Brazilian Journal of Probability and Statistics (2002), 16, pp. 39–46.

ROLE OF REGRESSION ESTIMATOR INVOLVING MEASUREMENT ERRORS

Maneesha and **R. Karan Singh**

Department of Statistics, Lucknow University, Lucknow 226 007, India.
Email: rkaransingh@yahoo.com

Summary
The effects of measurement errors on usual linear regression estimator are examined. A comparative study is made among the linear regression estimator, the mean per unit estimator and the ratio estimator in the presence of measurement errors.

Key words: Bias; efficiency; mean square error; observational error.

1 Introduction

For a simple random sampling scheme (see Sukhatme et al., 1984), let (x_i, y_i) be observed values instead of the true values (X_i, Y_i) on the two characteristics (X, Y) respectively for the ith $(i = 1, 2, \ldots, n)$ unit in the sample of size n. Let the observational or measurement errors be

$$u_i = y_i - Y_i \tag{1.1}$$
$$v_i = x_i - X_i, \tag{1.2}$$

which are stochastic in nature with mean zero and variances σ_U^2 and σ_V^2 respectively, and are independent.

Further, let the population means of (X, Y) be (μ_X, μ_Y), population variances of (X, Y) be (σ_X^2, σ_Y^2) and σ_{XY} and ρ be the population covariance and the population correlation coefficient between X and Y respectively.

Assuming μ_X to be known, for the estimation of population mean μ_Y, the ratio estimator t_R and the linear regression estimator $\overline{y}_{1r}$ are

$$t_R = \left(\frac{\overline{y}}{\overline{x}}\right)\mu_X \tag{1.3}$$

and

$$\overline{y}_{lr} = \overline{y} + b(\mu_X - \overline{x}), \tag{1.4}$$

where $(\overline{y}, \overline{x})$ are the means of the sample observations on (Y, X) respectively,

$$b = \frac{s_{xy}}{s_x^2} \qquad s_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

and

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

2 Bias and mean square error (MSE) of $\overline{y}_{1r}$

Let

$$\begin{aligned}
W_U &= \frac{1}{n^{1/2}}\sum_{i=1}^{n} u_i, \quad W_Y = \frac{1}{n^{1/2}}\sum_{i=1}^{n}(Y_i - \mu_Y),\\
W_V &= \frac{1}{n^{1/2}}\sum_{i=1}^{n} v_i, \quad W_X = \frac{1}{n^{1/2}}\sum_{i=1}^{n}(X_i - \mu_X),\\
c_X &= \frac{\sigma_X}{\mu_X} \quad \text{and} \quad c_Y = \frac{\sigma_Y}{\mu_Y}.
\end{aligned}$$

We have

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n}\left[(y_i - Y_i) + (Y_i - \mu_Y) + \mu_Y\right]$$

or

$$\begin{aligned}
\overline{y} - \mu_Y &= \frac{1}{n}\sum_{i=1}^{n}\left[(y_i - Y_i) + (Y_i - \mu_Y)\right]\\
&= \frac{1}{n}\sum_{i=1}^{n}\left[u_i + (Y_i - \mu_Y)\right]\\
&= \frac{1}{n^{1/2}}\left[\frac{1}{n^{1/2}}\sum_{i=1}^{n} u_i + \frac{1}{n^{1/2}}\sum_{i=1}^{n}(Y_i - \mu_Y)\right]\\
&= \frac{1}{n^{1/2}}[W_U + W_Y]
\end{aligned}$$

or

$$\overline{y} = \frac{1}{n^{1/2}}[W_U + W_Y] + \mu_Y. \tag{2.1}$$

Similarly,

$$\overline{x} = \frac{1}{n^{1/2}}[W_V + W_X] + \mu_X. \tag{2.2}$$

From (A.3) in Appendix and (2.1), substituting the values of $b(\mu_X - \overline{x})$ and $\overline{y}$ in $\overline{y}_{lr} = \overline{y} + b(\mu_X - \overline{x})$, we have

