Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 66(3) 2012 359-380

Estimation and Prediction under Nonignorable Nonresponse via Response and Nonresponse Distributions

Abdulhakeem A.H. Eideh
[1]*Department of Mathematics, Al-Quds University, Abu-Dies Campus, Palestine P.O. Box 20002, Jerusalem*

Received 07 February 2011; Revised 26 August 2011; Accepted 31 May 2012

SUMMARY

The response distribution is the distribution of the observed outcomes given the respondent set and sample units. We study the response and nonresponse distributions under nonignorable nonresponse. We give some new results that further favor the use of the response and nonresponse distributions for analytical inference of complex survey data under nonignorable nonresponse. We derive some new relationships between moments of the population distribution before sampling and the response and nonresponse distributions. Thus provides new justification for the broad use of probability-weighted estimators (design-based school) in estimating finite population parameters in case of ignorable nonresponse. In addition to the estimation problem we introduce new predictors of the finite population total, under common mean population model, simple ratio population model, and simple regression population model. These new predictors take into account the nonignorable nonresponse. Thus, also provides new justification for the broad use of best linear unbiased predictors (model-based school) in predicting finite population parameters in case of ignorable nonresponse. The main feature of the present estimators and predictors is their behaviours in terms of the nonignorable nonresponse parameters. Furthermore, we introduce two new tests for testing the ignorability of nonresponse.

Keywords: Nonignorable nonresponse, Response propensity, Response distribution.

1. INTRODUCTION

Survey data may be viewed as the outcome of two processes: the process that generates the values of units in the finite population, often referred as the superpopulation model, and the process of selecting the sample units from the finite population, known as the sample selection mechanism. Analytic inference from sample survey data refers to the superpopulation model. For more discussion under informative sampling design and full response; see Skinner *et al*. (1989), Pfeffermann *et al*. (1998), Pfeffermann and Sverchkov (2003) and Eideh (2010).

In addition to the effect of complex sample design, one of the major problems in the analysis of survey data is the inability to obtain useful data on all questionnaire items from all members of the sample. We call this problem nonresponse or missing value problem. In short, by nonresponse (or missing value) is meant that the desired data are not obtained for the entire sample. According to Särndal *et al*. (1992, pp 563-364) strategies for dealing with nonresponse can be classified into three categories: (a) Before and during data collection, effective measures are taken to reduce the nonresponse to insignificant levels. (b) Special, perhaps costly techniques for data collection and estimation are used to permit unbiased estimation. (c) Model assumptions about the response mechanism and about relations between variables are used to construct estimators that "adjust" for a nonresponse that cannot be considered harmless. In this paper we consider the approach based on modeling.

E-mail address : msabdul@science.alquds.edu

Let $U = \{1, ..., N\}$ denote a finite population consisting of N units. Let y be the study variable of interest and let y_i be the value of y for the i^{th} population unit. Let $\mathbf{x}_i = (x_{i1}, ..., x_{ip})'$, $i \in U$ be the values of a vector of auxiliary variables, $x_1, ..., x_p$. Consider the population values $y_1, ..., y_N$ as random variables, which are independent realizations from a distribution with probability density function (pdf) $f_p(y_i|\mathbf{x}_i, \theta)$ indexed by a vector of parameters θ. A probability sample s is drawn from U according to a specified sampling design. The sample size is denoted by n. In what follows, we consider a noninformative sampling design (hence, distribution of y_i given $\mathbf{x}_i$ for $i \in s$ is the same as distribution of y_i given $\mathbf{x}_i$ for $i \in U$; that is, $f_s(y_i|\mathbf{x}_i, \theta) = f_p(y_i|\mathbf{x}_i, \theta)$) with selection probabilities $\pi_i = \Pr(i \in s)$, and sampling weight $w_i = 1/\pi_i$; $i = 1, ..., N$. Denote by $\mathbb{R} = (R_1, ..., R_N)'$ the N by 1 response indicator (vector) variable such that $R_i = 1$ if unit $i \in s$ is observed and $R_i = 0$ if unit $i \in s$ is not observed. The response set is defined accordingly as $r = \{i|i \in s, R_i = 1\}$ and the nonresponse set by $\bar{r} = \{i|i \in s, R_i = 0\}$. The size of the response set is denoted by m, so that the size of the nonresponse set is $n - m$. We assume probability sampling, so that $\pi_i = \Pr(i \in s) > 0$ for all units $i \in U$. Let the response probability

$$\psi_i = \Pr(R_i = 1|i \in s, \mathbf{x}, \mathbf{y}) = \Pr(i \in r|i \in s, \mathbf{x}, \mathbf{y}) = \Pr(i \in r|\mathbf{x}, \mathbf{y}) \quad (1)$$

for all units $i \in s$ and $\phi_i = 1/\psi_i$ be the response weight for $i \in r$. The response probabilities are theoretical quantities and they are unknown, and its value lies between 0 and 1. Furthermore, the response indicators R_i are observed for sample elements only. These response probabilities can be estimated based on the sample. By using an appropriate model based on auxiliary information $\mathbf{x}_i = (x_{i1}, ..., x_{ip})'$, for all units $i \in s$, we can compute sample-based estimates of the response probabilities, that is

$$\hat{\psi}_i = \psi(\mathbf{x}_i) = \Pr(R_i = 1 | i \in s, \mathbf{x}_i) \quad (2)$$

for $i = 1, ..., n$. We refer to $\hat{\psi}_i = \psi(\mathbf{x}_i)$ as the response propensity. The response propensity is the estimated response probability conditional on the sample and the individual characteristics $\mathbf{x}_i$. Now, we describe two methods that can be used to compute response propensities.

Method 1. Probit model

$$\hat{\psi}_i = \psi(\mathbf{x}_i) = \Phi(\mathbf{x}_i'\hat{\boldsymbol{\beta}}) = \int_{-\infty}^{\mathbf{x}_i'\boldsymbol{\beta}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \quad (3)$$

where $\hat{\boldsymbol{\beta}}$ is a p by 1 vector of coefficients, and Φ is the cumulative distribution function of the standard normal distribution.

Method 2. Logit model

$$\hat{\psi}_i = \psi(\mathbf{x}_i) = L(\mathbf{x}_i'\hat{\boldsymbol{\beta}}) = \frac{\exp(\mathbf{x}_i'\hat{\boldsymbol{\beta}})}{1 + \exp(\mathbf{x}_i'\hat{\boldsymbol{\beta}})} \quad (4)$$

The logit transformation leads to the logistic regression model:

$$\ln \frac{\psi(\mathbf{x}_i)}{1 - \psi(\mathbf{x}_i)} = \mathbf{x}_i'\hat{\boldsymbol{\beta}} = x_{i1}\beta_1 + \cdots + x_{ip}\beta_p \quad (5)$$

The probit and logit models are the most common models but in fact any model with the right property can be used. From now on, if the response probabilities are unknown, we replace them by their estimates, the response propensities.

A key issue that must be confronted when dealing with missing data or nonresponse is the relationship between the response indicator (vector) variable, the sample selection indicator membership, the study variable, and the auxiliary population variable. Little and Rubin (2002) consider three types of nonresponse mechanism or missing data mechanism:

(a) Missing completely at random (MCAR): If the response probability does not depend on the study variable, or the auxiliary population variable, the missing data are MCAR. That is,

$$\Pr(i \in r | i \in s, \mathbf{x}_i, y_i) = \Pr(i \in r | i \in s) \quad (6)$$

for all possible values y_i and $\mathbf{x}_i$.

(b) Missing at random (MAR) given auxiliary population variable: If the response probability depends on the auxiliary population variable but not on the study variable, the missing data are MAR. That is,

$$\Pr(i \in r \mid i \in s, \mathbf{x}_i, y_i) = \Pr(i \in r \mid i \in s, \mathbf{x}_i) \quad (7)$$

for all possible values y_i.

(c) Not missing at random (NMAR): If the response probability depends on the value of a missing study variable, the missing data are NMAR. That is,

$$\Pr(i \in r \mid i \in s, \mathbf{x}_i, y_i) \neq \Pr(i \in r \mid i \in s, \mathbf{x}_i) \quad (8)$$

for all possible values y_i.

In this paper we make distinction between ignorable and nonignorable response mechanism.

(a) The response mechanism can be ignored conditional on $\mathbf{x}_i$ (or ignorable nonresponse) if:

$$\Pr(i \in r \mid i \in s, \mathbf{x}_i, y_i) = \Pr(i \in r \mid i \in s, \mathbf{x}_i)$$

for all possible values y_i

(b) The response mechanism cannot be ignored (nonignorable nonresponse) if:

$$\Pr(i \in r \mid s, \mathbf{x}_i, y_i) \neq \Pr(i \in r \mid i \in s, \mathbf{x}_i)$$

for all possible values y_i.

This distinction plays an important role when dealing with likelihood theory under ignorable nonresponse and nonignorable nonresponse models.

For more discussion on nonresponse; see Little and Rubin (2002), Schafer (1997), Little (1982), Rubin (1976), Särndal and Swensson (1987), Cobben (2009), and Chambers and Skinner (2003).

The plan of this paper is as follows. In Section 2 we discuss response and nonresponse distribution. Section 3 justifies unified probability weighted estimators via method of moments in case of nonignorable nonresponse. In Section 4 we introduce 4 different models for the conditional expectations of response probabilities. Section 5, 6 and 7 are devoted to response likelihood, estimation, Fisher information, and confidence interval under nonignorable nonresponse. Section 8 discussed the prediction of finite population total under nonignorable nonresponse. In Section 9 we introduce new tests for not missing at random mechanism or nonignorable nonresponse. We conclude with a brief discussion in Section 10.

2. RESPONSE AND NONRESPONSE DISTRIBUTIONS

Before defining the response and nonresponse distribution mathematically, let us introduce the following notations: f_r and $E_r(\cdot)$ denote the pdf and the mathematical expectation of the response distribution, respectively, and $f_{\bar{r}}$ and $E_{\bar{r}}(\cdot)$ denote the pdf and the mathematical expectation of the nonresponse distribution, respectively.

Eideh (2009) defined and studies the properties of response and nonresponse distributions when the sampling design is informative and missing value mechanism is nonignorable. In this paper, from now on, we assume that the sampling design is noninformative, that is, $f_s(y_i \mid \mathbf{x}_i, \theta, \gamma) = f_p(y_i \mid \mathbf{x}_i, \theta)$. Using the results derived in Eideh (2009), we have:

(a) The (marginal) response pdf of y_i is defined as:

$$f_r(y_i \mid \mathbf{x}_i, \theta, \eta) = f_p(y_i \mid \mathbf{x}_i, \theta, \eta, R_i I_i = 1)$$

$$= \frac{\Pr(i \in r \mid \mathbf{x}_i, y_i, \eta) f_p(y_i \mid \mathbf{x}_i, \theta)}{\Pr(i \in r \mid \mathbf{x}_i, \theta, \eta)} \quad (9)$$

where θ is the parameter of the population distribution, η is the parameter indexing $\Pr(i \in r \mid \mathbf{x}_i, y_i, \eta)$ - response mechanism, and

$$\Pr(i \in r \mid \mathbf{x}_i, \theta, \eta) = \int \Pr(i \in r \mid \mathbf{x}_i, y_i, \eta)\ f_p(y_i \mid \mathbf{x}_i, \theta) dy_i$$

Note that response pdf contains the population parameter, θ, that indexes, $f_p(y_i \mid \mathbf{x}_i, \theta)$, and the nonignorable nonresponse parameter, η, that indexes, $\Pr(i \in r \mid \mathbf{x}_i, y_i, \eta)$. Thus, the response pdf may contain more parameters than the population pdf. In addition to that, the (marginal) response pdf is different from the population pdf generating the finite population values, unless $\Pr(i \in r \mid \mathbf{x}_i, y_i, \eta = \Pr(i \in r \mid \mathbf{x}_i, \eta)$ for all possible values y_i, that is R_i and y_i are stochastically independent, in which case the response mechanism can be ignored conditional on $\mathbf{x}_i$. Also note that the marginal response distribution is a function of the population distribution and of the probability of responses or propensity score.

