Hacettepe Journal of Mathematics and Statistics
Volume 45 (4) (2016), 1297 – 1306

Generalized multi-phase regression-type estimators under the effect of measuemnent error to estimate the population variance

Saadia Masood*†, Javid Shabbir‡

Abstract

In this article, we suggest some regression-type estimators for the estimation of finite population variance using multi-variate auxiliary information under multi-phase sampling schemes when measurement error (ME) contaminates the study variable. An empirical study is also carried out to judge the merits of proposed estimators.

Keywords: Multi-phase sampling, measurement error, mean square error, efficiency.

2000 AMS Classification: 62D05

Received : 28.01.2014 *Accepted* : 06.04.2015 *Doi* : 10.15672/HJMS.201612116352

1. Introduction

In sample surveys, it is customary to exploit the auxiliary information to enhance the precision of estimators. Ratio and regression estimators provide one type of example. Sometimes the sample units are chosen with probability proportionate to some measure of size based on the auxiliary variate. In all these cases it is information on just one auxiliary variate that is used for reasons of sample selection or estimation. Pretty often we take information on several variates and it may be considered important to make use of the whole of the available material to improve the precision of at least some of the key items in the survey (see Raj [10]). Isaki [7] has discussed multi-variate ratio estimators to estimate finite population variance S_y^2. Singh and Solanki ([16], [17]) and Solanki and Singh [19] proposed the procedure for variance estimation using auxiliary information under simple random sampling.

Two-phase sampling of a finite population occurs when a sample from the population is itself sampled, with the goal of determining variates in the sub-sample not already

*Department of Statistics Quaid-i-Azam University, Islamabad, Pakistan,
Email: saadia2masood@yahoo.com
†Corresponding Author: Saadia Masood.
‡Department of Statistics Quaid-i-Azam University, Islamabad, Pakistan,
Email: javidshabbir@gmail.com

available in the sample. An important example is the regression estimator for means or totals, which uses values of an auxiliary variable from the full sample to estimate the mean of a variable of interest that is available only on the subsample. Multi-phase sampling is not widely discussed in literature. Mukerjee et al. [9] considered mainly three phases. Singh [18] proposed a class of estimators for population variance under two-phase sampling, whose composition was partially defined for the single auxiliary variable. Dorfman [5] proposed regression estimator for estimation of population variance under two-phase sampling scheme. Allen et al. [1] proposed a family of estimators of population mean using multi-auxiliary information in presence of measurement errors.

In most of the statistical studies, it is one of the common believes that the data are error-free but usually in realistic circumstances this statement is not absolutely met and the data are infected by errors. The consequences made for the error free data become invalid for the measurement error situation. Some important sources of measurement error are discussed in Cochran [3]. In sampling theory, the use of suitable auxiliary information results in considerable reduction in mean square error. Shukla et al. ([11],[12]) contributed by suggesting a mean estimator as well as class(es) of factor-type estimator(s) in the presence of measurement error. Singh and Karpe [13] have paid attention towards the estimation of population mean of the study variable y using the auxiliary information in presence of measurement error. Singh and Karpe [14] considered the problem of estimation of population variance S_y^2 under the assumptions: (i) when the study variable y is measured without error and auxiliary variable x is affected by error with known error variance S_v^2, (ii) when the study variable y is affected by error with known error variance S_u^2 and the auxiliary variable x is free from error. Furthermore, under the assumption of measurement error in study variable$y$, Singh and Karpe [15] paid attention towards the estimation of finite population variance S_y^2. Bhushan et al. [2] proposed two-phase generalized class of regression-type estimators using auxiliary information. Diana and Giordan [4] have proposed a family of estimators for the population variance S_y^2 by assuming error in both variables $y$and x under the regression approach. In practical application, let a psychiatrist wants to estimate the population variance of level of pathology in certain class of patients which depends upon the thinking disturbance , aggressive attitude, number of major miss-haps in life, ect.

In literature, the work on estimation of finite population variance using multi-auxiliary variables under multi-phase sampling is lacking especially when the study variable y is assumed to be contaminated with measurement error, so the present article is one of the steps to the solution of such situation.

Proposed Set-up: In the present study, we consider the following set-up:

(1) Complete Information Case (CIC): When information on all q auxiliary variables is *known* (we use single phase sampling).
(2) Incomplete Information Case (IIC): When information on some auxiliary variables is *known*.
(3) No Information Case (NIC): When information on all q auxiliary variables is *unknown*.

In Section 2, the symbols and notations used in this article are discussed. Section 3 presents the generalized regression-type variance estimator based on complete information of the multi-auxiliary variables about population variance, when the study variable is contaminated with measurement error. Section 4 present the generalized regression-type variance estimator when population variance of few auxiliary variables is known. In

Section 5, the generalized regression-type variance estimator is proposed when the population variance of all multi-auxiliary variable is unknown. Sections 6, 7 and 8 present the efficiency comparison, numerical analysis and concluding remarks respectively.

