

# Methods Used in Reduction of Bias Arising from Nonresponse

Müge Borazan ÇELİKBIÇAK<sup>1</sup>, Özlem Ege ORUÇ<sup>2</sup>

## Abstract

Various types of survey errors, especially nonresponse errors, may seriously deteriorate data quality. Nonresponse has more one reason to worry about its harmful effects on the survey estimates. So, nonresponse and methods dealing with nonresponse have increasingly become a standard part of survey sampling. A nonresponse occurs in a survey when, for any reason, a selected unit does not respond. The usual methods of estimation in the presence of nonresponse give biased results. Because of this special estimation techniques are required to deal with the problem. Imputation and reweighting are two standard methods provided by the literature for treating nonresponse. In the recent years, scientist became increased to concern with the calibration approach to reweighting method in the presence of nonresponse. The calibration approach generates the final weights which are as close as possible to specified design weights, while respecting known auxiliary population totals or unbiased estimates of these totals. This calibration procedure requires the formulation of a suitable auxiliary vector, through a selection from a possible larger set of auxiliary variables.

In this study standard methods for the reduction of bias and errors arising from nonresponse are explained. The calibration approach is examined as theoretically and simulation is performed by macro generated in C++ programming language to study how alternative specifications of the auxiliary vector affect the quality of estimators derived by the calibration technique. In the application, the calibration estimators and quality measures such as relative bias, variance are computed and interpreted for the population total.

**Keywords:** *Nonresponse, nonresponse bias, calibration, auxiliary information, nonresponse adjustment, weighting.*

---

<sup>1</sup> Jandarma Genel Komutanlığı Bakanlıklar, Ankara muge.borazan@deu.edu.tr

<sup>2</sup> Dokuz Eylül Üniversitesi Fen Edebiyat Fakültesi, İstatistik Bölümü, ozlem.ege@deu.edu.tr

## Özet

Anket arařtırmalarında karşılaşılan çeřitli problemlerden biri olan yanıtlanmama hatası veri kalitesini ciddi şekilde bozmaktadır. Yanıtlanmama, anket arařtırmalarından elde edilecek tahminler üzerindeki tehlikeli etkilerinden dolayı kaygı duyulan bir durumdur. Bu nedenle yanıtlanmama sorununu giderecek yöntemler örneklemenin standart bir bölümünü oluřturmaktadır. Yanıtlanmamanın varlıđı durumunda ideal kořullar için kullanılan tahmin yöntemleri yanlı sonuçlar vermektedir. Bu sorunun üstesinden gelmek için özel tahmin tekniklerinin kullanılması gerekmektedir. Yerine Atama (ikame) ve Yeniden Ađırlıklandırma yanıtlanmama sorunu ile karşılaşıldığında kullanılan ve literatürde de belirtilen iki standart yöntemdir. Son yıllarda, bilim adamları, yanıtlanmamanın varlıđında, yeniden ađırlıklandırma yönteminde kalibrasyon yaklaşımını daha sık kullanmaktadırlar. Çünkü yardımcı bilginin etkin kullanımında kalibrasyon uyarlanabilir bir yaklaşımdır. Kalibrasyon yaklaşımı, yardımcı bilgi kitle toplamını veya yansız tahminin bilinmesini gerektirirken, tasarım ađırlıklarına olabildiğince yakın olması gereken son ađırlıkları yaratır. Kalibrasyon yöntemi, olası yardımcı deđiřken kümesinden seçim ile oluřturulan uygun yardımcı deđiřken vektörünün formülasyonunu gerektirir.

Bu çalışmada, yanıtlanmadan kaynaklanan yanlılıđın ve hataların azaltılabilmesi için kullanılan yöntemler tanıtılmıřtır. Bu yöntemlerden biri olan Kalibrasyon yaklaşımı teorik olarak incelenmiřtir. Yardımcı deđiřken vektörünün deđiřik durumlarının kalibrasyon tekniđi ile hesaplanan tahmin edicilerin kalitesini nasıl etkilediđini görebilmek amacıyla C++ programlama dili kullanılarak bir benzetim çalışması yapılmıřtır. Uygulamada kitle toplamı için kalibrasyon tahminleri, görelilikler, varyanslar hesaplanarak yorumlanmıřtır.

**Anahtar Kelimeler:** Cevaplanmama, cevaplanmama yanlılıđı, kalibrasyon, yardımcı bilgi, cevaplanmama düzeltilmesi, ađırlıklandırma.