(b) If the response value mechanism is ignorable, that is

$$\Pr(i \in r \mid \mathbf{x}_i, y_i, \eta) = \Pr(i \in r \mid \mathbf{x}_i, \theta, \eta)$$

then

$$f_r(y_i \mid \mathbf{x}_i, \theta, \eta) = f_p(y_i \mid \mathbf{x}_i, \theta) \quad (10)$$

(c) The (marginal) nonresponse pdf of is defined as:

$$f_{\bar{r}}(y_i \mid \mathbf{x}_i, \theta, \eta) = f_p(y_i \mid \mathbf{x}_i, \theta, \eta, R_i = 0, I_i = 1)$$

$$= \frac{\Pr(i \in \bar{r} \mid \mathbf{x}_i, y_i, \eta) f_p(y_i \mid \mathbf{x}_i, \theta)}{\Pr(i \in \bar{r} \mid \mathbf{x}_i, \theta, \eta)} \quad (11)$$

Corollary 1.

$$\Pr(R_i = 1 \mid \mathbf{x}_i, y_i) = E_p(\psi_i \mid \mathbf{x}_i y_i)$$

Corollary 2. Alternative representation of the marginal response and nonresponse pdfs of y_i are given by:

$$f_r(y_i \mid \mathbf{x}_i) = \frac{E_p(\psi_i \mid \mathbf{x}_i, y_i) \; f_p(y_i \mid \mathbf{x}_i)}{E_p(\psi_i \mid \mathbf{x}_i)} \quad (12)$$

$$f_{\bar{r}}(y_i \mid \mathbf{x}_i) = \frac{\{1 - E_p(\psi_i \mid \mathbf{x}_i, y_i)\} \; f_p(y_i \mid \mathbf{x}_i)}{\{1 - E_p(\psi_i \mid \mathbf{x}_i)\}} \quad (13)$$

where

$$E_p(\psi_i \mid \mathbf{x}_i) = \int E_p(\psi_i \mid \mathbf{x}_i, y_i) f_p(y_i \mid \mathbf{x}_i) \, dy_i$$
$$= E_p\{E_p(\psi_i \mid \mathbf{x}_i, y_i)\}$$

Thus, given $f_p(y_i \mid \mathbf{x}_i)$, then $f_r(y_i \mid \mathbf{x}_i)$ and $f_{\bar{r}}(y_i \mid \mathbf{x}_i)$ are determined by specifying $E_p(\psi_i \mid \mathbf{x}_i, y_i)$.

Corollary 3. Let $\phi_i = 1/\psi_i$ be the response weight for $i \in r$. For vector of random variables $(y_i, \mathbf{x}_i)$, the following relationships hold:

$$E_r(\phi_i \mid y_i) = \{E_p(\psi_i \mid y_i)\}^{-1} \quad (14a)$$

$$E_p(y_i \mid x_i) = \{E_r(\phi_i \mid x_i)\}^{-1} E_r(\phi_i y_i \mid x_i) \quad (14b)$$

$$E_p(y_i) = \{E_r(\phi_i)\}^{-1} E_r(\phi_i y_i) \quad (14c)$$

$$E_r(\phi_i) = \{E_p(\psi_i)\}^{-1} \quad (14d)$$

$$E_{\bar{r}}(y_i \mid \mathbf{x}_i) = \frac{E_p\{(1-\psi_i) y_i \mid \mathbf{x}_i\}}{E_p\{(1-\psi_i) \mid \mathbf{x}_i\}}$$

$$= \frac{E_r\{(\phi_i - 1) y_i \mid \mathbf{x}_i\}}{E_r\{(\phi_i - 1) \mid \mathbf{x}_i\}} \quad (14e)$$

It should be emphasized that, in this paper, the proposed approach is model based.

Comment 0. Since we are assuming noninformative sampling design, therefore the distribution of y_i given $\mathbf{x}_i$ before sampling is the same as the distribution of y_i given $\mathbf{x}_i$ after sampling, hence all population moments before sampling and after sampling are the same. So that, when fitting models to survey data, the sampling weights are disappeared from all the formulas in the paper. This indicates that, when the sampling design is noninformative, we do not need to take into account the sampling weights in the analysis of survey data. I thing that the results obtained in this paper will raise the concept: the role of sampling weights when the sampling design is noninformative (or ignorable) and the missing data mechanism is not missing at random (or informative or nonignorable). We will leave this issue for future research.

3. UNIFIED PROBABILITY WEIGHTED ESTIMATORS UNDER NONIGNORABLE NONRESPONSE

In this section, we derive known results in probability sampling theory from the relationships given in Sections 1, 2 and 3. Also we prove that probability weighted estimator, in case of nonresponse, is just the method of moments estimator based the response and nonresponse distributions. So provides new justification for the broad use of probability-weighted estimators (design-based school) in estimating finite population parameters in case of ignorable nonresponse.

3.1 Estimation of Finite Population Mean and Finite Population Variance under Noninformative Sampling Design and Nonignorable Nonresponse

Let $y_1, \ldots, y_N$ be N independent and identically distributed random variable from a population with finite mean $E_p(y_i) = \mu$ and finite variance $Var_p(y_i) = \sigma^2$.

The method of moments estimates (MME) of μ and σ^2 are the solutions of the method of moments equations:

$$E_p(y_i) = \mu = \frac{1}{N}\sum_{i=1}^{N} y_i = \bar{Y}_U = \frac{1}{N}\sum_{i \in U} y_i$$

and

$$E_p\left(y_i^2\right) = \sigma^2 + \mu^2 = N^{-1}\sum_{i=1}^{N} y_i^2$$

which are:

$$\tilde{\mu} = N^{-1}\sum_{i=1}^{N} y_i = \bar{Y}_U \quad \text{and} \quad \tilde{\sigma}^2 = N^{-1}\sum_{i=1}^{N} y_i^2 - \bar{Y}_U^2$$

But $\sum_{i\in U} y_i$ and $\sum_{i\in U} y_i^2$ are unknown finite population parameters that need estimation.

Now using (14c), we have

$$E_p(y_i) = \mu = \frac{E_r(\phi_i y_i)}{E_r(\phi_i)} = \frac{\sum_{i\in r}\phi_i y_i}{\sum_{i\in r}\phi_i}$$

and

$$E_p\left(y_i^2\right) = \tilde{\sigma}^2 + \tilde{\mu}^2 = \frac{E_r\left(\phi_i y_i^2\right)}{E_r(\phi_i)} = \frac{\sum_{i\in r}\phi_i y_i^2}{\sum_{i\in r}\phi_i}$$

Therefore

$$\tilde{\mu} = \frac{\sum_{i\in r}\phi_i y_i}{\sum_{i\in r}\phi_i} = \bar{y}_\phi$$

and

$$\tilde{\sigma}_\phi^2 = \frac{\sum_{i\in r}\phi_i y_i^2}{\sum_{i\in r}\phi_i} - \left[\frac{\sum_{i\in r}\phi_i y_i}{\sum_{i\in r}\phi_i}\right]^2$$

$$= \frac{\sum_{i\in r}\phi_i\left(y_i - \tilde{\mu}\right)^2}{\sum_{i\in r}\phi_i} = \frac{\sum_{i\in r}\phi_i\left(y_i - \bar{y}_\phi\right)^2}{\sum_{i\in r}\phi_i}$$

Hence, the MM estimators of $\bar{Y}_U$ and $S_U^2 = N^{-1}\sum_{i\in U}\left(y_i - \bar{Y}\right)^2$ under nonignorable nonresponse are

$$\tilde{\bar{Y}}_U = \bar{y}_w = \frac{\sum_{i\in r}\phi_i y_i}{\sum_{i\in r}\phi_i} = \bar{y}_\phi \tag{15a}$$

and

$$\tilde{S}_U^2 = \frac{\sum_{i\in r}\phi_i\left(y_i - \bar{y}_\phi\right)^2}{\sum_{i\in r}\phi_i} \tag{15b}$$

which are the well known probability weighted estimator.

Similarly, we can show that, the MME of $\sum_{i\in U} y_i^k = NE_p\left(y_i^k\right)$ is $\sum_{i\in r}\phi_i y_i^k$.

If the nonresponse mechanism is ignorable: In this case, $E_p(y_i) = E_r(y_i)$, so that the MME of $\bar{Y}_U = N^{-1}\sum_{i\in U} y_i$ is given by

$$\hat{\hat{\bar{Y}}}_{UR} = \frac{\sum_{i\in r} y_i}{\sum_{i\in r} 1}$$

3.2 Estimation of the Multiple Linear Regression Parameters under Noninformative Sampling Design and Nonignorable Nonresponse

Let $y_i, x_{i1}, ..., x_{iq}$; $i = 1, ..., N$ be $(q + 1)$ random variables such that

$$E_p\left(y_i|\mathbf{x}_i\right) = \mathbf{x}_i'\boldsymbol{\beta} = x_{i1}\beta_1 + \cdots + x_{iq}\beta_q$$

and

$$V_p(y_i|\mathbf{x}_i) = \sigma^2;\ i = 1, ..., N$$

Then

$$\mathbf{x}_i E_p\left(y_i|\mathbf{x}_i\right) = \mathbf{x}_i\mathbf{x}_i'\boldsymbol{\beta}$$

and

$$E_p\{\mathbf{x}_i E_p(y_i|\mathbf{x}_i)\} = E_p\{E_p(\mathbf{x}_i y_i|\mathbf{x}_i)\}$$
$$= E_p\left(\mathbf{x}_i y_i\right) = E_p\left(\mathbf{x}_i\mathbf{x}_i'\right)\boldsymbol{\beta}$$

Solving this equation for β, we get

$$\boldsymbol{\beta} = \left\{E_p\left(\mathbf{x}_i\mathbf{x}_i'\right)\right\}^{-1} E_p\left(\mathbf{x}_i y_i\right).$$

The MME of β is

$$\boldsymbol{\beta} = \left\{E_p\left(\mathbf{x}_i\mathbf{x}_i'\right)\right\}^{-1} E_p\left(\mathbf{x}_i y_i\right) = \left\{\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i'\right\}^{-1}\sum_{i=1}^{N}\mathbf{x}_i y_i$$

Now using equation (14c), we have

$$\boldsymbol{\beta} = \left\{E_p\left(\mathbf{x}_i\mathbf{x}_i'\right)\right\}^{-1} E_p\left(\mathbf{x}_i y_i\right) = \left\{E_r\left(\boldsymbol{\varphi}_i\mathbf{x}_i\mathbf{x}_i'\right)\right\}^{-1} E_r\left(\boldsymbol{\varphi}_i\mathbf{x}_i y_i\right)$$

where $\boldsymbol{\varphi}_i = (0, ..., \phi_i, ..., 0)'$.

Thus MME of β is given by:

$$\tilde{\beta} = \left\{ \sum_{i \in r} \varphi_i \, \mathbf{x}_i \, \mathbf{x}_i' \right\}^{-1} \sum_{i \in r} \varphi_i \, \mathbf{x}_i \, y_i = \left(\mathbf{x}' \varphi \mathbf{x} \right)^{-1} \left(\mathbf{x} \varphi \mathbf{y} \right) \quad (16)$$

which is the probability weighted estimator of β.

If the nonresponse mechanism is ignorable: In this case, $E_p(y_i) = E_r(y_i)$, so that the MME of β is given by

$$\tilde{\beta} = \left\{ \sum_{i \in r} \mathbf{x}_i \, \mathbf{x}_i' \right\}^{-1} \sum_{i \in r} \mathbf{x}_i \, y_i = \left(\mathbf{x}' \mathbf{x} \right)^{-1} \left(\mathbf{xy} \right)$$

In particular if $E_p(y_i|x_i) = \beta_0 + \beta_1 x_i$; $i = 1, ..., N$ and $V_p(y_i) = \sigma^2$, then we have

$$\tilde{\beta}_{1r} = \frac{\sum_{i \in r} \phi_i y_i x_i - \sum_{i \in r} \phi_i \bar{x}_\phi \bar{y}_\phi}{\sum_{i \in r} \phi_i x_i^2 - \sum_{i \in r} \phi_i \bar{x}_\phi^2} = \frac{\sum_{i \in r} \phi_i \left(x_i - \bar{x}_\phi \right)\left(y_i - \bar{y}_\phi \right)}{\sum_{i \in r} \phi_i \left(x_i - \bar{x}_\phi \right)^2} \quad (17a)$$

and

$$\tilde{\beta}_{0r} = \bar{y}_\phi - \tilde{\beta}_{1r} \bar{x}_\phi \quad (17b)$$

where $\bar{x}_\phi = \sum_{i \in r} \phi_i x_i \Big/ \sum_{i \in r} \phi_i$, $\bar{y}_\phi = \sum_{i \in r} \phi_i y_i \Big/ \sum_{i \in r} \phi_i$.

3.3 Method of Moments Estimator of Census Log-Likelihood under Noninformative Sampling Design and Nonignorable Nonresponse

Let $y_1, ..., y_N$ be a random sample from $y \sim f_p(y;\, \theta)$. The census log-likelihood is: $l(\theta) = \sum_{i \in U} \log f_p\left(y_i;\, \theta\right)$. The census MLE of θ is the solution of the census log-likelihood equation: $U(\theta) = \partial l(\theta)/\partial\theta = \sum_{i \in U} \partial \log f_p\left(y_i;\, \theta\right) \big/ \partial\theta = 0$. The MM estimator of $U(\theta)$ is the solution of

$$E_p \left\{ \frac{\partial \log f_p\left(y_i;\, \theta\right)}{\partial \theta} \right\} = \frac{1}{N} \sum_{i=1}^{N} \frac{\partial \log f_p\left(y_i;\, \theta\right)}{\partial \theta} = \frac{U(\theta)}{N}$$

But using equation (14c), we have

$$E_p \left\{ \frac{\partial \log f_p\left(y_i;\, \theta\right)}{\partial \theta} \right\} = \left\{ E_r\left(\phi_i\right) \right\}^{-1} E_r \left\{ w_i \frac{\partial \log f_p\left(y_i;\, \theta\right)}{\partial \theta} \right\}$$

So that the method of moment estimate of satisfies

$$\frac{1}{\sum_{i \in r} \phi_i} \sum_{i \in r} \phi_i \left\{ \frac{\partial \log f_p\left(y_i;\, \theta\right)}{\partial \theta} \right\} = \frac{1}{N} \sum_{i=1}^{N} \frac{\partial \log f_p\left(y_i;\, \theta\right)}{\partial \theta}$$

Hence the MME of $U(\theta)$ is

$$\hat{U}(\theta) = N \frac{1}{\sum_{i \in r} \phi_i} \sum_{i \in r} \phi_i \left\{ \frac{\partial \log f_p\left(y_i;\, \theta\right)}{\partial \theta} \right\}$$

Thus the PML estimator of θ satisfies

$$\hat{U}(\theta) = \sum_{i \in r} \phi_i \left\{ \frac{\partial \log f_p\left(y_i;\, \theta\right)}{\partial \theta} \right\} = 0 \quad (18a)$$

Now, assume that $y_1, ..., y_N$ be a random sample from $y \sim f_p(y\,|\,x,\, \theta)$. The census log-likelihood is $l(\theta) = \sum_{i=1}^{N} \log f_p\left(y_i \middle| x_i,\, \theta\right)$. The census MLE of θ is the solution of the census log-likelihood equation

$$U(\theta) = \frac{\partial l(\theta)}{\partial \theta} = \sum_{i=1}^{N} \partial \log f_p\left(y_i \middle| x_i,\, \theta\right) \big/ \partial\theta = 0$$

Similarly, the PML estimator of θ satisfies

$$\hat{U}(\theta) = \sum_{i \in r} q_i \left\{ \frac{\partial \log f_p\left(y_i;\, \theta\right)}{\partial \theta} \right\} = 0 \quad (18b)$$

where $q_i = \dfrac{\phi_i}{E_r\left(\phi_i \middle| x_i\right)}$.

