains a brief discussion.

2. Notation

In the following, we attempt to estimate the mean of a variable Y for a certain population P. To do so, we select a sample S, and the variable Y is observed for only part of the sample. The sample of respondents is denoted R, and the sample of nonrespondents is denoted O. We assume that there is at least one variable that is observed for all the sampling units and correlated with Y.

The estimator of the mean, $\mu = \sum_{i \in P} Y_i / N$, where N is the size of the population, can be obtained by weighting the respondent units:

$$\mu_P^* = \frac{\sum_{i \in R} w_i \, w_{R,\,i}^* \, Y_i}{\sum_{i \in R} w_i \, w_{R,\,i}^*}, \tag{2.1}$$

where w_i denotes the sampling weights that correspond to the inverse selection probability and $w_{R,\,i}^*$ denotes the weights that correspond to the estimated inverse response probability. Another estimator of the mean can be obtained by imputing the missing values:

1. Jean-François Beaumont, Household Survey Methods Division, Statistics Canada, Ottawa, Ontario, Canada K1A 0T6.

$$\mu_I^* = \frac{\sum_{i\in R} w_i Y_i + \sum_{i\in O} w_i Y_i^*}{\sum_{i\in S} w_i}, \tag{2.2}$$

where Y_i^* denotes values that are imputed for the nonrespondent units.

For the sake of simplicity, we assume, in the following, that the sampling weights are constant for all units of the population. Thus, we can eliminate w_i from equations (2.1) and (2.2). We also assume that there is only one observed variable for all sampling units. This variable is denoted X.

3. Current Estimation Methods

In this section, equations (2.1) and (2.2) are developed under a variety of hypotheses concerning the response mechanism and the distribution of data, and appropriate estimation methods are described. In section (3.1), we assume a uniform response mechanism; in section (3.2), we assume a response mechanism that depends on X, while in section (3.3), we assume a response mechanism that depends on Y. The response mechanisms in sections (3.1) and (3.2) are ignorable, whereas the one in section (3.3) is not ignorable.

3.1 Uniform Response Mechanism

Assuming a uniform response mechanism, we have the same response probability for all sampling units. Thus, estimator (2.1) becomes:

$$\mu_{P,U}^* = \frac{\sum_{i\in R} Y_i}{n_R}, \tag{3.1}$$

where n_R is the total number of respondents. This estimator is the very same one we would have obtained by using equation (2.2) and by imputing the respondent mean for all nonrespondents.

3.2 Response Mechanism Dependent on X

When the response mechanism depends on variable X (correlated with Y), estimator (3.1) might be strongly biased. It is then preferable to use this variable as additional information for the estimation of mean μ.

Estimator (2.1) can be obtained by replacing $1/w_{R,i}^*$ by the estimated response probability using a logistic regression. A response probability model is therefore needed. If we only have one observed variable (X) for all sampling units, the model can be written as follows:

$$P(R_i = 1 \mid X_i) = \frac{1}{1 + \exp[-(\alpha_0 + \alpha_1 X_i)]},$$

where α_0 and α_1 are parameters to be estimated (using the maximum likelihood method, for example) and R_i is a dichotomous variable equal to 1 if unit i responds and to 0 otherwise. The estimator of the mean obtained in this way is denoted $\mu_{P,X}^*$.

If we prefer to use estimator (2.2) instead, the missing values can be imputed using the following model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \tag{3.2}$$

where β_0 and β_1 are unknown parameters and ε_i is a random error term of zero mean that is not correlated with X_i. The imputed values are given by: $Y_i^* = B_0^* + B_1^* X_i$, where B_0^* and B_1^* are estimates (obtained by means of the method of least squares using the respondent units) of B_0 and B_1 which are in turn estimates of β_0 and β_1. In fact, B_0 and B_1 are the estimates that would have been obtained (using the method of least squares) if we had observed all the units of sample S. The estimator obtained in this way is denoted $\mu_{I,X}^*$. Note that all the models considered in this document are assumed to be valid for all the units of sample S.