2. Symbols and notations

Let $U = \{1, 2, ..., j, ..., N\}$ be a finite population of N distinct and identifiable units. Let y and x_i $(i = 1, 2, ..., r, r+1, ..., q)$, be the study and the q auxiliary variables respectively, taking values y_j and x_{ij} for the j-th population unit. We are interested in estimating the finite population variance $\left(S_y^2\right)$ under multi-phase sampling schemes. Specifically we assume that a preliminary large sample $n_{(1)}$ is drawn with simple random sampling without replacement (SRSWOR) from a population and information on the auxiliary variable x_1 is taken. In second phase, a relatively small sample of size $n_{(2)}$ is drawn from $n_{(1)}$ $\left(n_{(2)} < n_{(1)}\right)$ and information on both auxiliary variables x_1 and x_2 is taken. This procedure goes up to the last phase when the smallest sample of size $n_{(q+1)}$ $\left(n_{(q+1)} < n_{(q)} < .... < n_{(1)}\right)$ is drawn. At this phase, all the q auxiliary variables as well as the variable of interest y are also observed. According to assumption, the measurement error is present in the variable of interest y denoted by $y^{\otimes}$ with known variance S_u^2. Moreover, let $S_{x_i}^2$ and $s_{x_{i(l)}}^2$denote the known population variance and sample variance of the i-th auxiliary variable $(i = 1, 2, ..., r, r+1, ..., q)$ at l-th phase $(l = i, ..., q, q+1)$, respectively. We limit our numerical study to two-phase sampling using *three* auxiliary variables.

The observational or measurement errors are defined as
$u_{j(l)} = y_{j(l)}^{\otimes} - y_{j(l)}$ and $v_{ij(l)} = x_{ij(l)}^{\otimes} - x_{ij(l)}$ $(i = 1, 2, ..., r, r+1, ..., q)$,
where $u_{j(l)}$ and $v_{ij(l)}$are assumed to be stochastic with zero mean and constant variances S_u^2 and $S_{v_i}^2$. As $\bar{y}_{(l)}$ and $\bar{x}_{i(l)}$ are unbiased estimators but $s_{y_{(l)}}^2$and $s_{x_{i(l)}}^2$ are biased estimators.
Let $\left(\bar{Y}, \bar{X}_i\right)$ and $\left(S_y^2, S_{x_i}^2\right)$ be the population means and population variances of the true values of $y_{j(l)}$ and $x_{ij(l)}$ respectively with corresponding sample means $\left(\bar{y}_{(l)}, \bar{x}_{i(l)}\right)$ and sample variances $\left(s_{y_{(l)}}^2, s_{x_{i(l)}}^2\right)$ at l-th phase. We know that $\bar{y}_{(l)}^{\otimes} = \frac{1}{n_{(l)}} \sum_{j=1}^{n_{(l)}} y_{i(l)}^{\otimes}$is unbiased estimator but $s_{y_{(l)}}^{\otimes 2} = \frac{1}{n_{(l)}-1} \sum_{j=1}^{n_{(l)}} \left(y_{j(l)}^{\otimes} - \bar{y}_{(l)}^{\otimes}\right)^2$ is biased estimator of S_y^2 due to measurement error. Similarly $\bar{x}_{ij(l)}^{\otimes} = \frac{1}{n_{(l)}} \sum_{j=1}^{n_{(l)}} x_{ij(l)}^{\otimes}$ is unbiased estimator but $s_{x_i}^{\otimes 2} = \frac{1}{n_{(l)}-1} \sum_{j=1}^{n_{(l)}} \left(x_{ij(l)}^{\otimes} - \bar{x}_{i(l)}^{\otimes}\right)^2$ $(i = 1, 2, ..., r, r+1, ..., q)$ is biased estimator of $S_{x_i}^2$ due to measurement error at l-th phase.
The expected values of $s_{y_{(l)}}^{\otimes 2}$ and $s_{x_i}^{\otimes 2}$are given by
$E\left(s_{y_{(l)}}^{\otimes 2}\right) = S_y^2 + S_u^2$ and $E\left(s_{x_{i(l)}}^{\otimes 2}\right) = S_{x_i}^2 + S_{v_i}^2$,
where S_u^2 and $S_{v_j}^2$are known, then the unbiased estimators of S_y^2 and $S_{x_j}^2$are
$\hat{S}_{y_{(l)}}^2 = s_{y_{(l)}}^{\otimes 2} - S_u^2$ and $\hat{S}_{x_{i(l)}}^2 = s_{x_{i(l)}}^{\otimes 2} - S_{v_i}^2 (i = 1, 2, ..., r, r+1, ..., q)$.

To obtain the properties of proposed estimators, we use the following approximations. For l-th and $(l+1)$-th phase, we define the notations as
Let $s_{y_{(l)}}^{\otimes 2} = S_y^2\left(1 + e_{y_{(l)}}^{\otimes}\right)$, $\quad s_{y_{(l+1)}}^{\otimes 2} = S_y^2\left(1 + e_{y_{(l+1)}}^{\otimes}\right)$, $\quad s_{y_{(l)}}^2 = S_y^2\left(1 + e_{y_{(l)}}\right)$,
$s_{x_{i(l)}}^{\otimes 2} = S_{x_i}^2\left(1 + e_{x_{i(l)}}^{\otimes}\right)$, $\quad s_{x_{i(l)}}^2 = S_{x_i}^2\left(1 + e_{x_{i(l)}}\right)$, $\quad s_{x_{i(l+1)}}^2 = S_{x_i}^2\left(1 + e_{x_{i(l+1)}}\right)$,