4. MODELING THE CONDITIONAL EXPECTATIONS OF RESPONSE PROBABILITIES

According to equation (12), for a given population distribution, the response distribution is completely determined by the specification of the conditional expectations of response probabilities, $E_p(\psi_i|y_i,\, \mathbf{x}_i,\, \eta)$. So in order to obtain the response pdf of y_i, we need to model these population conditional expectations. We consider the following four models for this population conditional expectation.

I. Exponential model

Suppose that the response probabilities have conditional expectations

$$E_p(\psi_i\,|\,y_i,\, \mathbf{x}_i) = \Pr(i \in r\,|\,y_i,\, \mathbf{x}_i,\, i \in s)$$
$$= \exp(a_0 + a_1 y_1 + h_1(\mathbf{x}_i)) \quad (19)$$

for some function $h_1(\mathbf{x})$, where $\{a_j,\, j = 0,\, 1\}$ are unknown parameters to be estimated from the respondent set.

Comment 1. Under (19), the marginal effect of y_i on $E_p(\psi_i | y_i, \mathbf{x}_i)$ is given by:

$$\frac{\partial E_p(\psi_i | y_i, \mathbf{x}_i)}{\partial y_i} = a_1 E_p(\psi_i | y_i, \mathbf{x}_i)$$

1. If $a_1 = 0$, then $E_p(\psi_i | y_i, \mathbf{x}_i)$ does not depend on y_i, so that the missing value mechanism is ignorable.
2. If $a_1 > 0$, then $E_p(\psi_i | y_i, \mathbf{x}_i)$ is an increasing function of y_i, so that larger values are more likely to be in the response set than smaller values.
3. If $a_1 < 0$, then $E_p(\psi_i | y_i, \mathbf{x}_i)$ is a decreasing function of y_i, so that smaller values are more likely to be in the response set than larger values.

II. Linear model

Suppose that the response probabilities have conditional expectations

$$\begin{aligned} E_p(\psi_i | y_i, \mathbf{x}_i) &= \Pr(i \in r | y_i, \mathbf{x}_i, i \in s) \\ &= (b_0 + b_1 y_i + h_2(\mathbf{x}_i)) \end{aligned} \quad (20)$$

for some function $h_2(\mathbf{x})$, where $\{b_j, j = 0, 1\}$ are unknown parameters to be estimated from the respondent set.

Comment 2. Under (20), the marginal effect of y_i on $E_p(\psi_i | y_i, \mathbf{x}_i)$ is given by

$$\frac{\partial E_p(\psi_i | y_i, \mathbf{x}_i)}{\partial y_i} = b_1$$

Similarly to the situation under exponential model, if $b_1 = 0$, so that the missing value mechanism is ignorable, if $b_1 > 0$ larger values are more likely to be in the response set than smaller values and vice versa if $b_1 < 0$.

Other standard ways of modeling response probabilities are obtained by the spirit of generalized linear models, via the logit and probit models.

III. Logit model

Suppose that the response probabilities have conditional expectations

$$\begin{aligned} E_p(\psi_i | y_i, \mathbf{x}_i) &= \Pr(i \in r | y_i, \mathbf{x}_i \cdot i \in s) \\ &= \frac{\exp(c_0 + c_1 y_i + h_3(\mathbf{x}_i))}{1 + \exp(c_0 + c_1 y_i + h_3(\mathbf{x}_i))} \end{aligned} \quad (21)$$

for some function $h_3(\mathbf{x})$, where $\{c_j, j = 0, 1\}$ are unknown parameters to be estimated from the response set.

IV. Probit model

Suppose that the response probabilities have conditional expectations

$$\begin{aligned} E_p(\psi_i | y_i, \mathbf{x}_i) &= \Pr(i \in r | y_i, \mathbf{x}_i \cdot i \in s) \\ &= \Phi(d_0 + d_1 y_i + h_4(\mathbf{x}_i)) \end{aligned} \quad (22)$$

where Φ denotes the cumulative distribution function of the standard normal distribution, for some function $h_4(\mathbf{x})$ and $\{d_j, j = 0, 1\}$ and are unknown parameters to be estimated from the response set.

An important aspect of the use of the response model for statistical inference is its sensitivity to wrong specification of the conditional expectations of response probabilities, $E_p(\psi_i | y_i, \mathbf{x}_i)$. This issue is investigated as follows: assume that response probabilities is based on the one of the models (exponential, linear, logit or probit) and then apply first the correct model and then the incorrect models for estimation, and compare the results obtained under the correct and incorrect models. In case of full response Eideh and Nathan (2006, Section 6.4) discussed this issue and found that: the estimators obtained by maximizing the sample log-likelihood functions based on the exponential and linear inclusion probability models, are very robust with respect to model assumptions and in fact there is no real difference between them.

As an illustration, in this paper we shall consider only the exponential model (19). From now on, we use the terms *'exponential response probabilities'* to denote that the conditional expectation of the response probabilities is an exponential function of the response variable and the available auxiliary variables, i.e. equation (19). Furthermore, as pointed out by Skinner (1994), this exponential approximation model for first order inclusion probabilities is appealing in the common situation where the sample selection is carried out in several stages so that the ultimate inclusion probabilities are the product of the selection probabilities at the various stages.

Now we have the following theorem, which gives the response pdf and their moments under the exponential response probabilities. In the following, we suppress the notation relating to the dependence of the response and population pdf's on the unknown parameters.

Theorem 1. Under the exponential response probabilities, we have

The response pdf y_i of y_i is given by

$$f_r(y_i|\mathbf{x}_i) = \frac{\exp(a_1 y_i) f_p(y_i|\mathbf{x}_i)}{M_p(a_1)} \tag{23}$$

where $M_p(a_1) = E_p(\exp(a_1 y_i)|\mathbf{x}_i$ is the moment generating function (mgf) of the population pdf of y_i.

Also the mgf and the mean of the sample pdf of y_i are given by

$$M_r(t) = \frac{M_p(t+a_1)}{M_p(a_1)} \tag{24a}$$

and

$$E_r(y_i|\mathbf{x}_i) = \frac{\partial \log M_p(a_1)}{\partial a_1} \tag{24b}$$

Proof: Using the definition of mgf and equation (12).

The following theorem gives the response distributions of the exponential family of distributions under the exponential response probabilities.

Theorem 2. Let the population distribution be a member of the exponential family of distributions

$$f_p(y_i|\mathbf{x}_i) = \exp\left(\frac{y_i\theta - h(\theta)}{\phi} + c(y_i,\, \phi)\right) \tag{25}$$

where $h(\cdot)$ and $c(\cdot)$ are known functions. The parameter θ is known as the natural parameter. Assume that $\theta = g(\mathbf{x}_i'\boldsymbol{\beta})$ where $g(\cdot)$ is a known increasing differentiable function and $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)'$ is a vector of parameter.

Under the exponential response probabilities, we have

$$f_r(y_i|\mathbf{x}_i) = \exp\left(\frac{y_i(a_1\phi+\theta) - h(a\phi_1+\theta)}{\phi} + c(y_i,\, \phi)\right) \tag{26}$$

The moment generating function of y_i is given by

$$M_r(t) = \exp\left(\frac{h(t\phi + a_1\phi + \theta) - h(\theta + a_1\phi)}{\phi}\right) \tag{27}$$

and

$$E_r(y_i|\mathbf{x}_i) = \frac{\partial h(a_1\phi+\theta)}{\partial a_1} \tag{28}$$

Proof. Using Theorem 1.

5. RESPONSE LIKELIHOOD AND ESTIMATION

Having derived the response distribution, and if the response measurements are independent, then the response likelihood for θ and η is given by

$$\begin{aligned} L_r(\theta,\, \eta) &= \prod_{i=1}^{m} f_r(y_i|\mathbf{x}_i,\, \theta,\, \eta) \\ &= \prod_{i=1}^{m} \frac{E_p(\psi_i|\mathbf{x}_i,\, y_i,\, \eta)}{E_p(\psi_i|\mathbf{x}_i,\, \theta,\, \eta)} f_p(y_i|\mathbf{x}_i,\, \theta) \end{aligned}$$

and the logarithm of the response likelihood for θ and η is

$$\begin{aligned} l_r(\theta,\, \eta) &= \sum_{i=1}^{m} \log f_r(y_i|\mathbf{x}_i,\, \theta,\, \eta) \\ &= l_{ign}(\theta) + \sum_{i=1}^{m} \log E_p(\psi_i|\mathbf{x}_i, y_i, \eta) \\ &\quad - \sum_{i=1}^{m} \log E_p(\psi_i|\mathbf{x}_i,\, \theta,\, \eta) \end{aligned} \tag{29}$$

where

$$l_{ign}(\theta) = \sum_{i=1}^{m} \log(f_p(y_i|\mathbf{x}_i, \theta)) \tag{30}$$

is the classical log-likelihood obtained under ignorable nonresponse.

The function given in equation (29) can be maximized with respect to θ and η to obtain the maximum response likelihood estimates of these parameters. Maximum response likelihood estimators of other parameters, which are the parameters of interest, (e.g. the parameter θ characterizing the population distribution of y) are defined using the invariance properties of the maximum likelihood (ML) approach.

The response likelihood function, $L_r(\theta, \eta)$, can be interpreted as a weighted likelihood, where the weights are ratios of the population conditional expectations of the response probabilities, given the values of y_i, and their unconditional expectations.

Standard estimation processes consider the case where the missing value mechanism is ignored and base the inference on the classical log-likelihood function, $l_{ign}(\theta)$. However, analysis using standard estimation methods, which ignores the last two terms of (29), leads to inconsistent estimates of θ. Thus the effect of the missing value mechanism must be taken into account.

Theorem 3. The joint response pdf of $\mathbf{y} = (y_1, ..., y_m)$ is defined by

$$f_r(\mathbf{y}) = f_p(y_1, ..., y_m|r) = \frac{\Pr(r|y_1, ..., y_n)}{\Pr(r)} f_p(y_1, ..., y_m)$$

If

$$\Pr(r|\mathbf{y}) = \exp(a_0 + \mathbf{a}'\mathbf{y}) = \prod_{i=1}^{m} \exp(a_0 + a_i y_i)$$

Then

$$f_r(\mathbf{y}) = \prod_{i=1}^{m} f_r(y_i) \tag{31}$$

where

$$f_r(y_i) = \frac{\exp(a_i y_i) f_p(y_i)}{E_p(\exp(a_i y_i))}$$

Thus under exponential response probabilities, the sample measurements are independent.

Comment 3. The response likelihood function given in (29) is a sum of two components, the first component is the classical likelihood function which ignores the missing value mechanism and just treats the response values of y as independent draws from the population distribution of y, while the second component reflects the effect of the mechanism of missing values.

Example 1. (Maximum likelihood Estimator of μ-Normal Population)

Assume $y_1, ..., y_N \underset{p}{\sim} N(\mu, 1)$ are independent. Let $y_1, ..., y_n$ be a sample of size n selected under noninformative sampling design. Assume that the observed outcomes set is $\{y_1, ..., y_m\}$ and the missing values set is $\{y_{m+1}, ..., y_n\}$. If we assume that $E_p(\psi_i|y_i) = \exp(\eta_0 + \eta_1 y_i)$, then it is easy to verify that $y_i \underset{r}{\sim} N(\mu + \eta_1, 1)$ are independents. In this case the parameters of the response pdf contains the parameters of the population pdf, μ, and the nonignorable nonresponse parameter, η_1. So that the MLE of μ and η_1 is the solution of the response likelihood equations

$$\begin{aligned} \frac{\partial l_r(\theta, \eta_1)}{\partial \theta} &= \sum_{i \in r} (y_i - \hat{\mu} - \hat{\eta}_1) = 0 \\ \frac{\partial l_r(\theta, \eta_1)}{\partial \eta_1} &= \sum_{i \in r} (y_i - \hat{\mu} - \hat{\eta}_1) = 0 \end{aligned} \tag{32}$$

Solving this system of response likelihood equations for μ and η_1 gives: $\hat{\mu}_r = \overline{y}_r - \hat{\eta}_1$.