We could also add a residual to the imputed values in order to better estimate the variance due to sampling (see for example Gagnon, Lee, Rancourt and Särndal 1996). However, this technique still does not make it possible to estimate the variance due to imputation. Moreover, it tends to produce estimates of the mean that are less precise than if no residual had been added. Since this paper does not deal with variance estimation, we have chosen not to add residuals to the imputed values. This has the added advantage of simplifying the calculation of $\mu_{I,X}^*$.

3.3 Response Mechanism Dependent on *Y*

All the estimators of the mean discussed so far can be strongly biased when the response mechanism depends on Y (nonignorable response mechanism). For such a response mechanism, the response probability can be modelled as follows:

$$P(R_i = 1 \mid Y_i) = \frac{1}{1 + \exp[-(\alpha_0 + \alpha_1 Y_i)]}. \tag{3.3}$$

Since variable Y is only observed for respondent units, it is impossible to obtain an estimated for α_0 and α_1 using the maximum likelihood method. Model (3.2) can also be used. However, the parameter estimates will not be consistent since $\mathrm{E}(\varepsilon_i \mid R_i = 1)$ and $\mathrm{E}(\varepsilon_i X_i \mid R_i = 1)$ are not zero. Even if we had consistent estimates, the missing values could not be imputed as described in section (3.2) since $\mathrm{E}(Y_i \mid R_i = 0, X_i) \neq \beta_0 + \beta_1 X_i$ (Greenlees, Reece and Zieschang 1982). If, for example, the response probability correlates positively with the variable of interest Y, then, for a given value of X, the mean of nonrespondent units will be lower than that of respondent units, and will therefore be lower than the mean of all units taken together. A similar argument can be presented if the response probability correlates negatively with the variable of interest. In fact, it can be shown that

$$\mathrm{E}(Y_i | R_i = 0, X_i) = \beta_0 + \beta_1 X_i - \frac{\operatorname{cov}(Y_i, p(Y_i) | X_i)}{1 - \mathrm{E}(p(Y_i) | X_i)},$$

where $p(Y_i) = P(R_i = 1 | Y_i)$.

The two approaches in section (3.2) are therefore invalid when the response mechanism is not ignorable. In such a situation, a better approach would be to estimate the parameters of models (3.2) and (3.3) simultaneously. The method of maximum likelihood can be used to this end. This method, however, requires as an additional hypothesis that errors ε_i follow a normal distribution (or any other distribution relevant to the type of data analyzed) with constant variance σ^2, and that they be mutually independent. The natural logarithm of the likelihood function l can be written as follows:

$$l = \sum_{i \in R} \ln[p(Y_i) f(Y_i | X_i)] + \sum_{i \in O} \ln[1 - \mathrm{E}(p(Y_i) | X_i)], \quad (3.4)$$

where $f(Y_i | X_i)$ is the probability density function of a normal distribution with a mean $\beta_0 + \beta_1 X_i$ and variance σ^2. The method of maximum likelihood consists in finding the parameter values which maximize l. To carry out the maximization, it must be possible to approximate $\mathrm{E}(p(Y_i) | X_i)$. This can be achieved by using a numerical integration method similar to that of Greenlees, Reece and Zieschang (1982). In this paper, the following approximation (Zeger, Liang and Albert 1988) has been used instead:

$$\mathrm{E}(p(Y_i) | X_i) \approx \frac{1}{1 + \exp\{-k[\alpha_0 + \alpha_1(\beta_0 + \beta_1 X_i)]\}}, \quad (3.5)$$

where $k = 1/\sqrt{c^2 \sigma^2 \alpha_1^2 + 1}$ and $c = 16\sqrt{3}/15\pi$. This approximation is based on the hypothesis that errors follow a normal distribution with constant variance. This approximation was preferred to a method of numerical integration because it is simpler and computationally faster, an advantage that must be considered seriously before any simulation study is undertaken. Finally, equation (3.4) was maximized using the Newton-Raphson algorithm and the NLIN procedure of the SAS software (SAS Institute Inc. 1990).