## 1. Introduction

The results of a sample survey are affected by many kinds of errors, one of the most important sources being nonresponse. Nonresponse has long been a matter of concern in survey sampling. Nonresponse means failure to obtain a measurement on one or more study variables for one or more elements  $k$  selected for the survey. The main problem caused by nonresponse is that estimators of population characteristics must be assumed to be biased unless convincing evidence to the contrary is provided. It is for this reason that survey researches are concerned about nonresponse.

In the best of surveys, nonresponse occurs, and special estimation techniques are required to deal with the problem. The principal methods for nonresponse adjustment are *reweighting* and *imputation*. Reweighting entails altering the weights of the respondents, compared to the weights that would have been used in the case of full

response. Imputation entails replacing missing values by proxy values. In recent years, there has been a great interest in the calibration approach to reweighting. The calibration procedure generates final weights which are as close as possible to specified initial (design) weights, while respecting known auxiliary population totals or unbiased estimates of these totals.

The aim of this study is investigation of standard methods for the reduction of bias arising from nonresponse and developing a macro using C++ computer programming language for one of this method to see reduction of nonresponse bias for estimators.

## 2. Nonresponse

By nonresponse it is meant that the desired data are not obtained for the entire set of elements  $s$ , designated for observation. By *full response* in the survey is meant that, after data collection and edit, the available data consist, for every  $k \in s$ , of a complete  $q$ -vector of observed values,  $y_k = (y_{1k}, \dots, y_{jk}, \dots, y_{qk})$ , these values form a data matrix of dimension  $n_s \times q$ , with no value missing. (Särndal, Swensson, and Wretman, 1991)

Nonresponse has more one reason to worry about its harmful effects on the survey estimates. The bias often increases with the rate of nonresponse. It is very difficult to get objective measures of the bias, but it is relatively simple to quantify the extent of the nonresponse. (Hunsen and Hurwitz, 1946) Different measures of the nonresponse are usually found in the quality declarations. A number of descriptive measures are used for the response, or its complement, the nonresponse. Two types of missing information for an element  $k$  can be distinguished from which response is solicited.

Item Nonresponse : The element  $k$  is an *item nonresponse* element if at least one, but not at all  $q$ , components of the vector  $y_k = (y_{1k}, \dots, y_{jk}, \dots, y_{qk})$  are missing. For example, the respondent returns a partially filled in questionnaire, or an interview results in responses to some but not all questions. (Särndal, Swensson, and Wretman, 1991)

Unit Nonresponse : The element  $k$  is a *unit nonresponse* element if the entire vector of  $y$  -values,  $\mathbf{y}_k = (y_{1k}, \dots, y_{jk}, \dots, y_{qk})$ , is missing. An example is when the respondent fails to return the questionnaire, even after one or more reminders, or if he or she refuses to participate in a personal interview. (Särndal, Swensson, and Wretman, 1991)

## **2.1 Sources of Nonresponse**

Nonresponse refers to many sources of failure to obtain observations on some elements selected and designated for the sample. If accurate accounts are kept of all eligible elements that fall into the sample, the nonresponse rate can be measured. These are necessary for understanding the sources of nonresponse, for its control and reduction, for predicting it in future surveys, and for estimating its possible effects on the surveys. If the many possible sources of nonresponse are sorted into a few meaningful classes, these aims can be better served. A good classification of nonresponse depends on the survey situation. (Kish, 1969)

In the literature, various classification types are seen about sources of nonresponse. For example, while Moser and Kalton (1971) evaluated that the states “not at homes” and “out of the city” in different classes, Kish (1969) classified two of them in the same group. Kish (1969) classified the sources of nonresponse as follows: *Not at homes, Refusals, Incapacity or inability, Not found, Lost schedules*. These categories refer to nonresponse involving the entire interview or questionnaire. Platek (1977) classifies sources of nonresponse as related to; *Survey content, Methods of data collection, Respondent characteristic. Survey content, Time of survey, Interviewers, Data-collection method, Questionnaire design, Respondent burden, Incentives and disincentives, Follow-up* are some factors that may influence response rate and data accuracy classified by Lohr (1999).