Since we have two unknowns and one equation, we have infinitely many solutions. Hence the parameters are not identifiable from the response observations of y alone. (A model is said to be nonidentifiable if it contains parameters that cannot be estimated uniquely, or, to put in another way, that have standard errors of infinity). The identifiability problem occurs here because we have only one sufficient statistic, which is $\overline{y}$, for two parameters. To solve this problem we consider two-step estimation, see below. For illustration, assume that the nonignorable nonresponse parameter is known, say $\eta_1 = \eta_1^0$, so that if this is the case, the MLE of μ is $\hat{\mu}_r = \overline{y}_r - \hat{\eta}_1^0 \sigma_0^2$. Here $\hat{\mu}_r$ underestimates the true value of μ if $\eta_1^0 > 0$ and overestimates the true value of μ if $\eta_1^0 < 0$.

To solve this problem a two-step estimation method is adopted. Based on the response data $\{y_i, \mathbf{x}_i, \phi_i; i \in r\}$ we can estimate the parameters of the population model in two steps:

Step-one: Estimate the nonignorable nonresponse parameters η using the following relationship

$$E_r(\phi_i|\mathbf{x}_i, y_i, \eta) = \frac{1}{E_p(\psi_i|\mathbf{x}_i, y_i, \eta)} \tag{33}$$

Thus the nonignorable nonresponse parameters can be estimated using regression analysis. Denoting the resulting estimate of η by $\tilde{\eta}$.

Step-two: Substitute $\tilde{\eta}$ in the response log-likelihood function, (29), and then maximize the resulting response log-likelihood function with respect to the population parameters, θ:

$$l_r(\theta, \tilde{\eta}) = \sum_{i=1}^{m} \log f_r(y_i|\mathbf{x}_i, \theta, \tilde{\eta})$$

$$= l_{ign}(\theta) + \sum_{i=1}^{m} \log E_p(\psi_i|\mathbf{x}_i, y_i, \tilde{\eta})$$

$$- \sum_{i=1}^{m} \log E_p(\psi_i|\mathbf{x}_i, \theta, \tilde{\eta}) \quad (34)$$

where $l_r(\theta, \tilde{\eta})$ is the response log-likelihood after substituting $\tilde{\eta}$ in the response log-likelihood function, (29).

But the second component of this response log-likelihood function does not contain θ, so we can just maximize

$$l_r(\theta, \tilde{\eta}) = l_{ign}(\theta) - \sum_{i=1}^{m} \log E_p(\psi_i|\mathbf{x}_i, \theta, \tilde{\eta})$$
$$= l_{ign}(\theta) + \sum_{i=1}^{m} \log E_r(\phi_i|\mathbf{x}_i, \theta, \tilde{\eta}) \quad (35)$$

Example 2. (Maximum likelihood Estimators of μ and σ^2 - Normal Population)

Assume $y_1, ..., y_N \underset{p}{\sim} N(\mu, \sigma^2)$ are independent. Let $y_1, \ldots, y_n$ be a sample of size n selected under noninformative sampling design. Assume that the observed outcomes set is $\{y_1, ..., y_m\}$ and the missing values set is $\{y_{m+1}, ..., y_n\}$. If we assume that $E_p(\psi_i|y_i) = \exp(\eta_0 + \eta_1 y_i)$, then it is easy to verify that $y_i \underset{r}{\sim} N(\mu + \eta_1\sigma^2, \sigma^2)$ are independents.

Step 1: Estimation of nonignorable nonresponse parameters. Using (33), we get

$$E_r(\phi_i|\mathbf{x}_i, y_i, \eta_0, \eta_1) = \frac{1}{\exp(\eta_0 + \eta_1 y_1)}$$

$$= \exp(-\eta_0 - \eta_1 y_1) \quad (36)$$

So that, approximately, the least squares estimators of η_0 and η_1 are given by

$$\tilde{\eta}_1 = -\left(\sum_{i=1}^{m}(y_i - \bar{y})^2\right)^{-1}\left(\sum_{i=1}^{m}(y_i - \bar{y})(\Phi_i - \bar{\Phi})\right) \quad (35a)$$

and $$\tilde{\eta}_0 = -(\bar{\Phi} - \tilde{b}_1\bar{y}_1) \quad (35b)$$

where $\Phi_i = \ln \phi_i$.

Step 2: Estimation of the population parameter, μ and σ^2. According to (34), the response log-likelihood function to be maximized is given by

$$l_r(\mu, \sigma^2; \tilde{\eta}_0, \tilde{\eta}_1) = l_r(\mu, \sigma^2; \tilde{\eta}_1)$$

$$= -\frac{m}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i\in r}(y_i - \mu - \tilde{\eta}_1\sigma^2)^2 \quad (36)$$

Now differentiating (36) with respect to μ and σ^2 and equating it to zero, we get

$$\hat{\mu}_{nign} = \hat{\mu}_r = \bar{y}_r - \tilde{\eta}_1 s^2_{nign} = \frac{1}{m}\sum_{i\in r} y_i - \tilde{\eta}_1 s^2_{nign} \quad (37)$$

and

$$s_r^2 = s^2_{nign} = \frac{1}{m}\sum_{i\in r}(y_i - \bar{y}_r)^2 \quad (38)$$

Again, if $\tilde{\eta}_1 = 0$, that is the missing value mechanism is estimated ignorable then

$$\hat{\mu}_{nign} = \hat{\mu}_r = \bar{y}_r \text{ and } s_r^2 = s^2_{nign} = \frac{1}{m}\sum_{i\in r}(y_i - \bar{y}_r)^2$$

Let us now consider the following theorem which is related to the effect of the normalizing factor on the estimation process, when modeling the population conditional expectation of the response probabilities, given the outcome variable and possibly auxiliary variables.

Theorem 4. Under the two-step estimation method. If $E_p(\psi_i|y_i) = k_e \exp(\eta_0 + \eta_1 y_i)$, $\eta_0, \eta_1 \neq 0$ where k_e is some constant, then

$$l_r(\theta, \tilde{\eta}_0, \tilde{\eta}_1) = l_{ign}(\theta) - n\log M_p(\tilde{\eta}_1)$$

Proof:

1. $E_p(\psi_i|y_i) = k_e \exp(\eta_0 + \eta_1 y_i)$, can be written as $E_p(\pi_i|y_i) = \exp(\eta_0^* + \eta_1 y_i)$, where $\eta_0^* = \eta_0 + \log(k_e)$. So that $E_p(\psi_i) = \exp(\eta_0^*)\, M_p(\eta_1)$ and $\log E_p(\psi_i) = \eta_0^* + \log M_p(\eta_1)$. Hence the estimated response

likelihood, using estimates of the nonignorable nonresponse parameters η_0^* and η_1 is given by

$$l_r(\theta, \tilde{\eta}_0, \tilde{\eta}_1) = l_{ign}(\theta) - n\log M_p(\tilde{\eta}_1)$$
$$= l_{ign}(\theta) - (\tilde{\eta}_0 + \log k_e) - n\log M_p(\tilde{\eta}_1)$$

Now since $(\tilde{\eta}_0 + \log k_e)$ does not depend on the parameters of the population distribution, therefore it just can be omitted from the response likelihood. Thus the estimated response likelihood is

$$l_r(\theta, \tilde{\eta}_0, \tilde{\eta}_1) = l_{ign}(\theta) - n\log M_p(\tilde{\eta}_1)$$

which is free of the normalized factor k_e. Also the estimate of the nonignorable nonresponse parameters, η_1, is not affected by the estimate of the parameter, η_0^*.

Corollary 4: Using Corollaries 2 and 3, the response pdf $y_i|\mathbf{x}_i$ can be written as

$$f_r(y_i|\mathbf{x}_i) = \frac{E_r(\phi_i|\mathbf{x}_i) f_p(y_i|\mathbf{x}_i)}{E_r(\phi_i|y_i, \mathbf{x}_i)} \quad (39)$$

The application of this response pdf requires the estimation of the conditional expectations $E_r(\phi_i|y_i, \mathbf{x}_i)$ and $E_r(\phi_i|\mathbf{x}_i)$. These conditional expectations can be estimated from the respondent data set, as follows:

1. Estimate $E_r(\phi_i|y_i, \mathbf{x}_i)$ by regressing ϕ_i against $(y_i, \mathbf{x}_i)$, $i \in r$.
2. Estimate $E_r(\phi_i|\mathbf{x}_i)$ in two steps:

Step-one: Using Corollaries 2 and 3 and the estimate of $E_r(\phi_i|y_i, \mathbf{x}_i)$ obtained above, we can estimate $E_p(\psi_i|\mathbf{x}_i)$ as follows:

$$E_p(\psi_i|\mathbf{x}_i) = \int E_p(\psi_i|y_i, \mathbf{x}_i) f_p(y_i|\mathbf{x}_i) dy_i$$
$$= \int \frac{1}{E_r(\phi_i|y_i, \mathbf{x}_i)} f_p(y_i|\mathbf{x}_i) dy_i \quad (40)$$

Step-two: Using Theorem 12 we get

$$E_r(\phi_i|\mathbf{x}_i) = \frac{1}{E_p(\psi_i|\mathbf{x}_i)} \quad (41)$$

The prominent feature of (40) is that, in order to fit a population model for survey data, obtained under an nonignorable nonresponse, we need only the response data set $\{(\mathbf{x}_i, y_i, \phi_i), i \in s\}$ and to specify the underlying population model of $y_i|\mathbf{x}_i$. However we do not need to specify the population conditional expectation of response probabilities

6. FISHER INFORMATION UNDER NONIGNORABLE NONRESPONSE

In this section we assume that the nonignorable nonresponse parameters η are held fixed at their estimated values, that is $\eta = \tilde{\eta}$ is fixed. Now, by maximizing the response log-likelihood function

$$l_r(\theta) = l_{ign}(\theta) - \sum_{i=1}^{m} \log E_p(\psi_i|\mathbf{x}_i, \theta, \tilde{\eta})$$
$$= l_{ign}(\theta) + \sum_{i=1}^{m} \log E_r(\phi_i|\mathbf{x}_i, \theta, \tilde{\eta}) \quad (42)$$

We get the ML estimator of the population parameter θ, which is the starting point for inference. We are usually interested in constructing confidence interval for parameter θ; see Section 7. In such case we need to calculate the value of Fisher information at $\theta = \hat{\theta}$. Let $sc_{pi}(\theta) = \partial \log f_p(y_i; \theta)/\partial\theta$ and $sc_{ri}(\theta) = \partial \log f_r(y_i; \theta)/\partial\theta$ be the score functions in one observation y_i evaluated under the population and response distributions, respectively.

So, we next arrive at two new results embodied in the following two theorems.

Theorem 5. The response score function is given by

$$E_r\left(\frac{\partial \log f_p(y_i; \theta)}{\partial\theta}\right) = \frac{\partial \log E_p(\pi_i|\theta, \eta)}{\partial\theta} \quad (43)$$

Proof:

Since $E_r\{\partial \log f_r(y_i; \theta)/\partial\theta\} = 0$, therefore using (12), we have

$$E_r\left(\frac{\partial \log E_p(\psi_i|y_i, \eta)}{\partial\theta} + \frac{\partial \log f_p(y_i; \theta)}{\partial\theta} - \frac{\partial \log E_p(\psi_i|\theta, \eta)}{\partial\theta}\right) = 0$$

But $E_p(\psi_i|y_i, \eta)$ is free of θ. Hence the result.

Theorem 6. The Fisher information with respect to the response distribution (or the response Fisher information) in one observation is

$$I_{ri}(\theta) = \mathrm{V}ar_r\left\{\frac{\partial \log f_p(y_i;\, \theta)}{\partial\theta}\right\} = \mathrm{V}ar_r\left\{sc_{pi}(\theta)\right\} \quad (44)$$

Proof: The Fisher information under the response distribution in one observation is given by

$$I_{ri}(\theta) = E_r\left\{\frac{\partial \log f_r(y_i;\, \theta, \eta)}{\partial\theta}\right\}^2 = -E_r\left\{\frac{\partial^2 \log f_r(y_i;\, \theta, \eta)}{\partial\theta^2}\right\}$$

Applying (12) and Theorem 3 we get

$$I_{ri}(\theta) = E_r\left\{\frac{\partial \log f_p(y_i;\, \theta)}{\partial\theta}\right\}^2 - \left\{E_r\left(\frac{\partial \log f_p(y_i;\, \theta)}{\partial\theta}\right)\right\}^2$$

$$= \mathrm{Var}_r\left(\frac{\partial \log f_p(y_i;\, \theta)}{\partial\theta}\right)$$

Corollary 5. The Fisher information with respect to the response distribution in a respondent set of size m is

$$I_r(\theta) = \sum_{i=1}^{m} I_{ri}(\theta) = \sum_{i=1}^{m} \mathrm{Var}_r\left\{\frac{\partial \log f_p(y_i;\, \theta)}{\partial\theta}\right\} \quad (45)$$

Example 3. Let $y_i \underset{p}{\sim} N(\mu, \sigma^2);\ i = 1, ..., N.$ Assume that σ^2 is known.