Once the parameters of models (3.2) and (3.3) have been estimated, estimators of the mean (2.1) or (2.2) can be chosen. Estimator (2.1) is obtained by replacing $w^*_{R,i}$ by $1/p^*(Y_i)$, where $p^*(Y_i)$ is the estimated response probability. This estimator is denote $\mu^*_{P,Y,ML}$. Estimator (2.2) can be obtained by determining imputed values Y^*_i in such a way that $\sum_{i \in S} e_i^2$ is minimized and that the constraints $\sum_{i \in S} e_i = 0$ and $\sum_{i \in S} e_i X_i = 0$ are met, where $e_i = Y_i - \beta^*_0 - \beta^*_1 X_i$, for $i \in R$, $e_i = Y^*_i - \beta^*_0 - \beta^*_1 X_i$, for $i \in O$, and β^*_0 and β^*_1 are the estimates of β_0 and β_1 respectively. The estimator of the mean can then be written as follows: $\mu^*_{I,Y,ML} = \beta^*_0 + \beta^*_1 \sum_{i \in S} X_i / n$, where n is the size of sample S. The reasoning behind this approach is that the two previous constraints would have been met if variable Y had been observed for all units in the sample and if this variable had been modelled using model (3.2).

4. Proposed Method of Estimation

This section describes the proposed method of estimation for a nonignorable response mechanism (section 4.1), as well as a graphic method (section 4.2) that can be used to verify the error normality hypothesis of model (3.2).

4.1 Method of Estimation for a Response Mechanism Dependent on *Y*

The method of maximum likelihood is valid when errors exhibit a normal distribution and have the same variance. When this hypothesis does not hold, it is preferable to use a more robust method of estimation.

If response probabilities $p(Y_i)$ were known and greater than zero for all sampling units, a robust method of estimation (in terms of both the error normality hypothesis and model 3.2) would consist in minimizing the error sum of squares weighted by the inverse response probability $p(Y_i)$. This minimization is equivalent to solving the system of equations

$$\sum_{i \in R} \frac{1}{p(Y_i)} (Y_i - \beta_0 - \beta_1 X_i) Z_{ik} = 0, \qquad k = 1, 2, \quad (4.1)$$

where $Z_{i1} = 1$ and $Z_{i2} = X_i$. This approach is considered robust with respect to the normality hypothesis since the method of least squares does not require that the distribution of errors be specified. Weighting by means of the inverse response probability also provides a certain robustness in terms of model (3.2). In fact, estimators β^*_0 and β^*_1 obtained using equation (4.1) are consistent with respect to the response mechanism for β_0 and β_1 (which are the estimators of β_0 and β_1 that would have been obtained if there had been no nonresponse) regardless of the validity of the model. A similar argument may be found in Särndal, Swensson and Wretman (1992, page 519), but in terms of the sample selection mechanism instead of the response mechanism.

Likewise, if the probability density function $f(Y_i | X_i)$ was known (not necessarily normal and yet not dependent on the parameters of model 3.3), we could then estimate parameters α_0 and α_1 of model (3.3) using the maximum likelihood method, for example, and solve the system of equations

$$\sum_{i \in R} \frac{\partial}{\partial \alpha_k} \ln[p(Y_i)] + \sum_{i \in O} \frac{\partial}{\partial \alpha_k} \ln[1 - \mathrm{E}(p(Y_i) | X_i)] = 0, \quad (4.2)$$

for $k = 0$ and $k = 1$.