## 2.2 The Importance of Auxiliary Information

Recent years have seen theoretical developments and increased use of methods that take account of substantial amounts of auxiliary information. The key to successful nonresponse adjustment lies in the use of “strong” auxiliary information. Such use will reduce both nonresponse bias and the variance. Auxiliary information can be used both at the design stage in constructing the sampling design and at the estimation stage in constructing the estimators. In the reweighting procedure for nonresponse second type of usage is valid. An auxiliary vector is made up of one or more auxiliary variables. There are two important steps in the process leading to the form of the auxiliary vector that will be ultimately used in the estimation. These are:

- a. Making an inventory of potential auxiliary variables;
- b. Selecting and preparing the most suitable of these for entry into the auxiliary vector.( Lundström and Särndal, 2001)

## 3. Methods for the Reduction of Nonresponse Errors

The usual methods of estimation in the presence of nonresponse give biased results. Because of this special estimation techniques are required to deal with the problem. *Imputation* and *reweighting* are two standard methods provided by the literature for treating nonresponse.

***Imputation*** : Imputation is the procedure whereby missing values for one or more study variables are ‘filled in’ with substitutes. These substitutes can be constructed according to some rule, or they can be observed values but for elements other than the nonrespondents. Thus imputed values are artificial and they contain error. Imputation error is similar to measurement error in that the true value is not recorded. (Lundström and Särndal, 2001) Imputed values can be classified into three major categories:

- (1) values constructed with the aid of a statistical prediction rule;
- (2) values observed not for the nonresponding elements themselves, but for responding elements

(3) values constructed by expert opinion or ‘best possible judgment’.

The imputed values must come as close as possible to the true unobserved values for which they are substitutes. Because of this, the construction of imputed values should be carried out with professional care. There are two frequently used approaches for imputation, both leading to rectangular data matrices, namely the ITIMP-approach and the UNIMP-approach.

(i) ITIMP-approach: Imputation is used to treat the item nonresponse only. In this procedure, values are imputed for the  $m$  elements for which at least one but not all  $y$  values are missing. The resulting rectangular data matrix has the dimensions  $m$  by  $J$  where  $J$  is the number of  $Y$  variables. Reweighting is then applied to compensate for the unit nonresponse.

(ii) UNIMP-approach: Imputation is used for both item nonresponse and unit nonresponse. In this procedure, values are imputed for all elements having at least one  $y$  value missing. The resulting completed rectangular data matrix has the dimensions  $n$  by  $J$ , where  $n$  is the sample size. There is no nonresponse weight adjustment.

**Reweighting** : Reweighting entails altering the weights of the respondents, compared to the weights that would have been used in the case of full response. Since observations are lost by nonresponse, reweighting will imply increased weights for all, almost all, of the responding elements. Reweighting is treated by Lundström and Särndal (2001) with the calibration approach, which has the favourable property of incorporating most “Standard” methods found in the different places in the literature.

**3.1 Calibration Approach:** The calibration estimators derived by Deville and Särndal (1992) are a family of estimators appealing a common base of auxiliary information. A calibration estimator uses calibrated weights, which are as close as possible, according to the given distance measure, to the original sampling design weights,  $d_k$ , while also respecting a set of constraints, the calibration equation. They discuss the merits of different metrics for the distance between  $w_k$  and  $d_k$ . For

every distance measure there is corresponding set of calibrated weights and a calibration estimator.

Lundström and Särndal (1999) can be practically certain that the nonresponse is not the result of a simple random selection mechanism so they try to adjust for the selection bias at the estimation stage. They suggests a simple and a unified approach is called calibration to the use of auxiliary information both the sampling error and the nonresponse bias in a survey. When population totals are used, the resulting point estimators are consistent in the sense that the final weights give perfect estimates when applied to each variable. This approach requires neither a response model nor a regression model but which nevertheless guarantees effective use of auxiliary information.

The calibration procedure generates final weights which are as close as possible to specified initial (design) weights , while respecting known auxiliary population totals or unbiased estimates of these totals. Calibration is used by Lundström and Särndal, 2001 as a main tool for nonresponse. This calibration approach requires the formulation of a suitable auxiliary vector, through a selection from a possible larger set of available auxiliary variables. This step follows a few basic and simple principles. The next step is computational.

The calibration approach has only a single computational step, in which the calibrated weights are produced. It requires no seperate modelling of a nonresponse mechanism. For these reasons, the calibration approach, is better suited for a routine treatment of nonresponse in organization.