(a) If $E_p(\psi_i | y_i) = \exp(a_0 + a_1 y_i)$, then $I_r(\mu) = \sum_{i=1}^{m} \mathrm{Var}_s\left\{\frac{1}{\sigma^2}(y_i - \mu)\right\} = \frac{m}{\sigma^2}.$

Note that $I_p(\mu) = \sum_{i=1}^{N} \mathrm{Var}_p\left\{\partial \log f_p(y_i;\, \mu)/\partial\mu\right\}$
$= N/\sigma^2 > I_r(\mu)$.

(b) If $E_p(\psi_i | y_i) = b_0 + b_1 y_i$, then $I_r(\mu) = n\{(1/\sigma^2) - (b_1/b_0 + b_1\mu)^2\}$.

If $b_1 = 0$, then $I_r(\mu) = n/\sigma^2 < N/\sigma^2 = I_p(\mu)$.

7. VARIANCE ESTIMATION AND CONFIDENCE INTERVAL

Let $\hat{\theta}$ be the MLE of θ defined by the solution of $\partial l_r(\theta)/\partial\theta = 0$. For the variance estimation of $\hat{\theta}$, we consider estimating the conditional variance of $\hat{\theta}$, given the nonignorable nonresponse parameters η are held fixed at their estimated values. The conditional Fisher information evaluated at $\theta = \hat{\theta}$ is given by

$$\hat{\mathrm{V}}\mathrm{ar}(\hat{\theta}) = \hat{V}_{response}(\hat{\theta}) = \left[\sum_{i=1}^{m} \mathrm{Var}_r\left\{\frac{\partial \log f_p(y_i;\, \theta)}{\partial\theta}\right\}\Bigg|_{\theta=\hat{\theta}}\right]^{-1} \quad (46)$$

Example 4. Assume that the population distribution of the outcome variable y_i, $i \in U$ is Poisson with parameter θ, so that

$$f_p(y_i|\theta) = \frac{\theta^{y_i}\exp(-\theta)}{y_i\,!};\ \theta > 0 \text{ and } y_i = 0, 1, 2, ...$$

Then $\partial \log f_p(y_i | \theta)/\partial\theta = (y_i/\theta) - 1$.

(a) If $E_p(\psi_i|y_i) = \eta y_i$, then

$$f_r(y_i|\theta) = \frac{\theta^{y_i - 1}\exp(-\theta)}{(y_i - 1)!};\ \theta > 0 \text{ and } y_i = 1, 2, ...$$

So that, under this response pdf, we can show that: $I_r(\theta) = m/\theta$. Consequently, According to (35), the MLE of θ is $\hat{\theta} = \bar{y}_r - 1$, so that $\hat{\mathrm{V}}\mathrm{ar}(\hat{\theta}) = (\bar{y}_r - 1)/m$. Hence an approximate $100(1 - \alpha)\%$ confidence interval for θ is given by

$$L_r = (\bar{y}_r - 1) - z_{\alpha/2}\sqrt{(\bar{y}_r - 1)/m}$$

and

$$U_r = (\bar{y}_r - 1) + z_{\alpha/2}\sqrt{(\bar{y}_r - 1)/m}$$

Note that the length of this confidence interval (L_s, U_s) is

$$U_r - L_r = 2z_{\alpha/2}\sqrt{(\bar{y}_r - 1)/m}$$

and the midpoint is

$$(U_r + L_r)/2 = \bar{y}_r - 1 = \left\{(U_{ign} + L_{ign})/2\right\} - 1$$

where

$$L_{ign} = \left\{\bar{y}_r - z_{\alpha/2}\sqrt{\bar{y}_r/m}\right\}$$

and

$$U_{ign} = \left\{\bar{y}_r + z_{\alpha/2}\sqrt{\bar{y}_r/m}\right\}$$

(b) If $E_p(\psi_i/y_i) = \exp(\eta_0 + \eta_1 y_i)$, then we can show that

$$f_r(y_i|\theta) = \frac{\{\theta\exp(\eta_1)\}^{y_i}\exp\{-\theta\exp(\eta_1)\}}{(y_i)!},\ y_i = 0,\ 1,\ 2,...$$

Note that, the response distribution is also Poisson but, under the response distribution, the parameter θ is $\theta^* = \{\theta\exp(\eta_1)\}$.

It is easy to verify that $I_{ri}(\theta) = \exp(\eta_1)/\theta$. Using (35), if η_1 is fixed, then the MLE of θ is $\hat{\theta} = \bar{y}_r/\exp(\eta_1)$. Hence

$$\hat{\mathrm{V}}\mathrm{ar}(\hat{\theta}) = \frac{\bar{y}_r}{n\exp(2\eta_1)}$$

Thus if η_1 is estimated by $\tilde{\eta}_1$, therefore an approximate 100(1 – α)% confidence interval for θ is given by

$$L_r = \{\exp(\tilde{\eta}_1)\}^{-1} L_{ign}$$

and

$$U_r = \{\exp(\tilde{\eta}_1)\}^{-1} U_{ign}$$

Note that the length of this confidence interval is

$$U_r - L_r = \{\exp(\tilde{\eta}_1)\}^{-1}(U_{ign} - L_{ign})$$

and the midpoint is

$$(U_r + L_r)/2 = \bar{y}_r\{\exp(\tilde{\eta}_1)\}^{-1} = \{\exp(\tilde{\eta}_1)\}^{-1}(U_r + L_r)/2$$

If $\eta_1 > 0$, that is, larger values from the population appear in the response set more often than smaller values, then the length of the confidence interval for θ based on the response distribution decreases by a factor $1/\exp(\eta_1)$.

If $\eta_1 < 0$, that is, smaller values from the population appear in the response set more often than larger values, then the length of the confidence interval for θ based on the response distribution increases by a factor $1/\exp(\eta_1)$.

8. PREDICTION OF FINITE POPULATION PARAMETER UNDER NONIGNORABLE NONRESPONSE

Sverchkov and Pfeffermann (2004) use sample and sample complement distributions for the prediction of finite population totals under informative sampling for single-stage sampling designs. Later Eideh and Nathan (2009) extend the theory to general linear functions of the population values and to two-stage informative cluster sampling. In this section we use the response and nonresponse distributions to predict the finite population total under noninformative sampling design and under nonignorable nonresponse. We consider the prediction for single-stage sampling and under three models namely, common mean population model, simple ratio population model, and simple regression population model.

8.1 Preliminaries

Assume single-stage population model. Let

$$T_\ell = \sum_{i=1}^{N} l_i y_i = \sum_{i\in s} l_i y_i + \sum_{i\in \bar{s}} l_i y_i = \sum_{i\in r} l_i y_i + \sum_{i\in \bar{r}} l_i y_i + \sum_{i\in \bar{s}} l_i y_i \tag{47}$$

where $(l_1, ..., l_N)$ is a vector of known constants, be the linear function of population values that we want to predict using the data from the response set and possibly values of auxiliary variables that may include some or all of the design variables.

Notice that T_ℓ can be decomposed into three components, the first component represents the total for observed units in the sample – response set, $\sum_{i\in r} l_i y_i$, the second component represents the total for unobserved units in sample – nonresponse set, $\sum_{i\in \bar{r}} l_i y_i$, and the third component represents the total for non-sample units, $\sum_{i\in \bar{s}} l_i y_i$.

For the prediction process we have the following available information:

(a) The information that comes from the sampling design denoted by

$O_s = [\{(x_i, I_i), i \in U\}, \{\pi_i, i \in s\}]$, where $I_i = 1$ for $i \in s$ and $I_i = 0$ for $i \notin s$.

(b) Information that comes from the response set denoted by

$O_r = [\{(y_i, \psi_i), i \in r\} | i \in s]$, N, n and m.

Thus the available information, from the sample and response set, for the prediction process is $O = O_s \cup O_r$.

Let $\hat{T}_\ell = \hat{T}_\ell(O)$ define the predictor of T_ℓ based on O. The mean square error (MSE) of $\hat{T}_\ell$ given with respect to the population pdf is defined by

$$\begin{aligned}\mathrm{MSE}_p(\hat{T}_\ell) &= E_p\{(\hat{T}_\ell - T_\ell)^2 | O\} \\ &= E_p[\{\hat{T}_\ell - E_p(T_\ell|O) + E_p(T_\ell|O) - T_\ell\}^2 | O] \\ &= \{\hat{T}_\ell - E_p(T_\ell|O)\}^2 + \mathrm{Var}_p(T_\ell|O)\end{aligned} \quad (48)$$

It is obvious from the last line of (48) that, (48) is minimized when $\hat{T}_\ell = E(T_\ell|O)$. Now we consider the following

$$\begin{aligned}E_p(T_\ell|O) &= E_p\left\{\left(\sum_{i\in s} l_i y_i + \sum_{i\in\bar{s}} l_i y_i\right)\middle| O\right\} \\ &= \sum_{i\in r} l_i y_i + \sum_{i\in\bar{r}} l_i E_{\bar{r}}(y_i|O) + \sum_{i\in\bar{s}} l_i E_p(y_i|O)\end{aligned} \quad (49)$$

Thus the general predictor for T_ℓ under nonignorable nonresponse is

$$\begin{aligned}\hat{T}_\ell &= E_p(T_\ell|O) = E_p\left\{\left(\sum_{i\in s} l_i y_i + \sum_{i\in\bar{s}} l_i y_i\right)\middle| O\right\} \\ &= \sum_{i\in r} l_i y_i + \sum_{i\in\bar{r}} l_i \hat{E}_{\bar{r}}(y_i|O) + \sum_{i\in\bar{s}} l_i \hat{E}_p(y_i|O)\end{aligned} \quad (50)$$

The predictor given in (50) represents the prediction of T_ℓ for single-stage sampling when the sampling mechanism in noninformative and missing value mechanism is nonignorable. The analysis that follows assumes known model parameters. In practice, the unknown model parameters are replaced under the frequentist approach by sample estimates, yielding the corresponding "empirical predictors." In the present case, maximum likelihood estimation of the model parameters must be based on the response distribution of the observed units in the sample – response set; see Sections 5-7.

We now consider the following special cases for prediction of the overall population total

$$T = \sum_{i=1}^{N} y_i = \sum_{i\in r} y_i + \sum_{i\in\bar{r}} y_i + \sum_{i\in\bar{s}} y_i$$

that is, for $(l_1, ..., l_N) = (1, 1, ..., 1)$.

According to (50), the general predictor for T under nonignorable nonresponse is

$$E_p(T|O) = \sum_{i\in r} y_i + \sum_{i\in\bar{r}} E_{\bar{r}}(y_i|O) + \sum_{i\in\bar{s}} E_p(y_i|O) \quad (51)$$

We know the values $\{y_i' \, s, i \in r\}$, so the sum $\sum_{i\in r} y_i$ is known. Thus to estimate for our response set, we need to predict the total for unobserved units in the sample – nonresponse set, $\sum_{i\in\bar{r}} y_i$, and the total for non-sample units, $\sum_{i\in\bar{s}} y_i$. That is, to predict T we need to predict values for the $\{y_i' \, s, i \in \bar{r}\}$ and values for the $\{y_i' \, s, i \in \bar{s}\}$.

According to (14e), we have

$$\begin{aligned}E_{\bar{r}}(y_i) &= \frac{E_p\{(1-\psi_i)y_i\}}{E_p(1-\psi_i)} = \frac{E_p(y_i) - E_p(y_i\psi_i)}{1 - E_p(\psi_i)} \\ &= \frac{E_p(y_i) - E_p(y_i\psi_i) - E_p(y_i)E_p(\psi_i) + E_p(y_i)E_p(\psi_i)}{1 - E_p(\psi_i)} \\ &= \frac{E_p(y_i)E_p(1-\psi_i) - \mathrm{Cov}_p(\psi_i, y_i)}{1 - E_p(\psi_i)} \\ &= E_p(y_i) - \frac{\mathrm{Cov}_p(\psi_i, y_i)}{1 - E_p(\psi_i)}\end{aligned} \quad (52)$$

So that

$$E_p(y_i) - E_{\bar{r}}(y_i) = \frac{\mathrm{Cov}_p(\psi_i, y_i)}{1 - E_p(\psi_i)} \quad (53)$$

Hence

$$\begin{aligned}\hat{T}_{nign} &= E_p(T|O) = \sum_{i\in r} y_i + \sum_{i\in\bar{r}} E_{\bar{r}}(y_i|O) + \sum_{i\in\bar{s}} E_p(y_i|O) \\ &= \sum_{i\in r} y_i + \sum_{i\in\bar{r}}\left\{E_p(y_i|O) - \frac{\mathrm{Cov}_p[(\psi_i, y_i)|O]}{1 - E_p(\psi_i)}\right\} \\ &\quad + \sum_{i\in\bar{s}} E_p(y_i|O)\end{aligned} \quad (54)$$

$$= \sum_{i\in r} y_i + \sum_{i\in\bar{r}} E_p\left(y_i|O\right) + \sum_{i\in\bar{s}} E_p\left(y_i|O\right)$$

$$-\sum_{i\in\bar{r}} \frac{\mathrm{Cov}_p\left[\left(\psi_i, y_i\right)|O\right]}{1-E_p\left(\psi_i\right)}$$

$$= \hat{T}_{ign} - \sum_{i\in\bar{r}} \frac{\mathrm{Cov}_p\left[\left(\psi_i, y_i\right)|O\right]}{1-E_p\left(\psi_i\right)}$$

where

$$\hat{T}_{ign} = \sum_{i\in r} y_i + \sum_{i\in\bar{r}} E_p\left(y_i|O\right) + \sum_{i\in\bar{s}} E_p\left(y_i|O\right) \qquad (55)$$

is the best linear unbiased predictor (BLUP) of $T = \sum_{i=1}^{N} y_i$ under ignorable nonresponse.