Thus, the estimates of parameters $\beta_0, \beta_1, \alpha_0$ and α_1 are obtained by solving the unbiased estimation equations (4.1) and (4.2). An algorithm that can be used to find the solution consists in solving alternately the systems of equations (4.1) and (4.2) until convergence is achieved. This requires the possibility of calculating $E(p(Y_i) \mid X_i)$ in equation (4.2). However, this last expectation requires that the distribution of errors ε_i be known, and in all likelihood it is unknown. To get around this problem, we must use an approximation, and a number of them can be considered, including approximation (3.5). Another option would be to develop a strategy based on the bootstrap method by selecting the respondent units proportionally to their inverse response probability. However, this method requires considerable computer processing time, and is not considered in this paper. Instead, we have chosen the following approximation, obtained by linearizing $p(Y_i)$ using a Taylor series assessed at point $E(Y_i \mid X_i)$ and by taking the expectation of the first two terms in this series:

$$E(p(Y_i) \mid X_i) \approx p(E(Y_i \mid X_i)) = p(\beta_0 + \beta_1 X_i). \quad (4.3)$$

It should be noted that the expectation of the second term in the series is zero. This approximation offers the advantage of requiring only the first moment of the distribution of Y_i conditional on X_i. In this sense, it should be robust with respect to the error normality hypothesis since it does not require that the error distribution be specified. Of course, if the distribution of errors is known or can be properly estimated, it will be possible to find better approximations than (4.3) although, in this case, it may be preferable to use the maximum likelihood method.

Another interesting property of approximation (4.3) is that alternately solving the systems of equations (4.1) and (4.2) might be achieved using the following algorithm:

1. determine initial values for the response probabilities (or for parameters α_0 and α_1), *e.g.*, let $p(Y_i)^{(0)} = 1$ for all the respondent units;
2. let $j = 1$, where j is the number of iterations;
3. solve the system of equations (4.1) by means of the current response probability estimates, $p(Y_i)^{(j-1)}$, using a weighted regression procedure to obtain $\beta_0^{(j)}$ and $\beta_1^{(j)}$;
4. impute the missing values using $Y_i^{(j)} = \beta_0^{(j)} + \beta_1^{(j)} X_i$ for $i \in O$;
5. solve the system of equations (4.2) by using a logistic regression procedure to obtain $p(Y_i)^{(j)}$;
6. stop once convergence has been achieved, otherwise let $j = j + 1$ and return to step 3.

It is sufficient then to simply have a linear regression procedure and a logistic regression procedure to obtain the desired estimates. This algorithm is a very efficient means of finding the solution although, in certain cases, many iterations might be needed before convergence is achieved. In actual practice, it did converge in all cases where it was used. It should also be noted that this algorithm shows certain similarities with the EM algorithm used by Dempster, Laird and Rubin (1977), except that here we do not maximize a likelihood function.

For the simulations in the next section, we selected instead the Newton-Raphson algorithm which converges more rapidly. However, the above-mentioned algorithm had to be used for the few cases in which the Newton-Raphson algorithm met with convergence problems.

The proposed algorithm might be very useful as a means of providing initial values for a more rapid algorithm such as the Newton-Raphson one. The proposed algorithm could simply be used with a not very demanding convergence criterion so that, after only a few iterations, it could provide sufficiently good initial values to ensure convergence of the Newton-Raphson algorithm. In a different context, Beaumont and Demnati (1998) used a similar approach by beginning the iterative process using an algorithm of the EM type so as to provide the initial values for a more rapid algorithm of the Newton-Raphson type. They were able to show empirically that the combination of the two algorithms represents a sound compromise between processing time and efficiency in finding a solution.

As in section (3.3), once the parameters of models (3.2) and (3.3) are estimated, we can select estimators of the mean (2.1) or (2.2). Estimator (2.1) is obtained by replacing $w^*_{R,i}$ by $1/p^*(Y_i)$, where $p^*(Y_i)$ is the estimated response probability. This estimator is denoted $\mu^*_{P,Y,ROB}$. Estimator (2.2) is also obtained as in section (3.3) by determining the imputed values Y_i^* in such a way that $\sum_{i \in S} e_i^2$ is minimized and the constraints $\sum_{i \in S} e_i = 0$ and $\sum_{i \in S} e_i X_i = 0$ are met, where $e_i = Y_i - B_0^* - B_1^* X_i$, for $i \in R$, and $e_i = Y_i^* - B_0^* - B_1^* X_i$, for $i \in O$. This estimator is denoted $\mu^*_{I,Y,ROB}$. The quality of these two estimators of the mean will depend largely on the validity of models (3.2) and (3.3) and on the quality of approximation (4.3).