### 3.1.1 Point Estimation Under Calibration Approach

It is first considered that the finite population of  $N$  elements  $U = \{1, \dots, k, \dots, N\}$ . The objective is to estimate the population total  $Y = \sum_U y_k$  , where  $y_k$  is the value of a study variable,  $y$ , for the  $k$ th element. The sample of size  $n$ ,  $s$ , drawn from  $U$  with the probability  $p(s)$ . When nonresponse occurs, the response set  $r$  of size  $m$  is obtained,

where  $r \subseteq s$  and  $m \leq n$ . It is concerned here with surveys with nonresponse, so  $y_k$  values are available only for the elements  $k$  in the response set  $r$ , a subset of the sample. Then, whatever the estimation technique, there will be some bias. Desirable properties of the chosen estimator are now: *a small nonresponse bias, a small total variance*. The calibration estimator is, formed as a linearly weighted sum of the observed  $y_k$  values. It is defined by

$$\hat{Y}_w = \sum_r w_k y_k \quad (1)$$

where  $w_k = d_k v_k$  and

$$v_k = 1 + q_k \left( \sum_U x_k - \sum_r d_k x_k \right) \left( \sum_r d_k c_k x_k x_k' \right)^{-1} x_k, \quad k \in r. \quad (2)$$

For a successful reduction of both the sampling error and the response bias, strong auxiliary information is a prerequisite. It is assumed that there exists an auxiliary vector,  $\mathbf{x}$ , containing such information. This vector's value for the  $k$ th element is denoted  $\mathbf{x}_k$ . Lundström and Särndal (1999) define the two "information levels" called Info-S and Info-U and they classified two information level in the following :

- a) Info-S:  $\mathbf{x}_k$  is known for all  $k \in s$ ,
- b) Info-U:  $\sum_U \mathbf{x}_k$  is known and moreover  $\mathbf{x}_k$  is known for all  $k \in s$ .

After having specified the auxiliary information, calibrated weights, denoted  $w_k$ , are computed and the estimator  $\hat{Y}_w = \sum_r w_k y_k$  of  $Y$  is constructed.

The principle behind the derivation that leads to calibrated weights  $w_k$  is to minimize a function measuring the distance between the old weights,  $d_k$ , and the new weights,  $w_k$ , subject to the calibration equation



$$\sum_r d_k v_k x_k = \sum_U x_k . \quad (3)$$

The calibrated weights are “as close as possible” with respect to the given distance measure,  $\sum_r (w_k - d_k)^2 / d_k q_k$ , to the design weights  $d_k$ , and they ensure consistency with the known auxiliary variable totals.

Calibration estimator for information at the sample level,  $\sum_s d_k x_k$ , is

$$\hat{Y}_{ws} = \sum_r w_{ks} y_k \quad (4)$$

where  $w_{ks} = d_k v_{sk}$

$$v_{sk} = 1 + \left( \sum_s d_k x_k - \sum_r d_k x_k \right) \left( \sum_s x_k x'_k d_k q_k \right)^{-1} x'_k q_k . \quad (5)$$

Calibration estimator for information at the sample level,  $\sum_U x_k$ , is

$$\hat{Y}_{wU} = \sum_r w_{kU} y_k$$

where  $w_{kU} = d_k v_{Uk}$

$$v_{Uk} = 1 + \left( \sum_U x_k - \sum_r d_k x_k \right) \left( \sum_s x_k x'_k d_k q_k \right)^{-1} x'_k q_k . \quad (7)$$

### 3.1.2 Examples of Calibration Estimators

The Simplest Auxiliary vector: The simplest formulation of the auxiliary vector is  $x_k = 1$  for all k. This vector recognizes no differences among elements. Specifying also  $q_k = 1$  for all k. Then the calibration weight, (7), for all k,

$$v_k = \frac{n}{m} \quad (8)$$

and the calibration estimator (6) becomes

$$\hat{Y}_w = \frac{N}{m} \sum_r y_k = \hat{Y}_{EXP} \quad (9)$$

As seen from that, when auxiliary vector is  $\mathbf{x}_k=1$ , the calibration estimator becomes the traditional expansion estimator. (Lundström and Särndal,1999)

One-way classification: In this case the target population is divided into non-overlapping and exhaustive groups,  $U_p$ ,  $p = 1, \dots, P$ , based on a specified classification criterion, for example age by sex groups. The auxiliary vector form is

$$\mathbf{x}_k = (\gamma_{1k}, \dots, \gamma_{pk}, \dots, \gamma_{Pk})' \text{ where}$$

$$\gamma_{pk} = \begin{cases} 1 & \text{if } k \in U_p \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