$$\begin{aligned} \overline{y}_{lr} - \mu_Y \quad = \quad & \frac{1}{n^{1/2}}(W_U + W_Y) + \frac{\sigma_{XY}}{\sigma_X^2}\left[-\frac{1}{n^{1/2}}(W_V + W_X)\right. \\ & -\frac{1}{n\sigma_{XY}}\{W_V(W_{XY} + W_{YV} + W_{XU} + W_{UV} \\ & -\mu_X W_U - \mu_X W_Y - \mu_Y W_V - \mu_Y W_X) \\ & +W_X(W_{XY} + W_{YV} + W_{XU} + W_{UV} - \mu_X W_U - \mu_X W_Y \\ & -\mu_Y W_V - \mu_Y W_X)\} + \frac{1}{n\sigma_X^2}\{W_V(W_{X^2} + 2W_{XV} \\ & +W_{V^2} - 2\mu_X W_V - 2\mu_X W_X) + W_X(W_{X^2} + 2W_{XV} \\ & \left. +W_{V^2} - 2\mu_X W_V - 2\mu_X W_X)\} + O_p\left(\frac{1}{n^{3/2}}\right)\right]. \end{aligned} \tag{2.3}$$

Taking expectation on both sides of (2.3), the bias of $\overline{y}_{lr}$ up to order $O(n^{-1})$ is

$$\begin{aligned} \text{Bias } (\overline{y}_{lr}) = & \frac{\sigma_{XY}}{n\sigma_X^2}\left[-\frac{1}{\sigma_{XY}}(\mu_{012} - \mu_Y\mu_{002} + \mu_{210} - \mu_X\mu_{110} - \mu_Y\mu_{200})\right. \\ & \left. +\frac{1}{\sigma_X^2}(2\mu_{102} + \mu_{003} - 2\mu_X\mu_{002} + \mu_{300} + \mu_{102} - 2\mu_X\mu_{200})\right], \end{aligned} \tag{2.4}$$

where $\mu_{rst} = E[(X-\mu_X)^r(Y-\mu_Y)^s V^t]$, $\mu_{110} = \sigma_{XY}$, $\mu_{200} = \sigma_X^2$, $\mu_{020} = \sigma_Y^2$ and $\mu_{002} = \sigma_V^2$.

Squaring both sides of (2.3) and taking expectation, the mean square error of $\overline{y}_{lr}$ up to order $O(n^{-1})$, is

$$\begin{aligned} \text{MSE}(\overline{y}_{lr}) \quad = \quad & \frac{1}{n}E(W_U^2 + W_Y^2 + 2W_U W_Y) \\ & +\frac{1}{n}\left(\frac{\sigma_{XY}}{\sigma_X^2}\right)^2 E(W_V^2 + W_X^2 + 2W_V W_X) \\ & -\frac{2}{n}\frac{\sigma_{XY}}{\sigma_X^2} E(W_U W_V + W_U W_X + W_Y W_V + W_Y W_X) \\ = \quad & \frac{1}{n}\left[(\sigma_U^2 + \sigma_Y^2) + \left(\frac{\sigma_{XY}}{\sigma_X^2}\right)^2 (\sigma_V^2 + \sigma_X^2) - 2\frac{\sigma_{XY}}{\sigma_X^2}\sigma_{XY}\right] \end{aligned}$$

$$\begin{aligned} &= \frac{1}{n}\left[(\sigma_U^2+\sigma_Y^2)+\frac{\rho^2\sigma_Y^2}{\sigma_X^2}(\sigma_V^2+\sigma_X^2)-2\rho^2\sigma_Y^2\right] \\ &= \frac{1}{n}\sigma_Y^2(1-\rho^2)+\frac{1}{n}\left[\sigma_U^2+\frac{\rho^2\sigma_Y^2}{\sigma_X^2}\sigma_V^2\right] \\ &= M_l^*+M_l, \end{aligned} \tag{2.5}$$

where $M_l^* = \sigma_Y^2(1-\rho^2)/n$ is the mean squared error of $\overline{y}_{lr}$ without measurement errors and

$$M_l = \frac{1}{n}\left[\sigma_U^2+\rho^2\left(\frac{\sigma_Y^2}{\sigma_X^2}\right)\sigma_V^2\right]$$

is the contribution of measurement errors in case of the linear regression estimator.