such that $E\left(e_{y(l)}^{\otimes 2}\right)=\varphi_{(l)}S_y^4A_{yy}^*,\ E\left(e_{y(l+1)}^{\otimes 2}\right)=\varphi_{(l+1)}S_y^4A_{yy}^*,\ E\left(e_{y(l)}^{2}\right)=\varphi_{(l)}S_y^4\lambda_{yy}^*,$
$E\left(e_{x_{i(l)}}^{2}\right)=\varphi_{(l)}S_{x_i}^4\lambda_{x_ix_i}^*,\ E\left(e_{x_{i(l+1)}}^{2}\right)=E\left(e_{x_{i(l)}}e_{x_{i(l+1)}}\right)=\varphi_{(l+1)}S_{x_i}^4\lambda_{x_ix_i}^*,$
$E\left(e_{y(l)}^{\otimes}e_{x_{i(l)}}\right)=\varphi_{(l)}S_y^2S_{x_i}^2\lambda_{yx_i}^*,\ E\left(e_{y(l)}^{\otimes}e_{x_{i(l+1)}}\right)=\varphi_{(l+1)}S_y^2S_{x_i}^2\lambda_{yx_i}^*,$

where $A_{yy}^*=\gamma_{2y}+\frac{2+\gamma_{2u}\left(1-\theta_y\right)^2}{\theta_y^2}$, $\theta_y=\frac{S_y^2}{S_y^2+S_u^2}$, $\gamma_{2y}=\beta_{2(y)}-3$ and $\gamma_{2u}=\beta_{2(u)}-3$, here $\beta_{2(y)}$ and $\beta_{2(u)}$ are the population co-efficients of kurtosis for the variable y and u.

Let $\lambda_{x_ix_i}^*=\lambda_{x_ix_i}-1$, $\lambda_{yx_i}^*=\lambda_{yx_i}-1$, $\mu_{yx_i}^*=\mu_{yx_i}-\mu_y\mu_{x_i}$, $\mu_y=S_y^2$, $\mu_{x_i}=S_{x_i}^2$ and $\varphi_{(l)}=\frac{1}{n_{(l)}}$.

Also $\lambda_{ts}=\frac{\mu_{22(t,s)}}{\mu_{20(t,s)}\mu_{02(t,s)}}=\frac{\mu_{ts}}{\mu_t\mu_s}$ or $t=y,x_i$ and $s=y,x_i$ $(i=1,2,....,r,r+1,....,q)$,

where $\mu_{ab(t,s)}=\frac{\sum_{i=1}^{N}\left(t_i-\bar{T}\right)^a\left(s_i-\bar{S}\right)^b}{N-1}$.

For $a=2$ and $b=2\Rightarrow\mu_{22(t,s)}=\frac{\sum_{i=1}^{N}\left(t_i-\bar{T}\right)^2\left(s_i-\bar{S}\right)^2}{N-1}$.

For $a=0$ and $b=2\Rightarrow\mu_{02(t,s)}=\frac{\sum_{i=1}^{N}\left(s_i-\bar{S}\right)^2}{N-1}$.

For $a=2$ and $b=0\Rightarrow\mu_{20(t,s)}=\frac{\sum_{i=1}^{N}\left(t_i-\bar{T}\right)^2}{N-1}$.

3. Generalized Regression-Type Estimators Using Multi-Auxiliary Variables

In this section, the estimators are formulated under the proposed setup.

3.1. Generalized regression-type estimators using multi-auxiliary variables under multi-phase sampling in the presence of ME under CIC. Let $s_{y(l)}^{\otimes 2}$ and $s_{x_{i(l)}}^{2}$ be the sample variances of the study variable y under measurement error and the i-th auxiliary variable $(i=1,2,...,r,r+1,...,q)$ respectively. The population variance $S_{x_i}^2$ $(i=1,2,...,r,r+1,...,q)$ of all multi-auxiliary variables is known. We consider the following generalized multi-phase regression-type estimator for population variance S_y^2 using α_i $(i=1,2,...,r,r+1,...,q)$ as unknown constants.

$$\hat{S}_{y1}^{\otimes 2}=s_{y(l)}^{\otimes 2}+\sum_{i=1}^{q}\alpha_i\left(S_{x_i}^2-s_{x_{i(l)}}^2\right). \tag{3.1}$$

In terms of e's, we have

$$\hat{S}_{y1}^{\otimes 2}-S_y^2=S_y^2e_{y(l)}^{\otimes}-\sum_{i=1}^{q}\alpha_iS_{x_i}^2e_{x_{i(l)}}. \tag{3.2}$$

Squaring (3.2) and then taking expectation, we get $MSE\left(\hat{S}_{y1}^{\otimes 2}\right)$ as

$$MSE\left(\hat{S}_{y1}^{\otimes 2}\right)=E\left(S_y^2e_{y(l)}^{\otimes}-\sum_{i=1}^{q}\alpha_iS_{x_i}^2e_{x_{i(l)}}\right)^2. \tag{3.3}$$