The component of the key vector totals are denoted as follows

$$\sum_U \mathbf{x}_k = (N_1, \dots, N_p, \dots, N_P)', \sum_s \mathbf{x}_k = (n_1, \dots, n_p, \dots, n_P)' \text{ and}$$

$$\sum_r x_k = (m_1, \dots, m_p, \dots, m_P)' \text{ When } q_k=1 \text{ for all } k, \text{ the calibration weight is}$$

$$v_k = \frac{N_p n}{Nm_p} \quad (11)$$

for  $k \in r_p$ , and the calibration estimator (6) becomes

$$\hat{Y}_w = \sum_{p=1}^P N_p \bar{y}_{r_p} = \hat{Y}_{PST} \quad (12)$$

where  $\bar{y}_{r_p} = \frac{1}{m_p} \sum_{r_p} y_k$  and  $m_p$  is the number of respondents in group  $p$ . This estimator commonly called the poststratified estimator and denotes as  $\hat{Y}_{PST}$ . When knowledge of the auxiliary vector  $\mathbf{x}_k = (\gamma_{1k}, \dots, \gamma_{pk}, \dots, \gamma_{pk})'$  is limited to the elements of the sample  $s$ , the calibration estimator (4) becomes

$$\hat{Y}_{ws} = \sum_{p=1}^P \hat{N}_p \bar{y}_{r_p} = \hat{Y}_{WCE} \quad (13)$$

with  $\hat{N}_p = \frac{N}{n} n_p$  and  $n_p$  is the number of sampled elements in group  $p$ . This estimator known as the weighting class estimator denoted as  $\hat{Y}_{WCE}$ . (Lundström and Särndal, 1999)

A single quantitative variable: It is assumed that a quantitative auxiliary variable  $\mathbf{x}_k$  is available. For example, the number of employees of enterprise  $k$  in a business survey,  $k = 1, \dots, N$ . It is assumed that its population total,  $\sum_u x_k$ , known. If

this is the only auxiliary variable, the auxiliary vector is uni-dimensional,  $\mathbf{x}_k = x_k$ .

When  $q_k$  is specified as  $q_k = x_k^{-1}$ , then the calibration estimator (4) is

$$\hat{Y}_w = \left( \sum_U x_k \right) \frac{\bar{y}_r}{\bar{x}_r} = \hat{Y}_{RA} \quad (14)$$

where  $\bar{y}_r = \frac{1}{m} \sum_r y_k$  and  $\bar{x}_r = \frac{1}{m} \sum_r x_k$ . This estimator has the well known

form of a ratio estimator and denoted as  $\hat{Y}_{RA}$ . (Lundström and Särndal, 1999)

With the same information it can be alternatively formulated the auxiliary vector as  $\mathbf{x}_k = (1, x_k)'$ . When  $q_k = 1$  for all  $k$ , the calibration estimator (4) becomes,

$$\hat{Y}_w = N \left\{ \bar{y}_r + (\bar{X} - \bar{x}_r) \hat{B} \right\} = \hat{Y}_{REG} \quad (15)$$

where  $\bar{X} = \frac{1}{N} \sum_U x_k$  and  $\hat{B} = \left[ \sum_r y_k x_k - \frac{1}{m} \sum_r y_k \right] / \left[ \sum_r x_k^2 - \frac{1}{m} \left( \sum_r x_k \right)^2 \right]$ . This

estimator has the well known form of a regression estimator and denoted as  $\hat{Y}_{REG}$ .

One –way classification and a quantitative variable: The auxiliary information concerns a  $P$ -valued categorical variable and a quantitative variable,  $x$ , that may be an indicator of the size of an element. It is assumed that every sampled element  $k$  is placed into the appropriate group, that its value  $x_k$  is known. And for each group the size  $N_p$  is known and the  $x$  total,  $\sum_{U_p} y_k$ . There are more than one way to use this information.