A modification of step (5) for the algorithm presented in this section was proposed by Beaumont (1999). The results of a simulation study show that this modification provides results that are slightly better than those obtained using the method proposed in this paper. However, this no longer involves using the maximum likelihood method to estimate the parameters of model (3.3), given that $f(Y_i \mid X_i)$ is known and a logistic regression procedure can no longer be used for step (5). It should nevertheless be mentioned that it is not absolutely necessary to use the method of maximum likelihood to estimate α_0 and α_1, although it is the method preferred in this paper.

4.2 Verifying the Error Normality Hypothesis

In order to use the method of maximum likelihood, we might be interested in verifying the error normality hypothesis (or rather the residual normality hypothesis since the errors are not observed). In the absence of nonresponse, a traditional method (D'Agostino 1986, page 25, equation 2.11) consists in producing the graph of $\Phi^{-1}[F_n(e_i)]$ in terms of residuals e_i, for $i \in S$, where $\Phi(.)$ is the distribution function for a random variable having the standard normal distribution, and $F_n(.)$ is the empirical distribution function. Whenever errors exhibit normal distribution, the points in this graph should more or less fall along a line having a slope $1/\sigma$ passing through the origin.

If there is nonresponse, the same strategy can be used as in the previous paragraph, but the empirical distribution function must be estimated using the respondent units. Since the units in the sample respond with unequal probabilities, the estimated empirical distribution function can be given by the formula (Särndal, Swensson and Wretman 1992, page 199):

$$F_n^*(e_i) = \frac{\sum_{j:\, j\in R \text{ and } e_j \le e_i} 1/p^*(Y_j)}{\sum_{j\in R} 1/p^*(Y_j)}.$$

Note that, in this last equation, the response probabilities are estimated as opposed to the Särndal, Swensson and Wretman formula, in which selection probabilities are known. Thus, the error normality hypothesis can be verified by producing the graph of $\Phi^{-1}[F_n^*(e_i)]$ in terms of residuals e_i, for $i \in R$. This method will be valid provided that $F_n^*(e_i)$ can correctly estimate $F_n(e_i)$, as is the case when the response probabilities are correctly estimated. When the nonresponse is not ignorable, and when the method of estimation proposed in this paper is used, the response probabilities should be properly estimated if models (3.2) and (3.3) are appropriate along with approximation (4.3).

5. Simulation Study

In order to compare the estimators of the mean presented in the two previous sections, we carried out a simulation study. We simulated 4 populations with a size of 1,000 according to model (3.2) with $\beta_0 = 2$ and $\beta_1 = 3$. Random variables X_i are independent of one another and they follow an exponential distribution of mean 1. Errors ε_i are independent of one another, are not correlated with the X_i and have a mean of zero and a variance σ^2. In two populations, the errors follow a normal distribution ($\varepsilon_i \sim \text{Nor}(0, \sigma^2)$), and in the other two populations, the errors follow an exponential distribution of mean σ recentred at $0\,(\varepsilon_i \sim \text{Exp}(\sigma) - \sigma)$. For each of these distributions, one population has a standard deviation σ equal to 1.5 corresponding to a squared coefficient of correlation (between X and Y) of 80% ($R^2 = 80\%$), and the other has a standard deviation equal to 3 corresponding to a square coefficient of correlation of 50% ($R^2 = 50\%$).

