A Generalized Class of Estimators for Finite Population Variance in Presence of Measurement Errors

Prayas Sharma
Banaras Hindu University, Varanasi, India, prayassharma02@gmail.com

Rajesh Singh
Banaras Hindu University, Varanasi, India, rsinghstat@gmail.com

Follow this and additional works at: http://digitalcommons.wayne.edu/jmasm

Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

Recommended Citation
Available at: http://digitalcommons.wayne.edu/jmasm/vol12/iss2/13

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized administrator of DigitalCommons@WayneState.
A Generalized Class of Estimators for Finite Population Variance in Presence of Measurement Errors

Prayas Sharma  
Banaras Hindu University  
Varanasi, India  

Rajesh Singh  
Banaras Hindu University  
Varanasi, India  

The problem of estimating the population variance is presented using auxiliary information in the presence of measurement errors. The estimators in this article use auxiliary information to improve efficiency and assume that measurement error is present both in study and auxiliary variable. A numerical study is carried out to compare the performance of the proposed estimator with other estimators and the variance per unit estimator in the presence of measurement errors.

Keywords: Population mean, study variate, auxiliary variates, mean squared error, measurement errors, efficiency.

Introduction

Over the past several decades, statisticians are paying their attention towards the problem of estimation of parameters in the presence of measurement errors. In survey sampling, the properties of estimators based on data usually presuppose that the observations are the correct measurements on characteristics being studied. However, this assumption is not satisfied in many applications and data is contaminated with measurement errors, such as non-response errors, reporting errors, and computing errors. These measurement errors make the result invalid, which are meant for no measurement error case. If measurement errors are very small and we can neglect it, then the statistical inferences based on observed data continue to remain valid. On the contrary, when they are not appreciably small and negligible, the inferences may not be simply invalid and inaccurate but may often lead to unexpected, undesirable and unfortunate consequences (see Srivastava and Shalabh, 2001). Some important sources of measurement errors in...
survey data are discussed in Cochran (1968), Shalabh (1997), and Sud and Srivastva (2000). Singh and Karpe (2008, 2010), Kumar et al. (2011a, b) studied some estimators of population mean under measurement error.

Many authors, including Das and Tripathi (1978), Srivastava and Jhajj (1980), Singh and Karpe (2009) and Diana and Giordan (2012), studied the estimation of population Variance \( \sigma_y^2 \) of the study variable \( y \) using auxiliary information in the presence of measurement errors. The problem of estimating the population variance and its properties are studied here in the presence of measurement errors.

Consider a finite population \( U = (U_1, U_2, \ldots, U_N) \) of \( N \) units. Let \( Y \) and \( X \) be the study variate and auxiliary variate, respectively. Suppose a set of \( n \) paired observations are obtained through simple random sampling procedure on two characteristics \( X \) and \( Y \). Further assume that \( x_i \) and \( y_i \) for the \( i^{th} \) sampling units are observed with measurement error as opposed to their true values \((X_i, Y_i)\) For a simple random sampling scheme, let \((x_i, y_i)\) be observed values instead of the true values \((X_i, Y_i)\) for \( i^{th} \) \((i=1,2,\ldots,n)\) unit, as

\[
\begin{align*}
u_i &= y_i - Y_i \\
v_i &= x_i - X_i
\end{align*}
\]

where \( u_i \) and \( v_i \) are associated measurement errors which are stochastic in nature with mean zero and variances \( \sigma^2_u \) and \( \sigma^2_v \), respectively. Further, let the \( u_i \)'s and \( v_i \)'s are uncorrelated although \( X_i \)'s and \( Y_i \)'s are correlated.

Let the population means of \( X \) and \( Y \) characteristics be \( \mu_x \) and \( \mu_y \), population variances of \((x, y)\) be \((\sigma_x^2, \sigma_y^2)\) and let \( \rho \) be the population correlation coefficient between \( x \) and \( y \) respectively (see Manisha and Singh (2002)).