For optimum value of $\alpha_i=(-1)^{i+1}\frac{\left|\Lambda_{yx_i}\right|_{(y\bar{x}_q)}}{\left|\Lambda_{xx(q\times q)}\right|}$ $(i=1,2,...,q)$, the resulting minimum $MSE\left(\hat{S}_{y1}^{\otimes 2}\right)$, to first order of approximation, is given by

$$(3.4)\qquad MSE\left(\hat{S}_{y1}^{\otimes 2}\right)_{\min} = \varphi_{(l)} S_y^4 \left[A_{yy}^* - \sum_{i=1}^{q} (-1)^{i+1} \frac{|\Lambda_{yx_i}|_{(y\tilde{x}_q)}}{|\Lambda_{xx(q\times q)}|} \frac{\mu_{x_i}\lambda_{yx_i}^*}{\mu_y} \right].$$

Let $\Re^2_{s_y^2 . s^2_{\tilde{x}_q}} = \sum_{i=1}^{q} (-1)^{i+1} \frac{|\Lambda_{yx_i}|_{(y\tilde{x}_q)}}{|\Lambda_{xx(q\times q)}|} \frac{\mu^*_{yx_i}}{\mu_y^2}$, then (3.4) can be written as

$$(3.5)\qquad MSE\left(\hat{S}_{y1}^{\otimes 2}\right)_{\min} = \varphi_{(l)} S_y^4 \left[A_{yy}^* - \Re^2_{s_y^2 . s^2_{\tilde{x}_q}} \right].$$

Remark 3.1.1: Single-phase sampling using q auxiliary variables

For full information case using q multi-auxiliary variables, we replace l by 1, which is the case of simple random sampling. The estimator given in (3.1) becomes

$$(3.6)\qquad \hat{S}_{y1}^{\otimes 2\dagger} = s_{y_{(1)}}^{\otimes 2} + \sum_{i=1}^{q} \alpha_i \left(S_{x_i}^2 - s^2_{x_{i(1)}} \right).$$

For the optimum values of $\alpha_i = (-1)^{i+1} \frac{|\Lambda_{yx_i}|_{(y\tilde{x}_q)}}{|\Lambda_{xx(q\times q)}|}$ $(i = 1, 2, ..., q)$, the resulting minimum $MSE\left(\hat{S}_{y1}^{\otimes 2\dagger}\right)$, to first order of approximation, is given by

$$(3.7)\qquad MSE\left(\hat{S}_{y1}^{\otimes 2\dagger}\right)_{\min} = \varphi_{(1)} S_y^4 \left[A_{yy}^* - \Re^2_{s_y^2 . s^2_{\tilde{x}_q}} \right].$$

Here $|\Lambda_{yx_i}|_{(y\tilde{x}_q)}$ is the determinant of matrix of population variances of the variables $y, x_1, ..., x_q$ and $|\Lambda_{xx(q\times q)}|$ is the determinant of matrix of population variances of the variables $x_1, ..., x_q$.

Remark 3.1.2: Single-phase sampling using q auxiliary variables in the absence of measurement error

Let the observations of variable of interest y be recorded without an error. Substituting $S_u^2 = 0$ in (3.5), we get $A_{yy}^* = \lambda_{yy}^*$,

$$(3.8)\qquad MSE\left(\hat{S}_{y1}^{2}\right)_{\min} = \varphi_{(1)} S_y^4 \left[\lambda_{yy}^* - \Re^2_{s_y^2 . s^2_{\tilde{x}_q}} \right].$$

3.2. Generalized regression-type estimators using multi-auxiliary information under multi-phase sampling in the presence of ME under IIC. Let $s^2_{x_{i(l)}}$ and $s^2_{x_{i(l+1)}}$ be sample variances of the auxiliary variables x_i $(i = 1, 2, \ldots, r, r+1, \ldots, q)$ at l-th and $(l+1)$-th phases respectively with the sample size $n_{(l)}$ and $n_{(l+1)}$ having the population variance $S^2_{x_i}$. Also $s^{\otimes 2}_{y_{(l+1)}}$ be the sample variances of the study variable y of size $n_{(l+1)}$ selected at $(l+1)$-th phase. The population variance $S^2_{x_i}$ $(i = 1, 2, ..., r, r+1, ..., q)$ on some auxiliary variables is known. We formulate the generalized regression-type estimator for the estimation of unknown finite population variance S_y^2 using α_i, δ_i $(i = 1, 2, ..., r)$ and γ_i $(i = r+1, r+2, ..., q)$ as unknown constants.

$$\begin{aligned}\hat{S}_{y2}^{\otimes 2} =& s^{\otimes 2}_{y_{(l+1)}} + \sum_{i=1}^{r} \alpha_i \left(S^2_{x_i} - s^2_{x_{i(l)}} \right) + \sum_{i=1}^{r} \delta_i \left(S^2_{x_i} - s^2_{x_{i(l+1)}} \right) \\ (3.9)\qquad & + \sum_{i=r+1}^{q} \gamma_i \left(s^2_{x_{i(l)}} - s^2_{x_{i(l+1)}} \right).\end{aligned}$$