For each population, we simulated 1,000 samples of respondents according to model (3.3) with $\alpha_1 = 0.5$. Parameter α_0 was determined separately for each of the 4 populations, so that the mean response rate would be 70%. This parameter varied between -1.185 and -0.958. Note that we have here a census ($n = N = 1{,}000$). The advantage of this is that we can concentrate solely on the nonresponse error since there is no sampling error. Moreover, the fact that populations of relatively large size (1,000) are generated makes it possible to emphasize the bias of the estimators instead of their variance, since the variance should diminish as the size of the population increases (for a fixed mean response rate).

For each of the 1,000 samples of respondents, we calculated the 7 estimates of the mean described in the two previous sections. We then calculated, for each population, the mean and the variance of these 1,000 estimates, denoted $\bar{\mu}^*$ and $S^2_{\mu^*}$ respectively. Finally, we calculated an estimate of the relative bias (expressed as a percentage), $\text{RB}^* = [(\bar{\mu}^* - \mu)/\mu] \times 100\%$, an estimate of the standard error associated with this relative bias, $\text{SE}^* = (100/\mu)\sqrt{S^2_{\mu^*}/1{,}000}$, and an estimate of the root mean square errors, $\text{RMSE}^* = \sqrt{S^2_{\mu^*} + (\bar{\mu}^* - \mu)^2}$.

The results of the simulation study are shown in Table 1. An analysis of this table indicates that, regardless of the error distribution, the relative bias and the mean square error of all the estimators is lower when the correlation between X and Y is greater, which is not surprising.

Table 1
Simulation Results Used to
Compare 7 Estimators of the Mean μ

Estimator	$R^2 = 80\%$			$R^2 = 50\%$		
	RB*(%)	SE*	RMSE*	RB*(%)	SE*	RMSE*
	Population with normally distributed errors					
$\mu^*_{P,U}$	16.90	0.03	0.84	26.68	0.04	1.33
$\mu^*_{P,X}$	5.65	0.02	0.28	18.02	0.03	0.90
$\mu^*_{P,Y,ML}$	-0.14	0.03	0.05	1.27	0.10	0.17
$\mu^*_{P,Y,ROB}$	1.14	0.03	0.08	10.12	0.06	0.51
$\mu^*_{I,X}$	5.50	0.02	0.27	17.74	0.03	0.89
$\mu^*_{I,Y,ML}$	0.13	0.03	0.04	1.03	0.07	0.13
$\mu^*_{I,Y,ROB}$	0.64	0.03	0.05	7.53	0.06	0.39
	Population with exponentially distributed errors					
$\mu^*_{P,U}$	17.83	0.04	0.86	26.60	0.05	1.29
$\mu^*_{P,X}$	5.44	0.02	0.26	16.06	0.04	0.78
$\mu^*_{P,Y,ML}$	-0.54	0.02	0.04	5.18	0.05	0.26
$\mu^*_{P,Y,ROB}$	1.31	0.02	0.07	7.43	0.03	0.36
$\mu^*_{I,X}$	5.19	0.02	0.25	15.41	0.03	0.75
$\mu^*_{I,Y,ML}$	-3.42	0.03	0.17	-25.47	0.05	1.23
$\mu^*_{I,Y,ROB}$	0.49	0.02	0.04	4.07	0.03	0.20

An analysis of the relative bias indicates that the method of maximum likelihood provides best results when the errors are normally distributed, followed by the robust estimation method described in section (4.1). Estimators which assume a nonignorable response mechanism have a lower relative bias than those which incorrectly assume an ignorable response mechanism. Among the latter estimators, the most biased is estimator $\mu^*_{P,U}$. For a given method, there is generally little difference between the weighted estimator (2.1) and the estimator that includes imputed values (2.2). However, the latter must be given a slight advantage.

The conclusions in the previous paragraph always apply when errors are exponentially distributed, except that the robust estimation method becomes the best. This observation should not be surprising since the method of maximum likelihood is based on the error normality hypothesis. However, the weighted estimator $\mu^*_{P, Y, ML}$ remains slightly biased, and this is more difficult to explain.