**Notations**

Let \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \), \( \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \), be the unbiased estimator of population means \( \bar{X} \) and \( \bar{Y} \), respectively but \( s^2_x = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \) and \( s^2_y = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2 \) are not unbiased estimator of \((\sigma_x^2, \sigma_y^2)\), respectively. The expected values of \( s^2_x \) and \( s^2_y \) in the presence of measurement error are, given by,
When the error variance $\sigma^2_v$ is known, the unbiased estimator of $\sigma^2_x$, is
\[ \hat{\sigma}^2_x = s^2_x - \sigma^2_v > 0, \] and when $\sigma^2_u$ is known, then the unbiased estimator of $\sigma^2_y$ is
\[ \hat{\sigma}^2_y = s^2_y - \sigma^2_u > 0. \]

Define
\[ \hat{\sigma}^2_y = \sigma^2_y (1 + e_o) \]
\[ \bar{x} = \mu_x (1 + e_t) \]
such that
\[ E(e_o) = E(e_t) = 0, \]
\[ E(e_i^2) = \frac{C^2}{n} \left( 1 + \frac{\sigma^2_u}{\sigma^2_x} \right) = \frac{C^2}{n\theta_x}, \]
and to the first degree of approximation (when finite population correction factor is ignored)
\[ E(e_o^2) = \frac{A_y}{n}, \quad E(e_y e_i) = \frac{\lambda C}{n}. \]

where,
\[ A_y = \left\{ \gamma_{2y} + \gamma_{2u} \frac{\sigma^4_u}{\sigma^4_y} + 2 \left( 1 + \frac{\sigma^2_u}{\sigma^2_y} \right) \right\}, \quad \lambda = \frac{\mu_2(x, y)}{\mu_2^2(x, y)}, \quad C_y = \frac{\sigma_y}{\mu_x}, \quad \theta_y = \frac{\sigma^2_y}{\sigma^2_x + \sigma^2_y}, \]
\[ \theta_u = \frac{\sigma^2_u}{\sigma^2_x + \sigma^2_u}, \quad \gamma_{2y} = \beta_2(y) - 3, \quad \gamma_{2u} = \beta_2(u) - 3, \quad \beta_2(u) = \frac{\mu_4(u)}{\mu_2^2(u)}, \quad \beta_2(y) = \frac{\mu_4(y)}{\mu_2^2(y)}, \]
\[ \mu_4(y) = E(Y_i - \mu_y)^4, \quad \mu_4(u) = E(u_i^4). \]
\(\theta_x\) and \(\theta_y\) are the reliability ratios of \(X\) and \(Y\), respectively, lying between 0 and 1.

**Estimator of population variance under measurement error**

According to Koyuncu and Kadilar (2010), a regression type estimator \(t_1\) is defined as

\[
t_1 = w_1 \sigma^2_y + w_2 (\mu_x - \bar{x})
\]

(3)

where \(w_1\) and \(w_2\) are constants that have no restriction.

Expression (3) can be written as

\[
t_1 - \sigma^2_y = (w_1 - 1) \sigma^2_y + w_1 \sigma^2_y e_0 - w_2 \mu_x e_1
\]

(4)

Taking expectation both sides of (4), results in

\[
Bias(t_1) = \sigma^2_y (w_1 - 1)
\]

(5)

Squaring both sides of (4)

\[
(t_1 - \sigma^2_y)^2 = [(w_1 - 1) \sigma^2_y + w_1 \sigma^2_y e_0 - w_2 \mu_x e_1]
\]

(6)

or

\[
(t_1 - \sigma^2_y)^2 = [(w_1 - 1)^2 \sigma^2_y + w_1^2 \sigma^2_y e_0^2 + w_2^2 \mu_x^2 e_1^2 + 2(w_1 - 1)w_1 \sigma^4_y e_0 e_1
\]

\[
-2(w_1 - 1)w_2 \sigma^2_y \mu_x e_1 - 2w_1 w_2 \sigma^2_y \mu_x e_1 e_0)]
\]

(7)