In terms of e's, we have

(3.10)
$$\hat{S}_{y2}^{\otimes 2}-S_y^2=\left[S_y^2 e_{y_{(l+1)}}^{\otimes}-\sum_{i=1}^{r}\alpha_i S_{x_i}^2 e_{x_{i(l)}}-\sum_{i=1}^{r}\delta_i S_{x_i}^2 e_{x_{i(l+1)}}+\sum_{i=r+1}^{q}\gamma_i S_{x_i}^2\left(e_{x_{i(l)}}-e_{x_{i(l+1)}}\right)\right].$$

Squaring (3.10) and then taking expectation, we get

$$\begin{aligned}MSE\left(\hat{S}_{y2}^{\otimes 2}\right)=&E\left[S_y^2 e_{y_{(l+1)}}^{\otimes}-\sum_{i=1}^{r}\alpha_i S_{x_i}^2 e_{x_{i(l)}}-\sum_{i=1}^{r}\delta_i S_{x_i}^2 e_{x_{i(l+1)}}\right.\\ &\left.+\sum_{i=r+1}^{q}\gamma_i S_{x_i}^2\left(e_{x_{i(l)}}-e_{x_{i(l+1)}}\right)\right]^2.\end{aligned} \tag{3.11}$$

For the optimum values of $\alpha_i=(-1)^{i+1}\left(\frac{|\Lambda_{yx_i}|_{(y\tilde{x}_q)}}{|\Lambda_{xx(q\times q)}|}-\frac{|\Lambda_{yx_i}|_{(y\tilde{x}_r)}}{|\Lambda_{xx(r\times r)}|}\right)$, $\delta_i=(-1)^{i+1}\frac{|\Lambda_{yx_i}|_{(y\tilde{x}_r)}}{|\Lambda_{xx(r\times r)}|}$ $(i=1,2,...,r)$ and $\gamma_i=(-1)^{i+1}\frac{|\Lambda_{yx_i}|_{(y\tilde{x}_q)}}{|\Lambda_{xx(q\times q)}|}$ $(i=r+1,r+2,...,q)$, the resulting minimum $MSE\left(\hat{S}_{y2}^{\otimes 2}\right)$, to first order of approximation, is given by

$$\begin{aligned}MSE\left(\hat{S}_{y2}^{\otimes 2}\right)_{\min}=&S_y^4\left[\varphi_{(l+1)}A_{yy}^{*}-\varphi_{(l)}\sum_{i=1}^{r}(-1)^{i+1}\frac{|\Lambda_{yx_i}|_{(y\tilde{x}_r)}}{|\Lambda_{xx(q\times q)}|}\frac{\mu_{yx_i}^{*}}{\mu_y^2}\right.\\ &\left.+\left(\varphi_{(l)}-\varphi_{(l+1)}\right)\sum_{i=1}^{q}(-1)^{i+1}\frac{|\Lambda_{yx_i}|_{(y\tilde{x}_q)}}{|\Lambda_{xx(q\times q)}|}\frac{\mu_{yx_i}^{*}}{\mu_y^2}\right].\end{aligned} \tag{3.12}$$

Let $\Re^2_{s_y^2.s_{\tilde{x}_r}^2}=\sum_{i=1}^{r}(-1)^{i+1}\frac{|\Lambda_{yx_i}|_{(y\tilde{x}_q)}}{|\Lambda_{xx(q\times q)}|}\frac{\mu_{yx_i}^{*}}{\mu_y^2}$. then (3.12) can be written as

$$MSE\left(\hat{S}_{y2}^{\otimes 2}\right)_{\min}=S_y^4\left[\varphi_{(l+1)}A_{yy}^{*}-\varphi_{(l)}\Re^2_{s_y^2.s_{\tilde{x}_r}^2}+\left(\varphi_{(l)}-\varphi_{(l+1)}\right)\Re^2_{s_y^2.s_{\tilde{x}_q}^2}\right]. \tag{3.13}$$

Remark 3.2.1: Two-phase sampling using q auxiliary variables

For the case of two-phase sampling using q multi-auxiliary variables, we replace l by 1. The estimator given in (3.9) becomes

(3.14)
$$\hat{S}_{y2}^{\otimes 2\dagger}=s_{y_{(2)}}^{\otimes 2}+\sum_{i=1}^{r}\alpha_i\left(S_{x_i}^2-s_{x_{i(1)}}^2\right)+\sum_{i=1}^{r}\delta_i\left(S_{x_i}^2-s_{x_{i(2)}}^2\right)+\sum_{i=r+1}^{q}\gamma_i\left(s_{x_{i(1)}}^2-s_{x_{i(2)}}^2\right).$$

For optimum values of $\alpha_i=(-1)^{i+1}\left(\frac{|\Lambda_{yx_i}|_{(y\tilde{x}_q)}}{|\Lambda_{xx(q\times q)}|}-\frac{|\Lambda_{yx_i}|_{(y\tilde{x}_r)}}{|\Lambda_{xx(r\times r)}|}\right)$, $\delta_i=(-1)^{i+1}\frac{|\Lambda_{yx_i}|_{(y\tilde{x}_q)}}{|\Lambda_{xx(q\times q)}|}$ $(i=1,2,...,r)$ and $\gamma_i=(-1)^{i+1}\frac{|\Lambda_{yx_i}|_{(y\tilde{x}_q)}}{|\Lambda_{xx(q\times q)}|}$ $(i=r+1,r+2,...,q)$, the minimum $MSE\left(\hat{S}_{y2}^{\otimes 2\dagger}\right)$, to first order of approximation, is given by