The conclusions drawn from an analysis of the relative bias still apply when analyzing the mean square error. In fact, estimators which are very biased show a strong tendency to having a high mean square error and vice versa.

6. Discussion

When the hypothesis of a nonignorable response mechanism is realistic, and when the hypothesis of error normality for linear regression model (3.2) is justified, using the method of maximum likelihood may be appropriate. However, when the latter hypothesis is not justified, the results of the simulation study described in section 5 show that the robust estimation method presented in this paper is preferable.

Moreover, Beaumont (1999) described the results of another simulation study indicating that the estimation method proposed in this paper is robust with respect to both the error normality hypothesis and model (3.2). As for the method of maximum likelihood, it has been shown to be even more sensitive to the validity of model (3.2) than to the hypothesis of error normality. The latter method should therefore only be used when all the hypotheses associated with models (3.2) and (3.3) are reasonable.

Obviously, all estimators show little bias when non-response is very low. Likewise, when the coefficient of correlation between X and Y is very high, all estimators show little bias, except for the estimator which assumes a uniform response mechanism $\mu^*_{P, U}$. In either case, the choice of an estimator should be based on the criterion of simplicity, which favours the estimators in section (3.2), specifically estimator $\mu^*_{I, X}$.

It should be noted that models (3.2) and (3.3) could be complexified according to the nature of the problem. For example, other independent variables could be included in these models. Variable Y could also be categorized using dummy variables, and these dummy variables could be used in model (3.3) instead of variable Y itself.

In this paper, we have dealt only with the problem of the estimation of a mean when the response mechanism is not ignorable. However, the methods described in section 3 and 4 apply to other types of estimation. For example, weights or imputed values could be used for the estimation of parameters in a given regression.

This paper has attempted to describe a robust estimation method with respect to the hypothesis of error normality for model (3.2), making it possible to reduce the bias due to a nonignorable response mechanism. In some future work, it would be interesting to evaluate simple methods of variance estimation using imputed data and this robust estimation method.

Acknowledgements

The author wishes to thank the Small Area and Administrative Data Division of Statistics Canada, which made this work possible. He also wishes to thank Eric Rancourt, the two referees as well as the associate editor for some useful comments which helped improve the quality of this paper.

References

Beaumont, J.-F., and Demnati, A. (1998). Parameter estimation for a finite mixture of distribution for dichotomous longitudinal data: Comparing algorithms. *Proceedings*, *Symposium 98*, *Longitudinal Analysis for Complex Survey*, Statistics Canada, 191-197.

Beaumont, J.-F. (1999). A robust estimation method in the presence of nonignorable nonresponse. *Proceedings of the Section on Survey Research Methods*, American Statistical Association. (To appear).

D'agostino, R.B. (1986). Graphical analysis. *Goodness-of-fit Techniques*, (Ed., R.B. D'Agostino and M.A. Stephens), 7-62. New York: Marcel Dekker.

Dempster, A.P., Laird, N.M. and Rubin, R.B. (1977). Maximum likelihood from incomplete data via the EM-algorithm. *Journal of the Royal Statistical Society B*, 39, 1-38.

Gagnon, F., Lee, H., Rancourt, E. and Särndal, C.-E. (1996). Estimating the variance of the generalized regression estimator in the presence of imputation for the Generalized Estimation System. *1996 Proceedings of the Survey Methods Section*, *Statistical Society of Canada*, 151-156.

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.

Rancourt, E., Lee, H. and Särndal, C.-E. (1994). Bias corrections for survey estimates from data with ratio imputed values for confounded nonresponse. *Survey Methodology*, 20, 137-147.

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

SAS Institute Inc. (1990). *SAS/STAT User's Guide*, 2, Version 6, Fourth Edition. Cary, NC: SAS Institute Inc.

Zeger, S.L., Liang, K. and Albert, P.S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics*, 44, 1049-1060.