Simplifying equation (7), taking expectations and using notations, results in

the mean square error of \(t_1\) up to first order of approximation, as

\[
MSE(t_1) = \left[ \sigma^4_y w_1^2 \left( \frac{A}{n} + 1 \right) + (1 - 2w_1) \sigma^4_y + w_2^2 \mu_x^2 \frac{C_x^2}{n\theta_x} - \frac{2w_1 w_2 \mu_x \sigma^2_y \lambda C_x}{n} \right]
\]

(8)
In the case, when the measurement error is zero, MSE of $t_i$ without measurement error is given by,

$$MSE^* (t_i) = \frac{\sigma^4_y}{n} \left\{ \gamma_{2y} + 2 + n \right\} + (1 - 2w_i)\sigma^4_y + w^2_i \mu^2_x \frac{C^2_x}{n} - 2w_i \mu_x \sigma^2_y \lambda \frac{C_x}{n}$$

and

$$M^*_e = \frac{\sigma^2_y}{n} \left[ \frac{\sigma^4_{2y} \gamma_{2u}}{\sigma^4_y} + 2 \left( \frac{\sigma^4_u}{\sigma^4_y} \right)^2 + 4 \frac{\sigma^4_u}{\sigma^4_y} \right] + w^2_i \mu^2_x \frac{C^2_x}{n} \frac{\sigma^2_y}{\sigma^2_y}$$

is the contribution of measurement errors in the MSE of estimator $t_i$.

Differentiating (8) with respect to $w_i$ and $w_2$ partially, equating them to zero and after simplification, results in the optimum values of $w_i$ and $w_2$, respectively as

$$w_i^* = -\frac{\sigma^4_y B}{C^2 - AB}, \quad w_2^* = -\frac{\sigma^4_y C}{C^2 - AB}$$

where, $A = \frac{(A_y + 1)\sigma^4_y}{n}$, $B = \frac{\mu_x^2 C_x^2}{n \theta_x}$ and $C = \frac{\sigma^2_y \mu_x C_x \lambda}{n}$.

Using the values of $w_i^*$ and $w_2^*$ from equation (11) into equation (8), gives the minimum MSE of the estimator $t_2$ in terms of $A$, $B$ and $C$ as

$$MSE(t_1)_{\text{min}} = \left( \frac{\sigma^4_y}{(C^2 - AB)} \right)^2 \left[ \frac{(C^2 - AB)^2}{\sigma^4_y} + 3BC^2 - AB^2 - 2BC \right]$$

**Another estimator under measurement error**

Based on Solanki and Singh (2012), an estimator $t_3$ is defined as
A CLASS OF ESTIMATORS FOR FINITE POPULATION VARIANCE

\[ t_2 = \sigma_y^2 \left\{ 2 - \left( \frac{x}{\mu_x} \right)^\alpha \exp \left[ \beta \left( \frac{\mu_x - \mu_y}{\mu_x + \mu_y} \right) \right] \right\} \]  

(13)

where \( \alpha \) and \( \beta \) are suitably chosen constants.

Expressing the estimator \( t_2 \), in terms of \( e \)'s is

\[ t_2 = \hat{\sigma}_y^2 \left[ 2 - \left( 1 + e_1 \right)^\alpha \exp \left\{ \frac{\beta e_1}{2} \left( 1 + \frac{e_1}{2} \right)^{-1} \right\} \right] \]  

(14)

Expanding equation (14) and simplifying results in

\[ t_2 - \sigma_y^2 = \sigma_y^2 \left[ e_0 - \frac{k}{2} \left( e_1 + e_0 e_1 \right) - \frac{e_1^2}{8} \left( k^2 - 2k \right) \right] \]  

(15)

where \( k = (\beta + 2\alpha) \).