3 Some remarks

(a) From Shalabh (1997), we have

$$V(\overline{y}) = \frac{1}{n}[\sigma_Y^2+\sigma_U^2] = \frac{1}{n}\sigma_Y^2+M, \tag{3.1}$$

where $M = \sigma_U^2/n$ is the contribution of measurement error for the case of mean per unit estimator $\overline{y}$.

From (2.5), the contribution of measurement error to the mean square error of regression estimator $\overline{y}_{lr}$ up to order $O(n^{-1})$ is

$$M_l = \frac{1}{n}\left[\sigma_U^2+\rho^2\left(\frac{\sigma_Y^2}{\sigma_X^2}\right)\sigma_V^2\right] = M+\left[\rho^2\left(\frac{\sigma_Y^2}{\sigma_X^2}\right)\sigma_V^2\right]. \tag{3.2}$$

The second term in (3.2) is non-negative, hence comparing (3.1) and (3.2), we see that the linear regression estimator $\overline{y}_{lr}$ is effected more by the measurement errors than the mean per unit estimator $\overline{y}$.

(b) Again, from Shalabh (1997), the mean square error of the ratio estimator t_R to the order $O(n^{-1})$, is

$$M(t_R) = \frac{\sigma_Y^2}{n}\left[1-\frac{c_X}{c_Y}\left(2\rho-\frac{c_X}{c_Y}\right)\right]+\frac{1}{n}\left[\sigma_U^2+\left(\frac{\mu_Y}{\mu_X}\right)^2\sigma_V^2\right] = M_R^*+M_R,$$

where

$$M_R^* = \frac{\sigma_Y^2}{n}\left[1-\frac{c_X}{c_Y}\left(2\rho-\frac{c_X}{c_Y}\right)\right]$$

is the mean squared error of the ratio estimator t_R without measurement errors and

$$M_R = \frac{1}{n}\left[\sigma_U^2 + \left(\frac{\mu_Y}{\mu_X}\right)^2 \sigma_V^2\right] \tag{3.3}$$

is the contribution of measurement errors in the case of the ratio estimator.

Comparing the contributions of measurement errors to the regression estimator and the ratio estimator, we see from (3.2) and (3.3) that

$$M_l < M_R$$

if

$$\frac{1}{n}\left[\sigma_U^2 + \left(\frac{\rho\sigma_Y}{\sigma_X}\right)^2 \sigma_V^2\right] < \frac{1}{n}\left[\sigma_U^2 + \left(\frac{\mu_Y}{\mu_X}\right)^2 \sigma_V^2\right]$$

or if

$$\left|\frac{\rho c_Y}{c_X}\right|^2 < 1$$

or if

$$\left|\frac{\rho c_Y}{c_X}\right| < 1. \tag{3.4}$$

From (3.4), it is clear that the regression estimator $\overline{y}_{lr}$ is less effected by the measurement errors than the ratio estimator t_R if

$$\left|\frac{\rho c_Y}{c_X}\right| < 1$$

otherwise, that is when

$$\left|\frac{\rho c_Y}{c_X}\right| > 1,$$

the regression estimator is more effected by measurement errors than the ratio estimator t_R.

Appendix

We have

$$\begin{aligned} s_x^2 &= n^{-1}\sum_{i=1}^{n} x_i^2 - \overline{x}^2 \\ &= n^{-1}\sum_{i=1}^{n}(X_i - V_i)^2 - \left\{n^{-1/2}(W_V + W_X) + \mu_X\right\}^2 \end{aligned}$$