The auxiliary vector is defined as

$$\mathbf{x}_k = (\gamma_{1k} x_k, \dots, \gamma_{pk} x_k, \dots, \gamma_{pk} x_k)' \quad (16)$$

where  $\gamma_{pk}$  is defined by (10). It leads to a well known estimator, because if  $q_k = x_k^{-1}$  is let (4) becomes

$$\hat{Y}_w = \sum_{p=1}^P \left( \sum_{U_p} x_k \right) \frac{\bar{y}_{r_p}}{\bar{x}_{r_p}} = \hat{Y}_{SEPR} \quad (17)$$

Where  $\bar{y}_{r_p} = \frac{1}{m_p} \sum_{r_p} y_k$  and  $\bar{x}_{r_p} = \frac{1}{m_p} \sum_{r_p} x_k$ . Thus,  $\hat{Y}_{SEPR}$  has the form of a

separate ratio estimator. Instead, the auxiliary vector is formulated as

$$\mathbf{x}_k = (\gamma_{1k}, \dots, \gamma_{pk}, \dots, \gamma_{pk}, \gamma_{1k} x_k, \dots, \gamma_{pk} x_k, \dots, \gamma_{pk} x_k)'$$

Then if  $q_k = 1$  for all  $k$ , the estimator (4) becomes

$$\hat{Y}_w = \sum_{p=1}^P N_p \left\{ \bar{y}_{r_p} + (\bar{X}_p - \bar{x}_{r_p}) \hat{B}_p \right\} = \hat{Y}_{SEPREG} \quad (18)$$

with  $\bar{X}_p = \frac{1}{N_p} \sum_{U_p} x_k$  and  $\hat{B}_p = \frac{Cov_{xyr_p}}{S_{xr_p}^2}$ . Covariance and variance term is

$$Cov_{xyr_p} = \frac{1}{m_p - 1} \left[ \sum_{r_p} y_k x_k - \frac{1}{m_p} \sum_{r_p} y_k \sum_{r_p} x_k \right] \quad (19)$$

and

$$S_{xr_p}^2 = \frac{1}{m_p - 1} \left[ \sum_{r_p} x_k^2 - \frac{1}{m_p} \left( \sum_{r_p} x_k \right)^2 \right]. \quad (20)$$

The estimator (18) is another well known form separate regression estimator. (Lundström and Särndal, 1999)

Several special cases of the general calibration estimator was obtained, corresponding to different formulations of the auxiliary vector  $\mathbf{x}_k$  and the factor  $q_k$ . Lundström and Särndal (1999) revisited these estimators with the purpose of showing how the theoretical result can guide the selection of relevant auxiliary information.

#### 4. Application

In the presence of available auxiliary information the calibration approach is flexible for the reduction of nonresponse bias. In application, it is given through the simulation that how alternative specifications of the auxiliary vector  $\mathbf{x}_k$  affect the quality of estimators derived by the calibration technique.

At the beginning of the application, a population in the size of  $N = 1000$  was generated by the Minitab macro program. Then, 100 random samples were drawn from this population with sample size of  $n = 400$ . Throughout the application section the only one type of parameter estimated is a total for the entire population. The SRS design is used for drawing samples from population. In order to both computation of point estimators and quality measures that relative bias and variance for the population total, generated by different  $\mathbf{x}_k$  vector specifications, a macro was written using C++ programming language.

The population was generated by the Minitab macro program with the size of  $N = 1000$ . The study variable  $y_k$  is a numerical variable measuring such as expense. The first auxiliary variable,  $\Gamma_k$ , is categorical, indicating such as one out of four possible group or regions. The second auxiliary variable, denoted  $x_k$ , is numerical too and defined such as revenues. In the simulation the square root of  $x_k$  was used as second auxiliary variable.

It is assumed that 850 of the 1000 data of population was responded. So the population used in the simulation consist of 850 responding elements. The response rate was thus 85 per cent. From the population consisting of the 850 responding elements, repeated simple random samples are drawn. 100 random samples with the size of  $n = 400$  was selected by the following Minitab macro program 'SAMPLING.MTW'.

SAMPLING.MTW

```
SAMPLE 400 C1 CK1
LET K1=K1+1
END
```

The macro was completed when the Minitab macro command was written as follows:

```
MTB > LET K1=2
MTB > EXEC 'C:\MTBWIN\DATA\SAMPLING.MTW' 100
```

**Table 1.** Some key characteristics of the study variable  $Y$ .

Characteristics for the entire population		Group ( Regions)			
		1	2	3	4
Total ( $Y$ )	1,282,671	353,209	306,795	305,314	317,353
Mean ( $\bar{Y}$ )	1,509	1,542	1,511	1,511	1,469
Number of elements( $N$ )	850	229	203	202	216