On taking expectations of both sides of (15), the bias of the estimator \( t_3 \) up to the first order of approximation is obtained as

\[ \text{Bias}(t_2) = \sigma_y^2 \left[ e_0 + \frac{k}{2} \frac{C_x}{n} - \frac{k^2 - 2k}{8} \frac{C_x^2}{n} \right] \]  

(16)

Squaring both sides of (15) and after simplification,

\[ \left( t_2 - \sigma_y^2 \right)^2 = \sigma_y^4 \left[ e_0 + \frac{k^2}{4} e_1^2 - ke_0 e_1 \right] \]  

(17)

Taking expectations of (17) and using notations, the \( \text{MSE} \) of estimator \( t_2 \) is calculated as

\[ \text{MSE}(t_2) = \frac{\sigma_y^4}{n} \left[ \frac{A_y \theta_x}{\theta_x} + \frac{k^2}{4} C_x^2 - k \lambda C_x \theta_x \right] \]  

(18)
Differentiating equation (18) with respect to k and equating to zero and after simplification the optimum value of k is

\[ k^* = 2 \frac{\lambda \theta_x}{C_x} \]  

(19)

Putting the optimum value of k from (19) to (18), results in the minimum MSE of estimator \( t_2 \) as

\[ MSE(t_2)_{\min} = \frac{\sigma^4}{n} \left[ A_y - \lambda^2 \theta_x \right] \]  

(20)

**Remark:**

Singh and Karpe (2009) defined a class of estimator for \( \sigma^2_y \) as

\[ t_d = \hat{\sigma}^2_y d(b) \]  

(21)

where, \( d(b) \) is a function of \( b \) such that \( d(1) = 1 \), and certain other conditions, similar to those given in Srivastava (1971). The minimum MSE of \( t_d \) is given by,

\[ MSE(t_d)_{\min} = \frac{\sigma^4}{n} \left[ A_y - \lambda^2 \theta_x \right] \]  

(22)

which is the same as the minimum MSE of estimator \( t_2 \), given in equation (20).

**A General Class of Estimators**

A general class of estimator \( t_3 \) is proposed as

\[ t_3 = \left[ m_1 \hat{\sigma}^2_y + m_2 \left( \mu_x - \bar{x} \right) \right] \left\{ 2 \left( \frac{\bar{x}}{\mu_x} \right)^\alpha \exp \left[ \frac{\beta (\bar{x} - \mu_x)}{(\bar{x} + \mu_x)} \right] \right\} \]  

(23)

Where \( m_1 \) and \( m_2 \) are constants chosen so as to minimize the mean squared error of the estimator \( t_3 \).

Equation (23) can be expressed in terms of \( e \)'s as
A CLASS OF ESTIMATORS FOR FINITE POPULATION VARIANCE

\[
t_3 = \left[ m_i \sigma_y^2 + m_i \sigma_y^2 e_0 - m_2 \mu_e e_i \right] \left[ 1 - \frac{k}{2} e_i - \frac{(k^2 - 2k)}{8} e_i^2 \right]
\]  
(24)

Expanding equation (24) and subtracting \( \sigma_y^2 \) from both sides, results in

\[
(t_3 - \sigma_y^2) = \left[ (m_i - 1) \sigma_y^2 - \frac{k}{2} m_i \sigma_y^2 e_i + m_i \sigma_y^2 e_0 - m_2 \mu_e e_i \right.

\frac{e_i^2}{8} \sigma_y^2 m_i \left( k^2 - 2k \right) - \frac{\sigma_y^2 m_k}{2} e_0 e_i + \frac{k}{2} m_2 \mu_e e_i^2 \left] \right.
\]  
(25)

On taking expectations of both sides of (25) the bias of the estimator \( t_3 \) up to the first order approximation is obtained as

\[
\text{Bias}(t_3) = (m_i - 1) \sigma_y^2 - \frac{1}{8} \sigma_y^2 m_i \left( k^2 - 2k \right) \frac{C_x^2}{n \theta_x} - \frac{\sigma_y^2 m_k}{2} \frac{\lambda C_x}{n} + \frac{k}{2} m_2 \mu_e \frac{C_x^2}{n \theta_x}
\]  
(26)

Squaring both sides of (25), results in

\[
(t_3 - \sigma_y^2)^2 = \left[ (m_i - 1) \sigma_y^2 - \frac{k}{2} m_i \sigma_y^2 e_i + m_i \sigma_y^2 e_0 - m_2 \mu_e e_i \right]^2
\]  
(27)