The nonresponse bias of $\hat{T}_{nign}$ is

$$B\left(\hat{T}_{nign}\right) = E_p\left(\hat{T}_{nign} - T\right) = -\left\{\sum_{i\in\bar{s}} E_p\left[y_i - E_p\left(y_i\right)\right] + \sum_{i\in\bar{r}} E_p\left[y_i - E_{\bar{r}}\left(y_i\right)\right]\right\} \qquad (56)$$

$$= -\sum_{i\in\bar{r}}\left[E_p\left(y_i\right) - E_{\bar{r}}\left(y_i\right)\right] = -\sum_{i\in\bar{r}} \frac{\mathrm{Cov}_p\left(\psi_i, y_i\right)}{E_p\left(1-\psi_i\right)}$$

Hence, the predictor $\hat{T}_{nign}$ is unbiased if there is no correlation between the study variable and the response probabilities ψ_i. The stronger the relationship between the study variable and the response probability, the larger the bias. Similar result was obtained by Bethlehem (1988).

Using equations (14a-e), we obtain

$$E_p\left(y_i\right) - E_{\bar{r}}\left(y_i\right) = \frac{E_r\left(\phi_i y_i\right)}{E_r\left(\phi_i\right)} - \frac{E_r\left\{\left(\phi_i - 1\right)y_i\right\}}{E_r\left\{\left(\phi_i - 1\right)\right\}}$$

$$= \frac{E_r\left(\phi_i y_i\right)\left[E_r\left(\phi_i\right) - 1\right] - \left[E_r\left(\phi_i y_i\right) - E_r\left(y_i\right)\right]E_r\left(\phi_i\right)}{E_r\left(\phi_i\right)E_r\left\{\left(\phi_i - 1\right)\right\}}$$

$$= \frac{E_r\left(\phi_i y_i\right)E_r\left(\phi_i\right) - E_r\left(\phi_i y_i\right) - E_r\left(\phi_i y_i\right)E_r\left(\phi_i\right) + E_r\left(y_i\right)E_r\left(\phi_i\right)}{E_r\left(\phi_i\right)E_r\left\{\left(\phi_i - 1\right)\right\}}$$

$$= \frac{-\left\{E_r\left(\phi_i y_i\right) - E_r\left(y_i\right)E_r\left(\phi_i\right)\right\}}{E_r\left(\phi_i\right)E_r\left\{\left(\phi_i - 1\right)\right\}} = \frac{\mathrm{Cov}_r\left(\phi_i, y_i\right)}{E_r\left(\phi_i\right)E_r\left\{\left(\phi_i - 1\right)\right\}} \qquad (57)$$

Hence (54) and (56) can be written as

$$\hat{T}_{nign} = E_p\left(T|O\right) = \hat{T}_{ign} - \sum_{i\in\bar{r}} \frac{\mathrm{Cov}_r\left(\phi_i, y_i\right)}{E_r\left(\phi_i\right)E_r\left\{\left(\phi_i - 1\right)\right\}} \qquad (58)$$

and

$$B\left(\hat{T}_{nign}\right) = -\sum_{i\in\bar{r}} \frac{\mathrm{Cov}_r\left(\phi_i, y_i\right)}{E_r\left(\phi_i\right)E_r\left\{\left(\phi_i - 1\right)\right\}} \qquad (59)$$

Thus, if the for unobserved units in sample and the response weights ϕ_i are correlated, then ignoring the sampling scheme yields biased predictors.

Schaible (1983), defined the incomplete data bias in a predictor $\hat{T}_{nign}(r)$ as the total bias in the predictor minus the bias that would occur if the sample were complete $\hat{T}_n$

$$B\left(\hat{T}_{nign(r)}\right) = E_p\left(\hat{T}_{nign(r)} - T\right) - E_p\left(\hat{T}_n - T\right)$$

$$= E_p\left(\hat{T}_{nign(r)} - \hat{T}_n\right)$$

$$= E_p\left\{\left(\sum_{i\in r} y_i + \sum_{i\in\bar{r}} E_{\bar{r}}\left(y_i\right) + \sum_{i\in\bar{s}} E_p\left(y_i\right)\right) - \left(\sum_{i\in r} y_i + \sum_{i\in\bar{r}} y_i + \sum_{i\in\bar{s}} E_p\left(y_i\right)\right)\right\}$$

$$= E_p\left(\sum_{i\in r} y_i + \sum_{i\in\bar{r}} E_{\bar{r}}\left(y_i\right) + \sum_{i\in\bar{s}} E_p\left(y_i\right)\right) - E_p\left(\sum_{i\in r} y_i + \sum_{i\in\bar{r}} y_i + \sum_{i\in\bar{s}} E_p\left(y_i\right)\right)$$

$$= -E_p\sum_{i\in\bar{r}}\left(y_i - E_{\bar{r}}\left(y_i\right)\right)$$

$$= -\sum_{i\in\bar{r}} E_p\left(y_i - E_{\bar{r}}\left(y_i\right)\right)$$

$$= -\sum_{i\in\bar{r}}\left(E_p\left(y_i\right) - E_{\bar{r}}\left(y_i\right)\right)$$

$$= -\sum_{i\in\bar{r}} \frac{\mathrm{Cov}_p\left(\psi_i, y_i\right)}{E_p\left(1-\psi_i\right)} \qquad (60)$$

Hence $E_p\left(\hat{T}_{nign(r)} - T\right) - E_p\left(\hat{T}_n - T\right) = E_p\left(\hat{T}_{nign(r)} - T\right)$.

8.2 Common Mean Population Model

The common mean population model (C) stating that $y_i \underset{p}{\sim} N\left(\mu,\, \sigma^2\right),\ i = 1, ..., N$ are independent normal random variable, with mean $E_p(y_i) = \mu$ and variance $V_p(y_i) = \sigma^2$.

Under the exponential response probability model

$$E_p(\psi_i | y_i) = \exp(\eta y_i) \tag{61}$$

we have, the response distribution as $y_i \underset{r}{\sim} N\left(\mu + \eta\sigma^2, \sigma^2\right),\ i = 1, ..., m.$ That is,

$$E_r(y_i) = \mu + \eta\sigma^2 = E_p(y_i) + \eta\sigma^2 \text{ and } \mathrm{Var}_r(y_i) = \sigma^2 \tag{62}$$

Accordingly,

$$E_p(\psi_i) = E_p E_p(\psi_i | y_i) = E_p\{\exp(\eta y_i)\} \tag{63}$$

$$= M_p(\eta) = \exp\left(\eta\mu + \frac{\eta^2\sigma^2}{2}\right)$$

and

$$E_p(y_i\psi_i) = E_p(E_p(\psi_i y_i | y_i)) = E_p(y_i E_p(\psi_i | y_i))$$

$$= E_p\left\{y_i \exp\left(\eta y_i\right)\right\} = \frac{d}{d\eta} M_p(\eta) \tag{64}$$

$$= \left(\mu + \eta\sigma^2\right)\exp\left(\mu\eta + \frac{\eta^2\sigma^2}{2}\right)$$

$$= E_r(y_i)\, M_p(\eta)$$

Also,

$$\mathrm{Cov}_p(\psi_i, y_i) = (\eta\sigma^2) M_p(\eta) \tag{65a}$$

and

$$\frac{\mathrm{Cov}_p(\psi_i,\ y_i)}{E_p(1-\psi_i)} = \left(\eta\sigma^2\right)\frac{M_p(\eta)}{1-M_p(\eta)} \tag{65b}$$

Hence, according to (52) and (65), we have,

$$E_{\bar{r}}(y_i) = \mu - \left(\eta\sigma^2\right)\frac{M_p(\eta)}{1-M_p(\eta)} \tag{65c}$$

Using the results obtained in Section 8.1, and results (62-65) we obtain the following.

(a) The BLUP for T under nonignorable nonresponse is

$$\hat{T}_{C,nign} = \sum_{i\in r} y_i + (n-m)\left\{\mu - \left(\eta\sigma^2\right)\frac{M_p(\eta)}{1-M_p(\eta)}\right\} + (N-n)\mu \tag{66}$$

Under the response distribution given in (62), we can show that the ML estimators of the common population model are given by

$$\hat{\mu}_{C,nign} = \bar{y}_r - \tilde{\eta} s^2_{C,nign} = \frac{1}{m}\sum_{i\in r} y_i - \tilde{\eta} s^2_{C,nign} \tag{67}$$

and

$$s^2_{C,nign} = \frac{1}{m}\sum_{i\in r}\left(y_i - \bar{y}_r\right)^2 \tag{68}$$

where $\tilde{\eta}$ is the least square estimator obtained via the following relationship in (14a)

$$E_r(\phi_i | y_i) = \{E_p(\psi_i | y_i)\}^{-1}$$

So that

$$\hat{T}_{C,nign} = N\bar{y}_r - \tilde{\eta} s^2_{C,nign}\left\{(N-m) + \frac{\hat{M}_p(\tilde{\eta})}{1-\hat{M}_p(\tilde{\eta})}\right\} \tag{69}$$

where

$$\hat{M}_p(\tilde{\eta}) = \exp\left(\hat{\mu}_{C,nign}\tilde{\eta} + \frac{\tilde{\eta}^2 s^2_{C,nign}}{2}\right) \tag{70}$$

If $\eta = 0$, that is, the missing value mechanism is ignorable, then

$$\hat{T}_{c,ign} = \sum_{i\in r} y_i + (n-m)\{\mu\} + (N-n)\mu$$

$$= \sum_{i\in r} y_i + (N-m)\mu \tag{71}$$

Now, under ignorable nonresponse, the MLE of μ is

$$\hat{\mu}_{C,ign} = \bar{y}_r = \frac{1}{m}\sum_{i\in r} y_i$$

then

$$\hat{T}_{c,ign} = \sum_{i\in r} y_i + (N-m)\bar{y}_r = m\bar{y}_r + (N-m)\bar{y}_r$$

$$= N\bar{y}_r = \frac{N}{m}\sum_{i\in r} y_i \tag{72}$$

which is the classical BLUP known in sampling surveys.

Also, the estimate of the bias of $\hat{T}_{nign}$ is given by

$$B\left(\hat{T}_{C,nign}\right)=-\sum_{i\in\bar{r}}\left(\tilde{\eta}s^2_{C,nign}\right)\frac{\hat{M}_p\left(\tilde{\eta}\right)}{1-\hat{M}_p\left(\tilde{\eta}\right)} \quad (73)$$

Using (59), the method of moment estimate of the bias of $\hat{T}_{C,nign}$ is given by

$$\begin{aligned} B\left(\hat{T}_{C,nign}\right) &= -\sum_{i\in\bar{r}}\frac{Cov_r\left(\phi_i, y_i\right)}{E_r\left(\phi_i\right)E_r\left\{(\phi_i-1)\right\}} \\ &=-(n-m)\frac{m^{-1}\sum_{i\in r}\left(\phi_i-\bar{\phi}_r\right)\left(y_i-\bar{y}_r\right)}{\left(m^{-1}\sum_{i\in r}\phi_i\right)\left(m^{-1}\sum_{i\in r}(\phi_i-1)\right)} \\ &= -(n-m)\frac{m^{-1}\sum_{i\in r}\left(\phi_i-\bar{\phi}_r\right)\left(y_i-\bar{y}_r\right)}{\left(m^{-1}\sum_{i\in r}\phi_i\right)\left(m^{-1}\sum_{i\in r}(\phi_i-1)\right)} \\ &=-m\frac{m^{-1}\sum_{i\in r}\left(\phi_i-\bar{\phi}_r\right)\left(y_i-\bar{y}_r\right)}{\left(m^{-1}\sum_{r}\phi_i\right)} \\ &= -m\frac{C_r\left(\phi_i, y_i\right)}{\bar{\phi}_r}=-\frac{m^2}{n}C_r\left(\phi_i, y_i\right) \quad (74) \end{aligned}$$

where $\sum_{i\in r}\phi_i = n.$

8.3 Simple Ratio Population Model

The simple ratio population model (R) stating that $y_i|x_i \underset{p}{\sim} N\left(\beta x_i,\ \sigma^2 x_i\right)$, $i = 1, \ldots, N$ are independent normal random variable, with mean $E_p(y_i|x_i) = \beta x_i$ and variance $\mathrm{Var}_p(y_i|x_i) = \sigma^2 x_i$.