In the previous section, examples of calibration estimators obtained by the specifications of auxiliary vectors for the population total were given. These point estimators were computed through the C++ programming language. For response set  $j$  ( $j=1, \dots, 100$ ),  $\hat{Y}_{w(j)}$  is calculated, where  $\hat{Y}_{w(j)}$  is the value of the point estimator  $\hat{Y}_w$  for response set  $j$ . Here,  $\hat{Y}_w = \hat{Y}_{ws}$  given by (4) for Info-S, and  $\hat{Y}_w = \hat{Y}_{wU}$  given by (6) for Info-U. And two quality measures were studied for see the affects of auxiliary vectors on estimators. The first is simulation relative bias in per cent,

$$RB_{SIM}(\hat{Y}_w) = 100 \frac{[E_{SIM}(\hat{Y}_w) - Y]}{Y} \quad (21)$$

where  $E_{SIM}(\hat{Y}_w)$  is the simulation expectation of  $\hat{Y}_w$  denote as,

$$E_{SIM}(\hat{Y}_w) = \sum_{j=1}^{100} \frac{\hat{Y}_{w(j)}}{100} \quad (22)$$

and the second is simulation variance,

$$V_{SIM}(\hat{Y}_w) = \sum_{j=1}^{100} \frac{[\hat{Y}_{w(j)} - E_{SIM}(\hat{Y}_w)]^2}{99} \quad (23)$$

The simulation results are given in Table 2.

**Table 2.** Simulation relative bias and simulation variance for different point estimators, SRS with  $n = 400$

Auxiliary vector	Info Level	Estimators	$RB_{SIM}(\hat{Y}_w)$	$V_{SIM}(\hat{Y}_w) \times 10^6$
1	U-S	$\hat{Y}_{EXP}$	6.96	9.105
$(\gamma_{1k}, \dots, \gamma_{pk}, \dots, \gamma_{pk})'$	U	$\hat{Y}_{PST}$	0.13	6.346
	S	$\hat{Y}_{WCE}$	0.21	10.342
$x_k$	U	$\hat{Y}_{RA}$	1.57	1.614
	S	$\hat{Y}_{RA}$	5.97	13.912
$(1, x_k)'$	U	$\hat{Y}_{REG}$	0.09	2.492
	S	$\hat{Y}_{REG}$	1,37	3.556
$(\gamma_{1k}x_k, \dots, \gamma_{pk}x_k, \dots, \gamma_{pk}x_k)'$	U	$\hat{Y}_{SEPPRA}$	0.10	1.372
	S	$\hat{Y}_{SEPPRA}$	0.15	4.358
$(\gamma_{1k}, \dots, \gamma_{pk}, \dots, \gamma_{pk}, \gamma_{1k}x_k, \dots, \gamma_{pk}x_k, \dots, \gamma_{pk}x_k)'$	U	$\hat{Y}_{SEPPREG}$	0.11	1.342
	S	$\hat{Y}_{SEPPREG}$	0.13	2.686

### 5. Conclusion

The aim of this paper is to introduce methods for the reduction of bias and errors arising from survey with nonresponse. In the recent years, scientist became increased to concern with the calibration approach to reweighting method in the presence of nonresponse. Because of this reason, we interested in especially reweighting method with calibration approach. Calibration procedure requires the formulation of a suitable auxiliary vector, through a selection from a possible larger set of auxiliary variables. Different specifications of auxiliary variables leads to different form of calibration estimators.



In application, we studied the point estimators generated by different  $\mathbf{x}_k$  vector specifications, namely,  $\mathbf{x}_k = 1$  for all  $k$ ,  $\mathbf{x}_k = (\gamma_{1k}, \dots, \gamma_{pk}, \dots, \gamma_{pk})'$ ,  $\mathbf{x}_k = x_k$ ,  $\mathbf{x}_k = (1, x_k)'$ ,  $\mathbf{x}_k = (\gamma_{1k}x_k, \dots, \gamma_{pk}x_k, \dots, \gamma_{pk}x_k)'$ ,  $\mathbf{x}_k = (\gamma_{1k}, \dots, \gamma_{pk}, \dots, \gamma_{pk}, \gamma_{1k}x_k, \dots, \gamma_{pk}x_k, \dots, \gamma_{pk}x_k)'$

Application shows that the calibration approach is highly flexible in its use of auxiliary information for the reduction of nonresponse bias. When  $\mathbf{x}_k = 1$  there is no auxiliary information. In this reason simulation results for estimator  $(\hat{Y}_{EXP})$  has more bias and variance. We compared the other estimators  $(\hat{Y}_{PST}, \hat{Y}_{RA}, \hat{Y}_{REG}, \hat{Y}_{SEPPRA}, \hat{Y}_{SEPREG})$ , which generated by different  $\mathbf{x}_k$  vector specifications, with  $\mathbf{x}_k = 1$ . We would expect the nonresponse bias to diminish with increasing amounts of auxiliary information, and this is confirmed by Table 2.