(3.15)
$$MSE\left(\hat{S}_{y2}^{\otimes 2\dagger}\right)_{\min}=S_y^4\left[\varphi_{(2)}\left(A_{yy}^{*}-\Re^{\otimes 2}_{s_y^2.s_{\tilde{x}_r}^2}\right)+\left(\varphi_{(1)}-\varphi_{(2)}\right)\left(\Re^{\otimes 2}_{s_y^2.s_{\tilde{x}_q}^2}-\Re^{\otimes 2}_{s_y^2.s_{\tilde{x}_r}^2}\right)\right].$$

Remark 3.2.2: Two-phase sampling using q auxiliary variables in the absence of measurement error

Let the observations of variable of interest y be recorded without an error. Substituting $S_u^2 = 0$ in (3.13), we get $A_{yy}^* = \lambda_{yy}^*$, so

$$MSE\left(\hat{S}_{y2}^2\right)_{\min} = S_y^4\left[\varphi_{(2)}\lambda_{yy}^* - \varphi_{(1)}\Re^2_{s_y^2.s_{\tilde{x}_r}^2} + \left(\varphi_{(1)} - \varphi_{(2)}\right)\Re^2_{s_y^2.s_{\tilde{x}_q}^2}\right]. \tag{3.16}$$

3.3. Generalized regression-type estimators using multi-auxiliary information under multi-phase sampling in the presence of ME under NIC. Let $s_{y(l+1)}^{\otimes 2}$ and $s_{x_{i(l+1)}}^2$ be the sample variances of the study variable y under measurement error and the i-th auxiliary variable $(i = 1, 2, ..., r, r+1, ..., q)$ respectively at $(l+1)$-th phase, whereas $s_{x_{i(l)}}^2$ be the sample variance of i-th auxiliary variable at l-th phase. The population variance $S_{x_i}^2$ $(i = 1, 2, ..., r, r+1, ..., q)$ of all multi-auxiliary variables is unknown. We consider the following generalized regression-type estimator for population variance S_y^2 under no information case using α_i $(i = 1, 2, ..., r, r+1, ..., q)$ as unknown constant.

$$\hat{S}_{y3}^{\otimes 2} = s_{y(l+1)}^{\otimes 2} + \sum_{i=1}^{q}\alpha_i\left(s_{x_{i(l)}}^2 - s_{x_{i(l+1)}}^2\right). \tag{3.17}$$

To the first order of approximation, we write (3.17) as

$$\hat{S}_{y3}^{\otimes 2} - S_y^2 = S_y^2 e_{y(l+1)}^{\otimes} + \sum_{i=1}^{q} S_{x_i}^2\alpha_i\left(e_{x_{i(l)}} - e_{x_{i(l+1)}}\right). \tag{3.18}$$

Squaring (3.18) and then taking expectation, we get MSE as

$$MSE\left(\hat{S}_{y3}^{\otimes 2}\right) = S_y^4 E\left[e_{y(l+1)}^{\otimes} + \sum_{i=1}^{q}\alpha_i\left(e_{x_{i(l)}} - e_{x_{i(l+1)}}\right)\right]^2. \tag{3.19}$$

For optimum value of $\alpha_i = (-1)^{i+1}\frac{|\Lambda_{yx_i}|_{(y\tilde{x}_q)}}{|\Lambda_{xx(q\times q)}|}(i = 1, 2, ..., q)$,the resulting minimum $MSE\left(\hat{S}_{y3}^{\otimes 2}\right)$, to first order of approximation, is given by

$$\begin{aligned} MSE\left(\hat{S}_{y3}^{\otimes 2}\right)_{\min} =& S_y^4\Bigg[\varphi_{(l+1)}\left(A_{yy}^* - \sum_{i=1}^{q}(-1)^{i+1}\frac{|\Lambda_{yx_i}|_{(y\tilde{x}_q)}}{|\Lambda_{xx(q\times q)}|}\frac{\mu_{yx_i}^*}{\mu_y^2}\right) \\ &+\varphi_{(l)}\sum_{i=1}^{q}(-1)^{i+1}\frac{|\Lambda_{yx_i}|_{(y\tilde{x}_q)}}{|\Lambda_{xx(q\times q)}|}\frac{\mu_{yx_i}^*}{\mu_y^2}\Bigg]. \end{aligned} \tag{3.20}$$

We can write (3.20) in compact form as

$$MSE\left(\hat{S}_{y3}^{\otimes 2}\right)_{\min} = S_y^4\left[\varphi_{(l+1)}A_{yy}^* + \left(\varphi_{(l)} - \varphi_{(l+1)}\right)\Re^2_{s_y^2.s_{\tilde{x}_q}^2}\right]. \tag{3.21}$$

Remark 3.3.1: Two-phase sampling using q auxiliary variables

For no information case using q auxiliary variables, we replace l by 1, which is the case of two-phase sampling. The estimator given in (3.17) becomes

$$\hat{S}_{y3}^{\otimes 2\dagger} = s_{y(2)}^{\otimes 2} + \sum_{i=1}^{q}\alpha_i\left(s_{x_{i(1)}}^2 - s_{x_{i(2)}}^2\right). \tag{3.22}$$