Simplifying equation (27) and taking expectations both sides the \( \text{MSE} \) of estimator \( t_3 \) up to the first order of approximation is obtained as

\[
\text{MSE}(t_3) = \left[ (1 - 2m_i) \sigma_y^4 + m_i^2 P + m_i^2 Q - m_i m_2 R \right]
\]  
(28)

where \( P = \left( 1 + \frac{A_i}{n} + \frac{k^2 C_x^2}{4 n \theta_x} - \frac{k}{n} \lambda C_x \right) \sigma_y^4 \), \( Q = \mu_x \frac{\sigma_y^2 C_x}{n \theta_x} \) and \( R = \sigma_y^2 \left( \frac{C_x^2}{\theta_x} + 2 \lambda C_x \right) \mu_x \).

Minimizing \( \text{MSE} \) \( t_3 \) with respect to \( m_1 \) and \( m_2 \) the optimum values of \( m_i \) and \( m_2 \) is
Putting the optimum values of \( m_1 \) and \( m_2 \) in equation (28) results in the minimum \( MSE \) of estimator \( t_3 \) as

\[
MSE(t_3) = \sigma_y^4 \left[ 1 + \frac{4\sigma_y^4 Q}{4PQ - R^2} \right] \tag{29}
\]

### Empirical Study

#### Data Statistics:

The data used for empirical study was taken from Gujrati and Sangeetha (2007) - pg, 539., where,

\begin{align*}
Y_i &= \text{True consumption expenditure}, \\
X_i &= \text{True income}, \\
y_i &= \text{Measured consumption expenditure}, \\
x_i &= \text{Measured income}.
\end{align*}

From the data given we get the following parameter values:

| Table 1. Parameter values from empirical data |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| \( N \)         | \( \mu_y \)     | \( \mu_x \)     | \( \sigma_y^2 \) | \( \sigma_x^2 \) | \( \rho \)       |
| 10              | 127             | 170             | 1278            | 3300            | 0.964           |
|                 |                 |                 |                 |                 | 36.0            |
|                 |                 |                 |                 |                 | 36.0            |

| Table 2. Showing the \( MSE \) of the estimators with and without measurement errors |
|------------------------------------|--------------------|-----------------|--------------------|
| Estimators                         | \( MSE \) without meas. Error | Contribution of meas. Errors in \( MSE \) | \( MSE \) with meas. Errors |
| \( \hat{\sigma}_y^2 \)            | 245670             | 35458           | 281128             |
| \( t_1 \)                          | 229734             | 30354           | 260088             |
Table 2 continued.

<table>
<thead>
<tr>
<th>Estimators</th>
<th>MSE without meas. Error</th>
<th>Contribution of meas. Errors in MSE</th>
<th>MSE with meas. Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_{2 \min}$</td>
<td>245411</td>
<td>35461</td>
<td>280872</td>
</tr>
<tr>
<td>$t_{3 \min}(\alpha = 1, \beta = 0)$</td>
<td>247440</td>
<td>30442</td>
<td>277862</td>
</tr>
<tr>
<td>$\alpha = 0, \beta = 1$</td>
<td>234402</td>
<td>30555</td>
<td>267957</td>
</tr>
<tr>
<td>$\alpha = 1, \beta = 1$</td>
<td>268144</td>
<td>30219</td>
<td>298363</td>
</tr>
<tr>
<td>$\alpha = 1, \beta = 1$</td>
<td>234402</td>
<td>33555</td>
<td>267957</td>
</tr>
<tr>
<td>$\alpha = 0, \beta = 1$</td>
<td>231969</td>
<td>30600</td>
<td>262569</td>
</tr>
<tr>
<td>$\alpha = -0.9, \beta = 2$</td>
<td>229145</td>
<td>30365</td>
<td>259510</td>
</tr>
</tbody>
</table>

**Conclusion**

Table 2 shows that the MSE of proposed estimator $t_3$ (for $\alpha = -0.9, \beta = 2$) is minimum among all other estimators considered. It is also observed that the effect due to measurement error on the estimator $t_1$ and usual estimators is less than the effect on the estimator $t_2$ under measurement error for this given data set.

**References**