For instance, as shown in Table 2 when auxiliary information are not used its calculated that the relative bias of estimator is  $[RB_{SIM}(\hat{Y}_{EXP}) = 6.96\%]$  and variance is  $[V_{SIM}(\hat{Y}_{EXP}) = 9,105 \times 10^6]$ . A large simulation relative bias is observed for the uninformative specification  $\mathbf{x}_k = 1$ , for all  $k$ . The bias drops when we use more informative auxiliary vector,  $\mathbf{x}_k = (\gamma_{1k}, \dots, \gamma_{pk}, \dots, \gamma_{pk})'$ , and it's relative bias and variance are obtained  $[RB_{SIM}(\hat{Y}_{PST}) = 0.13\%]$ ,  $[V_{SIM}(\hat{Y}_{PST}) = 6,346 \times 10^6]$ . However the bias is nearly eliminated when the numerical variable  $x_k$  is also included in the auxiliary vector. The auxiliary vectors with  $x_k$  are  $\mathbf{x}_k = (\gamma_{1k}x_k, \dots, \gamma_{pk}x_k, \dots, \gamma_{pk}x_k)'$ , and  $\mathbf{x}_k = (\gamma_{1k}, \dots, \gamma_{pk}, \dots, \gamma_{pk}, \gamma_{1k}x_k, \dots, \gamma_{pk}x_k, \dots, \gamma_{pk}x_k)'$ . Their relative bias and variances are much smaller than other auxiliary vectors that used in estimation stage. This denotes that using the auxiliary information decreases the nonresponse bias and variance.

Calibration approach may play considerable role in reduction of nonresponse bias and it is better suited for a routine treatment of nonresponse in organization.

## REFERENCES

- Çelikbıçak, B.M. 'Methods Used In Reduction of Bias Arising From Nonresponse.(2006),D.E.Ü.,Y.Lisans Tezi
- Deville, J.C., & Särndal, C. E., 'Calibration estimators in survey sampling' *Journal of the American Statistical Association* 87(1992), pp.376-382.
- Hansen, M.H., Hurwitz, W.N., 'The problem of nonresponse in sample surveys', *Journal of American Statistical Association*, 41(1946), pp.517-529.
- Kish, L., *Survey Sampling*. (New York: Wiley, 1969).
- Kalton, G. and Cervantes, I.F., 'Weighting methods', *Journal of Official Statistics*, 19 (2) (2003), pp.81-97.
- Lundström, S., and Särndal, C. E., 'Calibration as a standard method for treatment of nonresponse', *Journal of Official Statistics*, 15(1999),pp. 305-327.
- Lundström, S., *Calibration as a standard method for treatment of nonresponse* (Doctoral dissertation, Stockholm University, 1997)
- Lundström, S. and Särndal, C. E., *Estimation in the presence of nonresponse and frame imperfections* (Printed in Sweden, SCB-tyrck,Örebro, 2001)
- Lohr, S. L., *Sampling: Design and Analysis* (Duxbury Press, 1999).
- Moser, C.A. and Kalton, G. (1971) *Survey Methods in Social Investigation* (second edition). Aldershot: Gower.
- Platek, R. (1977), Some factors affecting non-Response, *Survey Methodology*,3,191–214.
- Särndal, C. E., Estevao, V. M., 'A functional form to calibration', *Journal of Official Statistics*, 16 (4) (2000), pp.379-399.
- Särndal, C. E., Swensson, B., Wretman, J., *Model Assisted Survey Sampling*. (New York: Springer-Verlag, 1991)
- Teksoy, N., 'Cevaplamama hatalarının ve ikamelerin tahminler üzerindeki etkileri' *DİE Uzmanlık Tezi*, (1991).
- Théberge, A., 'Extensions of calibration estimators in survey sampling' *Journal of the American Statistical Association*, 94 (446) (1999), pp. 635-644.