For optimum value of $\alpha_i = (-1)^{i+1} \frac{|\Lambda_{yx_i}|_{(y\bar{x}_q)}}{|\Lambda_{xx(q\times q)}|}$ $(i = 1, 2, ..., q)$, the resulting minimum $MSE\left(\hat{S}_{y3}^{\otimes 2\dagger}\right)$, to first order of approximation, is given by

$$(3.23) \quad MSE\left(\hat{S}_{y3}^{\otimes 2\dagger}\right)_{\min} = S_y^4 \left[\varphi_{(2)} A_{yy}^* + \left(\varphi_{(1)} - \varphi_{(2)}\right) \Re_{s_y^2 \cdot s_{\bar{x}_q}^2}^{\otimes 2}\right].$$

Remark 3.3.2: Two-phase sampling using q auxiliary variables in the absence of measurement error

Let the observations of variable of interest y be recorded without an error. Substituting $S_u^2 = 0$ in (3.21), we get $A_{yy}^* = \lambda_{yy}^*$,

$$(3.24) \quad MSE\left(\hat{S}_{y3}^{2}\right)_{\min} = S_y^4 \left[\varphi_{(2)} \lambda_{yy}^* + \left(\varphi_{(1)} - \varphi_{(2)}\right) \Re_{s_y^2 \cdot s_{\bar{x}_q}^2}^{2}\right].$$

4. Efficiency Comparison

To obtain the efficiency of proposed estimators, we compare the mean square errors of proposed multi-phase regression-type variance estimators under measurement error with the estimators assumed to be free of error.
By (3.7) and (3.8), (3.15) and (3.16), (3.23) and (3.24), it is evident that

$$(4.1) \quad \lambda_{yy}^* < A_{yy}^*.$$

Note: The Condition (4.1) is always true.

Remark: The numerical comparison is made under the efficiency conditions given above.

5. Data Description

Population 1: (Source: Mukherjee et al. [8])
The fertility data is based on 64 countries. Let y =Total fertility rate, 1980–1985, the average number of children born to a woman, using age specific fertility rates for a given year, x_1 = Child mortality, the number of deaths of children under age 5 in a year per 1000 live births, x_2 =Female literacy rate, (percent) and x_3 =Per capita GNP (in billions) in 1980.

$$N = 64, S_y^2 = 2.277, S_{x_1}^2 = 5772.670, S_{x_2}^2 = 676.409, S_{x_3}^2 = 7429417.00,$$
$$\bar{Y} = 5.549, \bar{X}_1 = 141.500, \bar{X}_2 = 51.188, \bar{X}_3 = 1401.250, S_u^2 = 1.255,$$
$$\lambda_{yy} = 2.773, \lambda_{x_1x_1} = 2.341, \lambda_{x_2x_2} = 1.631, \lambda_{x_3x_3} = 34.046,$$
$$\lambda_{yx_1} = 1.458, \lambda_{yx_2} = 1.069, \lambda_{yx_3} = 0.540, \lambda_{x_1x_2} = 1.415,$$
$$\lambda_{x_1x_3} = 1.921, \lambda_{x_2x_3} = 0.372, A_{yy} = 1.234.$$

Population 2:(Source: Gujarati [6])
The data is baesd on the demand for chicken in USA, 1960-1982. Let y = Per capita consumption of chickens in pounds, x_1 = Real disposable income per capita in dollars, x_2 = Real retail price of chicken per pound (in cents) and x_3 = Real retail price of pork per pound (in cents).

$$N = 23, S_y^2 = 54.360, S_{x_1}^2 = 381735.00, S_{x_2}^2 = 123.592, S_{x_3}^2 = 1240.710,$$
$$\bar{Y} = 39.669, \bar{X}_1 = 1035.065, \bar{X}_2 = 47.995, \bar{X}_3 = 90.400, S_u^2 = 3.987,$$
$$\lambda_{yy} = 2.03, \lambda_{x_1x_1} = 2.696, \lambda_{x_2x_2} = 1.756, \lambda_{x_3x_3} = 1.951, \lambda_{yx_1} = 2.094,$$
$$\lambda_{yx_2} = 1.541, \lambda_{yx_3} = 1.758, \lambda_{x_1x_2} = 1.997, \lambda_{x_1x_3} = 2.145, \lambda_{x_2x_3} = 1.755,$$

$A_{yy} = 1.033$.

Population 3:(Source: Vandaele [20])
The data is based on the crime rate data of USA in 1960. Let y =Number of offenses reported to police per million population, x_1 =Number of males of age 14-24 per 1000 population, x_2 =Indicator variable for southern states and x_3 =Mean number of years of schooling times 10 for persons age 25 or older.

$N = 47, S_y^2 = 1495.853, S_{x_1}^2 = 151.516, S_{x_2}^2 = 0.229, S_{x_3}^2 = 124.076,$
$\bar{Y} = 90.508, \bar{X}_1 = 137.511, \bar{X}_2 = 0.340, \bar{X}_3 = 105.406, S_u^2 = 1428.881,$
$\lambda_{yy} = 3.859, \lambda_{x_1x_1} = 3.684, \lambda_{x_2x_2} = 1.423, \lambda_{x_3x_3} = 1.896, \lambda_{yx_1} = 0.456,$
$\lambda_{yx_2} = 0.743, \lambda_{yx_3} = 1.041, \lambda_{x_1x_2} = 1.354, \lambda_{x_1x_3} = 1.356, \lambda_{x_2x_3} = 1.220,$
$A_{yy} = 2.703$.