Under the exponential response probability model

$$E_p(\psi_i|y_i) = \exp(\eta\, y_i) \quad (75)$$

We have, the response distribution is $y_i|x_i \underset{r}{\sim} N\ ((\eta\sigma^2 + \beta)\, x_i,\ \sigma^2 x_i)$, $i = 1, \ldots, m$. That is,

$$E_r(y_i|x_i) = (\eta\sigma^2 + \beta)x_i \quad (76a)$$

and

$$\mathrm{Var}_r(y_i|x_i) = \sigma^2 x_i \quad (76b)$$

Accordingly,

$$E_p(\psi_i) = M_p(\eta)=\exp\left(\eta\left(\beta x_i\right)+\frac{\eta^2\sigma^2 x_i}{2}\right) \quad (77)$$

and

$$\begin{aligned} E_p(y_i\psi_i) &= \frac{d}{d\eta}M_p(\eta) \\ &= \left(\beta x_i+\eta\sigma^2 x_i\right)\exp\left(\eta\left(\beta x_i\right)+\frac{\eta^2\sigma^2 x_i}{2}\right) \quad (78) \\ &= E_r(y_i)\ M_p(\eta) \end{aligned}$$

Also,

$$\mathrm{Cov}_p(\psi_i, y_i) = \left(\eta\sigma^2 x_i\right)\exp\left(\eta\left(\beta x_i\right)+\frac{\eta^2\sigma^2 x_i}{2}\right) \quad (79a)$$

and

$$\frac{\mathrm{Cov}_p\left(\psi_i,\ y_i\right)}{E_p\left(1-\psi_i\right)} = \left(\eta\sigma^2 x_i\right)\frac{\exp\left(\eta\left(\beta x_i\right)+\frac{\eta^2\sigma^2 x_i}{2}\right)}{1-\exp\left(\eta\left(\beta x_i\right)+\frac{\eta^2\sigma^2 x_i}{2}\right)} \quad (79b)$$

Hence, according to (52) and (65), we have

$$E_{\bar{r}}\left(y_i\right)=\beta x_i-\left(\eta\sigma^2 x_i\right)\frac{\exp\left(\eta\left(\beta x_i\right)+\frac{\eta^2\sigma^2 x_i}{2}\right)}{1-\exp\left(\eta\left(\beta x_i\right)+\frac{\eta^2\sigma^2 x_i}{2}\right)} \quad (80)$$

Thus

$$\hat{T}_{nign,r} = \sum_{i\in r}y_i+\sum_{i\in\bar{r}}\left\{\beta x_i-\left(\eta\sigma^2 x_i\right)\frac{\exp\left(\eta\left(\beta x_i\right)+\frac{\eta^2\sigma^2 x_i}{2}\right)}{1-\exp\left(\eta\left(\beta x_i\right)+\frac{\eta^2\sigma^2 x_i}{2}\right)}\right\} + \sum_{i\in\bar{s}}\beta x_i$$

$$= \sum_{i\in r} y_i + \sum_{i\in \bar{r}} \beta x_i + \sum_{i\in \bar{s}} \beta x_i$$

$$- \sum_{i\in \bar{r}} \left\{ \left(\eta\sigma^2 x_i\right) \frac{\exp\left(\eta\left(\beta x_i\right) + \frac{\eta^2\sigma^2 x_i}{2}\right)}{1-\exp\left(\eta\left(\beta x_i\right) + \frac{\eta^2\sigma^2 x_i}{2}\right)} \right\}$$

$$= \hat{T}_{R,ign} - \sum_{i\in \bar{r}} \left\{ \left(\eta\sigma^2 x_i\right) \frac{\exp\left(\eta\left(\beta x_i\right) + \frac{\eta^2\sigma^2 x_i}{2}\right)}{1-\exp\left(\eta\left(\beta x_i\right) + \frac{\eta^2\sigma^2 x_i}{2}\right)} \right\} \tag{81}$$

where

$$\hat{T}_{R,ign} = \sum_{i\in r} y_i + \sum_{i\in \bar{r}} \beta x_i + \sum_{i\in \bar{s}} \beta x_i \tag{82}$$

Under the response distribution given in (76), we can show that the ML estimators of the simple ratio population model are given by

$$\tilde{\eta}\hat{\sigma}^2_{R,nign} + \hat{\beta}_{R,nign} = \frac{\bar{y}_r}{\bar{x}_r} \tag{83a}$$

$$\hat{\sigma}^2_{R,nign} = \frac{1}{r}\sum_{i\in r}\frac{1}{x_i}\left(y_i - \left(\tilde{\eta}\hat{\sigma}^2_{R,nign} + \hat{\beta}_{R,nign}\right)x_i\right)^2$$
$$= \frac{1}{r}\sum_{i\in r}\frac{1}{x_i}\left(y_i - \left(\frac{\bar{y}_r}{\bar{x}_r}\right)x_i\right)^2 \tag{83b}$$

$$\hat{\beta}_{R,nign} = \frac{\bar{y}_r}{\bar{x}_r} - \tilde{\eta}\hat{\sigma}^2_{R,nign}$$
$$= \frac{\bar{y}_r}{\bar{x}_r} - \tilde{\eta}\frac{1}{r}\sum_{i\in r}\frac{1}{x_i}\left(y_i - \left(\frac{\bar{y}_r}{\bar{x}_r}\right)x_i\right)^2 \tag{83c}$$

where $\tilde{\eta}$ is the least square estimator obtained via the relationship in (14a).

So that, if $\eta = 0$, that is, the missing value mechanism is ignorable, then

$$\hat{\beta}_{R,ign} = \frac{\bar{y}_r}{\bar{x}_r} \tag{84a}$$

and

$$\hat{\sigma}^2_{R,ign} = \frac{1}{r}\sum_{i\in r}\frac{1}{x_i}\left(y_i - \left(\frac{\bar{y}_r}{\bar{x}_r}\right)x_i\right)^2 \tag{84b}$$

Therefore,

$$\hat{T}_{R,ign} = \sum_{i\in r} y_i + \sum_{i\in \bar{r}} \hat{\beta}_{R,ign} x_i + \sum_{i\in \bar{s}} \hat{\beta}_{R,ign} x_i$$
$$= \frac{\bar{y}_r}{\bar{x}_r} N\bar{X} \tag{85}$$

which is the classical ratio estimator under nonignorable nonresponse and under noninformative sampling design.

Now, the Bias of $\hat{T}_{R,nign}$ is

$$B\left(\hat{T}_{nign}\right) = E_p\left(\hat{T}_{nign} - T\right)$$
$$= -\sum_{i\in \bar{r}} \left\{ \left(\eta\sigma^2 x_i\right) \frac{\exp\left(\eta\left(\beta x_i\right) + \frac{\eta^2\sigma^2 x_i}{2}\right)}{1-\exp\left(\eta\left(\beta x_i\right) + \frac{\eta^2\sigma^2 x_i}{2}\right)} \right\} \tag{86}$$

8.4 Simple Regression Population Model

The simple regression population model (L) stating that $y_i|x_i \underset{p}{\sim} N(\beta_0 + \beta x_i, \sigma^2)$, $i = 1, ..., N$ are independent normal random variable, with mean $E_p(y_i|x_i) = \beta_0 + \beta x_i$ and variance $\mathrm{Var}_p(y_i|x_i) = \sigma^2$.

Under the exponential response probability model

$$E_p(\psi_i|y_i) = \exp(\eta y_i) \tag{87}$$

We have, the response distribution is

$$y_i|x_i \underset{s}{\sim} N(\eta\sigma^2 + \beta_0) + \beta_1 x_i, \sigma^2), \ i = 1, ..., m.$$

That is,

$$E_r(y_i|x_i) = (\eta\sigma^2 + \beta_0) + \beta_1 x_i \text{ and } \mathrm{Var}_r(y_i|x_i) = \sigma^2 \tag{88}$$

Accordingly,

$$E_p\left(\psi_i\right) = M_p\left(\eta\right) = \exp\left(\eta\left(\beta_0 + \beta_1 x_i\right) + \frac{\eta^2\sigma^2}{2}\right) \tag{89}$$

and

$$E_p(y_i\psi_i) = \frac{d}{d\eta_1} M_p\left(\eta\right)$$

$$= \left(\left(\beta_0 + \beta_1 x_i\right) + \eta\sigma^2\right)\exp\left(\eta\left(\beta_0 + \beta_1 x_i\right) + \frac{\eta^2\sigma^2}{2}\right) \quad (90)$$

$= E_r(y_i)\ M_p(\eta)$

Hence,

$$\mathrm{Cov}_p\left(\psi_i, y_i\right) = \left(\eta\sigma^2\right)\exp\left(\eta\left(\beta_0 + \beta_1 x_i\right) + \frac{\eta^2\sigma^2}{2}\right) \quad (91a)$$

and

$$\frac{\mathrm{Cov}_p\left(\psi_i, y_i\right)}{E_p\left(1-\psi_i\right)} = \left(\eta\sigma^2\right)\frac{\exp\left(\eta\left(\beta_0 + \beta_1 x_i\right) + \frac{\eta^2\sigma^2}{2}\right)}{1-\exp\left(\eta\left(\beta_0 + \beta_1 x_i\right) + \frac{\eta^2\sigma^2}{2}\right)} \quad (91b)$$

Therefore,

$$E_{\bar{r}}\left(y_i\right) = \left(\beta_0 + \beta_1 x_i\right) - \left(\eta\sigma^2\right)\frac{\exp\left(\eta\left(\beta_0 + \beta_1 x_i\right) + \frac{\eta^2\sigma^2}{2}\right)}{1-\exp\left(\eta\left(\beta_0 + \beta_1 x_i\right) + \frac{\eta^2\sigma^2}{2}\right)} \quad (92)$$

Thus,

$$\begin{aligned}\hat{T}_{L,nign} &= \sum_{i\in r} y_i + \sum_{i\in\bar{r}}\left\{\left(\beta_0 + \beta_1 x_i\right) - \left(\eta\sigma^2\right) \times \frac{\exp\left(\eta\left(\beta_0 + \beta_1 x_i\right) + \frac{\eta^2\sigma^2}{2}\right)}{1-\exp\left(\eta\left(\beta_0 + \beta_1 x_i\right) + \frac{\eta^2\sigma^2}{2}\right)}\right\} + \sum_{i\in\bar{s}}\left(\beta_0 + \beta_1 x_i\right)\\ &= \sum_{i\in r} y_i + \sum_{i\in\bar{r}}\left(\beta_0 + \beta_1 x_i\right) + \sum_{i\in\bar{s}}\left(\beta_0 + \beta_1 x_i\right) - \sum_{i\in\bar{r}}\left\{\left(\eta\sigma^2\right)\frac{\exp\left(\eta\left(\beta_0 + \beta_1 x_i\right) + \frac{\eta^2\sigma^2}{2}\right)}{1-\exp\left(\eta\left(\beta_0 + \beta_1 x_i\right) + \frac{\eta^2\sigma^2}{2}\right)}\right\}\\ &= \hat{T}_{L,ign} - \sum_{i\in\bar{r}}\left\{\left(\eta\sigma^2\right)\frac{\exp\left(\eta\left(\beta_0 + \beta_1 x_i\right) + \frac{\eta^2\sigma^2}{2}\right)}{1-\exp\left(\eta\left(\beta_0 + \beta_1 x_i\right) + \frac{\eta^2\sigma^2}{2}\right)}\right\} \quad (93)\end{aligned}$$

where

$$\hat{T}_{L,ign} = \sum_{i\in r} y_i + \sum_{i\in\bar{r}}\left(\beta_0 + \beta_1 x_i\right) + \sum_{i\in\bar{s}}\left(\beta_0 + \beta_1 x_i\right) \quad (94)$$

Under the response distribution given in (88), we can show that the ML estimators of the simple regression population model are given by

$$\tilde{\eta}_1\hat{\sigma}^2_{L,nign} + \hat{\beta}_{L,0nign} = \bar{y}_r - \hat{\beta}_{L,1nign}\bar{x}_r \quad (95a)$$

$$\hat{\beta}_{L,0nign} = \left(\bar{y}_r - \hat{\beta}_{L,1nign}\bar{x}_r\right) - \tilde{\eta}_1\hat{\sigma}^2_{L,nign} \quad (95b)$$

$$\hat{\beta}_{L,1nign} = \frac{\sum_{i\in r}\left(y_i - \bar{y}_r\right)\left(x_i - \bar{x}_r\right)}{\sum_{i\in r}\left(x_i - \bar{x}_r\right)^2} \quad (95c)$$

and

$$\begin{aligned}\hat{\sigma}^2_{L,nign} &= \frac{1}{r}\sum_{i\in r}\left\{\left(y_i - \bar{y}_r\right) - \hat{\beta}_{L,1nign}\left(x_i - \bar{x}_r\right)\right\}^2\\ &= \frac{1}{r}\sum_{i\in r}\left(y_i - \hat{y}_i\right)^2 \quad (95d)\end{aligned}$$

Hence,

$$\begin{aligned}\hat{y}_i &= \left(\tilde{\eta}_1\hat{\sigma}^2_{L,nign} + \hat{\beta}_{L,0nign}\right) + \hat{\beta}_{1nign}x_i\\ &= \bar{y}_r - \hat{\beta}_{L,1nign}\bar{x}_r + \hat{\beta}_{L,1nign}x_i\\ &= \bar{y}_r + \hat{\beta}_{L,1nign}\left(x_i - \bar{x}_r\right) \quad (95e)\end{aligned}$$

where $\tilde{\eta}$ is the least square estimator obtained via the following relationship in (14a).