Table 1. MSE of proposed ratio-type estimators $\hat{S}_{y1}^{\otimes 2\dagger}$, $\hat{S}_{y2}^{\otimes 2\dagger}$ and $\hat{S}_{y3}^{\otimes 2\dagger}$

Estimators	Pop.1	Pop.2	Pop.3
$\hat{S}_{y1}^{\otimes 2\dagger}$	0.408	119.840	17.957
	0.112	**2.05**	**42.106**
$\hat{S}_{y2}^{\otimes 2\dagger}$	0.709	453.212	62.581
	0.134	**14.781**	**1.408**
$\hat{S}_{y3}^{\otimes 2\dagger}$	0.554	280.270	19.282
	0.379	**2.050**	**42.106**

*The results written in Table 1 in bold format are the absolute values of measurement error.

6. Conclusion

In general, the presence of measurement error in the survey data invalidates the results. The goal of this study was to show how measurement error is to be seperated in case of multi-phase sampling using multi-auxiliary variables for estimation of population variance S_y^2. The values of absolute measurement error are shown in Table 1. It is also evident that the condition (4.1) holds for all the populations. Hence, the use of proposed estimators are highly preferred in the cases of multi-phase sampling under CIC, IIC and NIC.

Acknowledgments
The authors wish to thank the editor and the anonymous referees for their suggestions which led to improvement in the earlier version of the manuscript.

References

[1] Allen, J., Singh, H.P. and Smarandache, F. *A family of estimators of population mean using multi-auxiliary information in presence of measurement error,* International Journal of Social Economics, 30 (7): 837-849, 2003.

[2] Bhushan, S., Singh, R. K. and Pandey, A. *Some generalized classes of double sampling regression type estimators using auxiliary information,* Science Vision, 11 (1): 2-6, 2011.

[3] Cochran W.G. Errors of measurement in statistics. *Technometrics*, 10: 637-666, 1968.

[4] Diana, T. and Giordan, M. *Finite population variance estimation in presence of measurement error,* Communications in Statistics-Theory and Methods, 41: 4302-431, 2012.

[5] Dorfman, A. H. *A note on variance estimation for regression estimator for double sampling*, Journal of American Statistical Association, 89: 137- 140, 1994.

[6] Gujarati, D. M. *Basic Econometrics. The McGraw-Hill Companies*, 2004.

[7] Isaki, C.T. *Variance estimation using auxiliary information*,Journal of American Statistical Association, 381: 117–123, 1983.

[8] Mukherjee, C., White, H. and Whyte, M. *Econometrics and Data Analysis for Developing Countries, Routledge, London*, 1998.

[9] Mukerjee R., Rao T.J. and Vijayan K. *Regression-type estimators using multiple auxiliary information*, Australian Journal of Statistics, 29 (3): 244–254, 1998.

[10] Raj, D. *On a method of using multi-auxiliary information in sample surveys*, Journal of American Statistical Association, 60: 270-277, 1965.

[11] Shukla, D., Pathak, S. and Thakur, N. S. *An estimator for mean estimation in presence of measurement error*, Research and Reviews: A Journal of Statistics, 2 (1), 1-8, 2012 a.

[12] Shukla, D., Pathak, S. and Thakur, N. S. *Class(es) of factor-type estimator(s) in presence of measurement error*, Journal of Applied Statistical Methods, 11 (2), 336- 347, 2012 b.

[13] Singh, H.P. and Karpe, N. *Ratio-product estimator for population mean in the presence of measurement errors*, Journal of Applied Statistical Science, 16: 49-64, 2008 a.

[14] Singh, H.P. and Karpe, N. *Estimation of population variance using auxiliary information in the presence of measurement errors*, Statistics in Transition, 9 (3): 443-470, 2008b.

[15] Singh, H.P. and Karpe, N. *A class of estimators using auxiliary information for estimating finite population variance in presence of measurement errors*, Communications in Statistics-Theory and Methods, 38: 734-741, 2009.

[16] Singh H. P. and Solanki, R.S. *Improved estimation of finite population variance using auxiliary information,* Communications in Statistics-Theory and Methods, 42 (15), 2718-2730, 2013 a.

[17] Singh H. P. and Solanki, R.S. *A new procedure for variance estimation in simple random sampling using auxiliary information,* Statistical Papers, 54 (2), 479-497, 2013 b.

[18] Singh, S. *Estimation of finite population variance using double sampling*, Aligarh Journal of Statistics, 11: 53-56, 1991.

[19] Solanki R. S. and Singh H. P. *An improved class of estimators for the population variance,* Model Assisted Statistics and Applications, 8 (3), 229-238, 2013.

[20] Vandaele, W. *Participation in Illegitimate Activities: Erlich Revisted. Deterrence and Incapacitation*, National Academy of Sciences. 270–335, 1978.