So that, if $\eta = 0$, that is, the missing value mechanism is ignorable, then

$$\hat{\beta}_{L,0ign} = \bar{y}_r - \hat{\beta}_{L,1ign}\bar{x}_r \quad (96a)$$

$$\hat{\beta}_{L,1ign} = \frac{\sum_{i\in r}\left(y_i - \bar{y}_r\right)\left(x_i - \bar{x}_r\right)}{\sum_{i\in r}\left(x_i - \bar{x}_r\right)^2} \quad (96b)$$

$$\hat{\sigma}^2_{L,ign} = \frac{1}{r}\sum_{i\in r}\left\{\left(y_i - \bar{y}_r\right) - \hat{\beta}_{1ign}\left(x_i - \bar{x}_r\right)\right\}^2$$

$$= \frac{1}{r}\sum_{i\in r}\left(y_i - \hat{y}_i\right)^2 \tag{96c}$$

and

$$\begin{aligned} \hat{y}_i &= \hat{\beta}_{L,0ign} + \hat{\beta}_{L,1ign}x_i \\ &= \bar{y}_r - \hat{\beta}_{L,1ign}\bar{x}_r + \hat{\beta}_{L,1ign}x_i \\ &= \bar{y}_r + \hat{\beta}_{L,1ign}\left(x_i - \bar{x}_r\right) \end{aligned} \tag{96d}$$

Therefore

$$\begin{aligned} \hat{T}_{L,ign} &= \hat{T}_{ign} = \sum_{i\in r} y_i + \sum_{i\in \bar{r}}\left(\hat{\beta}_{L,0ign} + \hat{\beta}_{L,1ign}x_i\right) \\ &\quad + \sum_{i\in \bar{s}}\left(\hat{\beta}_{L,0ign} + \hat{\beta}_{L,1ign}x_i\right) \\ &= \sum_{i\in r} y_i + \hat{\beta}_{L,1ign}\sum_{i\in \bar{r}} x_i + \hat{\beta}_{L,1ign}\sum_{i\in \bar{s}} x_i \\ &\quad + \left(N - m\right)\hat{\beta}_{L,0ign} \\ &= \sum_{i\in r} y_i + \hat{\beta}_{L,1ign}\left(N\bar{X} - m\bar{x}_r\right) \\ &\quad + \left(N - m\right)\left(\bar{y}_r - \hat{\beta}_{L,1ign}\bar{x}_r\right) \\ &= N\bar{y}_r + N\hat{\beta}_{L,1ign}\left(\bar{X} - \bar{x}_r\right) \\ &= N\left[\bar{y}_r + \hat{\beta}_{L,1ign}\left(\bar{X} - \bar{x}_r\right)\right] \end{aligned} \tag{97}$$

which is the classical regression estimator under nonignorable nonresponse and under noninformative sampling design.

Now, the Bias of $\hat{T}_{L,nign}$

$$\begin{aligned} B\left(\hat{T}_{L,nign}\right) &= E_p\left(\hat{T}_{L,nign} - T\right) = -\sum_{i\in \bar{r}}\left[E_p\left(y_i\right) - E_{\bar{r}}\left(y_i\right)\right] \\ &= -\sum_{i\in \bar{r}}\left\{\left(\eta\sigma^2\right)\frac{\exp\left(\eta\left(\beta_0 + \beta_1 x_i\right) + \dfrac{\eta^2\sigma^2}{2}\right)}{1 - \exp\left(\eta\left(\beta_0 + \beta_1 x_i\right) + \dfrac{\eta^2\sigma^2}{2}\right)}\right\} \end{aligned} \tag{98}$$

9. TEST OF NONIGNORABLE NONRESPONSE

Pfeffermann and Sverchkov (1999) use conventional t-correlation test for testing ignorability of sampling design. Later Eideh and Nathan (2006), use the another test based on the Kullback-Leibler information measure. In this section we extend these two tests for testing not at random missing value mechanism.

A natural question arising for nonignorable nonresponse is how to test if the messing value mechanism in not missing at random or nonignorable, so we cannot be ignored for the inference process, given the available design information. Under the assumptions of Section 2, it is easy to verify that $f_r(y_i|\theta, \eta) = f_p(y_i|\theta)$, that is the messing value mechanism is ignorable, if and only if

$$f_r(y_i|\theta, \eta) = f_p(y_i|\theta) \tag{99}$$

9.1 Conventional Correlation t-Statistic

Condition (99) allows us to use the t-correlation test for testing ignorable nonresponse. If we have no auxiliary variables, then using (14c), we can show that the test of ignorable nonresponse is equivalent to testing the set of hypotheses

$$H_{0k} = Corr_r\left(y_i^k,\ \phi_i\right) = 0,\ \ k = 1,\ 2,\ \ldots \tag{100}$$

where $Corr_r$ is the correlation under the response distribution.

The conventional t-statistic can be used to test this set of hypotheses. This test requires that the all moments of the distribution exist.

If the size of the response set is m, and response probabilities are unknown, then an estimate of $Corr_r\left(y_i^k, \phi_i\right)$ can be computed as

$$\hat{C}orr_r\left(y_i^k,\ \hat{\phi}_i\right) = \frac{\sum_{i=1}^{m}\left(y_i - \bar{y}_r\right)\left(\hat{\phi}_i - \bar{\hat{\phi}}\right)}{\sqrt{\sum_{i=1}^{m}\left(y_i - \bar{y}_r\right)^2}\sqrt{\sum_{i=1}^{m}\left(\hat{\phi}_i - \bar{\hat{\phi}}\right)^2}} \tag{101}$$

where $\bar{y}_r = m^{-1}\sum_{i=1}^{m} y_i$, $\hat{\phi}_i = 1/\hat{\psi}_i$ and $\bar{\hat{\phi}} = m^{-1}\sum_{i=1}^{m}\hat{\phi}_i$. So we don't need the values of the study or outcome variable for units in the nonrespondents set.

9.2 Kullback-Leibler Information Test

A new test for response ignorability we propose is based on the Kullback-Leibler information measure; see Kullback (1978). For instance, under the exponential response probability model, the condition

(99) implies that the test of response ignorability is equivalent to testing the null hypothesis

$$H_0 : f_r(y_i|\theta, \eta) = f_r(y_i|\theta) \text{ or } \eta = 0 \qquad (102)$$

against the alternative hypothesis

$$H_1 : f_r(y_i|\theta, \eta) \neq f_p(y_i|\theta \text{ or } \eta \neq 0 \qquad (103)$$

We can show, under the common mean population model, see Section 8.2, that the minimum discrimination information (for a single observation) from the response log- likelihood is given by

$$\begin{aligned} I\left(f_r : f_p\right) &= E_r\left[\log \frac{f_r\left(y_i|\theta, \eta\right)}{f_p\left(y_i|\theta\right)}\right] \\ &= \frac{\eta}{2}\left(\eta\sigma^2\right) \end{aligned} \qquad (104)$$

Note that $I(f_r : f_p)$ is a product of $\eta/2$ and the amount of change in the location parameter, and is equal zero if and only if $f_r(y_i|\theta, \eta) = f_p(y_i|\theta)$, that is, if $\eta = 0$.

Now let r denote the values of the response data set $y_1, \ldots, y_m$ of m independent and identically distributed observations. Then the estimate of the minimum discrimination information given in (104) is

$$\hat{I}\left(f_r : f_p\right) = \hat{I}\left(H_1 : H_0\right) = E_r\left[\log \frac{f_r\left(y_i|\theta, \eta\right)}{f_p\left(y_i|\theta\right)}\right]$$

$$= \frac{\tilde{\eta}}{2}\left(\tilde{\eta}\hat{\sigma}^2_{C,nign}\right) \qquad (105)$$

where $\tilde{\eta}$ and $\hat{\sigma}^2_{C,nign}$ are the appropriate estimators of η and σ^2 respectively; see equations (67) and (68). Asymptotically, under certain regularity conditions and under the null hypothesis H_0, $2\hat{I}\left(H_1 : H_0\right)$ given in (105) has an asymptotic chi-square distribution with one degree of freedom, see Kullback (1978, Section 5.5). Thus, asymptotically

$$\Pr\left(2\hat{I}\left(H_1 : H_0\right) \geq \chi^2_{2\alpha,1}\right) = \alpha \qquad (106)$$

10. CONCLUSIONS

In this paper we consider a new method of estimating the parameters of the superpopulation model for single-stage sampling from a finite population when the sampling design is noninformative and the response mechanism is nonignorable. We derive some new relationships between moments of the population distribution before sampling and the response and nonresponse distributions. Thus provides new justification for the broad use of probability-weighted estimators and pseudo likelihood estimator in estimating finite population parameters in case of ignorable nonresponse. We study Fisher information and confidence intervals under the response distribution. Furthermore we fit three population models, namely: common mean population model, simple ratio population model, and simple regression population model, under noninformative sampling design and under nonignorable nonresponse. In addition to the estimation problem we introduce new predictors of the finite population total for common mean population model, simple ratio population model, and simple regression population model. These new predictors take into account the nonignorable nonresponse. Thus, also provides new justification for the broad use of best linear unbiased predictors (model-based school) in predicting finite population parameters in case of ignorable nonresponse.

The main features of the present predictors and estimators are their behaviours in terms of the nonignorable nonresponse parameters. Also the use of the best linear unbiased predictors and estimators that ignore the nonignorable nonresponse yield biased predictors and bias estimators.

Finally, we introduce two new tests: conventional t-test, and the Kullback-Leibler information test for testing ignorability of missing value mechanism.

The paper is purely mathematical; it shows the role of missing value mechanism in adjusting various estimator, and predictors, for bias reduction, under different population models and under exponential response probabilities. Other modeling of conditional expectation of probability of response given the values of study variable can be studied in the same way.

I hope that the new mathematical results obtained and the issue of the role of sampling weights when the sampling design is noninformative (or ignorable) and the missing data mechanism is not missing at random (or informative or nonignorable) will encourage further theoretical, empirical and practical research in these directions.

ACKNOWLEDGEMENTS

The author is grateful to the coordinating editor Rajender Parsad and to the referees for their valuable comments.

REFERENCES

Bethlehem, J.G. (1988). Reduction of nonresponse bias through regression estimation. *J. Official Statist.,* **4,** 251-260.

Binder, D.A. (1983). On the variances of asymptotically normal estimators from complex surveys. *Intern. Statist. Rev.,* **51(2)**, 79-292.

Chambers, R. and Skinner, C. (2003). *Analysis of Survey Data*. John Wiley, New York.

Cobben, F. (2009). *Nonresponse in Sample Survey – Method for Analysis and Adjustment*. Statistics Netherlands.

Eideh A.H. (2009). On the use of the sample distribution and sample likelihood for inference under informative probability sampling. *DIRASAT (Natural Science)*, **36(1)**, 18-29.

Eideh, A.H. (2010). Analytic inference of complex survey data under informative probability sampling. *Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I. The Islamic Countries Society of Statistical Sciences, Lahore, Pakistan*. Eds.: Zeinab Amin and Ali S. Hadi. The American University in Cairo, 507-536.

Eideh, A.H. and Nathan, G. (2006). Fitting time series models for longitudinal survey data under informative sampling. *J. Statist. Plann. Inf.,* **136(9)**, 3052-3060. [Corrigendum, 137 (2007), 628].

Eideh, A.H. and Nathan, G. (2009). Two-stage informative cluster sampling with application in small area estimation. *J. Statist. Plann. Inf.,* **139**, 3088-3101.

Kullback, S. (1978). *Information Theory and Statistics*. Gloucester, Mass, Peter Smith.

Little, R.J.A. (1982). Models for nonresponse in sample surveys. *J. Amer. Statist. Assoc.*, **77**, 237-250.

Little, R.J.A. and Rubin, D.B. (2002). *Statistical Analysis with Missing Data*. Wiley, New York.

Pfeffermann, D., Krieger, A.M. and Rinott, Y. (1998). Parametric distributions of complex survey data under informative probability sampling. *Statistica Sinica*, **8,** 1087-1114.

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

Pfeffermann, D. and Sverchkov, M. (2003). Fitting generalized linear models under informative sampling. In: *Analysis of Survey Data*, Eds. C. Skinner and R. Chambers. Wiley, New York, 175-195.

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

Särndal, C.E. and Lundstorm, S. (2005). *Estimation in Surveys with Nonresponse*. John Wiley, New York.

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

Särndal C.E. and Swensson B. (1987). A general view of estimation for two phases of selection with applications to two-phase sampling and nonresponse. *Intern. Statist. Rev.*, **55**, 279-294.

Schafer, J.L. (1997). *Analysis of Incomplete Multivariate Data*. Chapman and Hall, London.

Schaible, W.L. (1983). Estimation of finite population totals from incomplete sample data: Prediction approach. In: *Incomplete Data in Sample Surveys*, Vol 3, Proceedings of the symposium, Eds. Maow, W.G., and Olkin, I. Academic, New York, 131-141.

Skinner, C.J., Holt, D. and Smith, T.M.F (eds.) (1989). *Analysis of Complex Surveys*. Wiley, New York.

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