$$
\begin{aligned}
&= n^{-1}\sum_{i=1}^{n} X_i^2 + 2n^{-1}\sum_{i=1}^{n} X_iV_i + n^{-1}\sum_{i=1}^{n} V_i^2 \\
&\quad - \left\{ n^{-1}(W_V+W_X)^2 + 2n^{-1/2}(W_V+W_X)\mu_X + \mu_X^2 \right\} \\
&= n^{-1/2}\left\{ n^{-1/2}\sum_{i=1}^{n}[X_i^2 - (\sigma_X^2+\mu_X^2) + (\sigma_X^2+\mu_X^2)] \right. \\
&\quad \left. +2n^{-1/2}\sum_{i=1}^{n} X_iV_i + n^{-1/2}\sum_{i=1}^{n} V_i^2 \right\} \\
&\quad - \left\{ n^{-1}(W_V+W_X)^2 + 2n^{-1/2}(W_V+W_X)\mu_X + \mu_X^2 \right\} \\
&= n^{-1/2}W_{X^2} + 2n^{-1/2}W_{XV} + n^{-1/2}W_{V^2} + \sigma_X^2 + \mu_X^2 \\
&\quad - \left\{ n^{-1}(W_V+W_X)^2 + 2n^{-1/2}(W_V+W_X)\mu_X + \mu_X^2 \right\} \\
&= \sigma_X^2 + n^{-1/2}W_{X^2} + 2n^{-1/2}W_{XV} + n^{-1/2}W_{V^2} \\
&\quad - \left\{ n^{-1}(W_V+W_X)^2 + 2n^{-1/2}(W_V+W_X)\mu_X \right\}, \qquad \text{(A.1)}
\end{aligned}
$$

where

$$
\begin{aligned}
W_{X^2} &= n^{-1/2}\sum_{i=1}^{n}\{X_i^2 - (\sigma_X^2+\mu_X^2)\}, \\
W_{XV} &= n^{-1/2}\sum_{i=1}^{n} X_iV_i
\end{aligned}
$$

and

$$
W_{V^2} = n^{-1/2}\sum_{i=1}^{n} V_i^2.
$$

Further,

$$
\begin{aligned}
s_{xy} &= n^{-1}\sum_{i=1}^{n} x_iy_i - \overline{xy} \\
&= n^{-1}\sum_{i=1}^{n}(X_i+V_i)(Y_i+U_i) \\
&\quad - \left\{ n^{-1/2}(W_V+W_X) + \mu_X \right\}\left\{ n^{-1/2}(W_U+W_Y) + \mu_Y \right\} \\
&= n^{-1}\sum_{i=1}^{n}(X_iY_i + V_iY_i + X_iU_i + U_iV_i)
\end{aligned}
$$

$$
\begin{aligned}
&-n^{-1}(W_V+W_X)(W_U+W_Y)-n^{-1/2}\mu_X(W_U+W_Y)\\
&-n^{-1/2}\mu_Y(W_V+W_X)-\mu_X\mu_Y\\
=\;& n^{-1/2}\left\{n^{-1/2}\sum_{i=1}^{n}(X_iY_i-\sigma_{XY}-\mu_X\mu_Y)\right.\\
&\left.+n^{-1/2}\sum_{i=1}^{n}Y_iV_i+n^{-1/2}\sum_{i=1}^{n}X_iU_i+n^{-1/2}\sum_{i=1}^{n}U_iV_i\right\}\\
&+\sigma_{XY}+\mu_X\mu_Y-n^{-1}(W_V+W_X)(W_U+W_Y)\\
&-n^{-1/2}\mu_X(W_U+W_Y)-n^{-1/2}\mu_Y(W_V+W_X)-\mu_X\mu_Y\\
=\;& \sigma_{XY}+n^{-1/2}W_{XY}+n^{-1/2}W_{YV}+n^{-1/2}W_{XU}\\
&+n^{-1/2}W_{UV}-n^{-1}(W_V+W_X)(W_U+W_Y)\\
&-n^{-1/2}\mu_X(W_U+W_Y)-n^{-1/2}\mu_Y(W_V+W_X), \qquad \text{(A.2)}
\end{aligned}
$$

where

$$
\begin{aligned}
W_{XY} &= n^{-1/2}\sum_{i=1}^{n}(X_iY_i-\sigma_{XY}-\mu_X\mu_Y),\\
W_{YV} &= n^{-1/2}\sum_{i=1}^{n}Y_iV_i,\\
W_{XU} &= n^{-1/2}\sum_{i=1}^{n}X_iU_i
\end{aligned}
$$

and

$$
W_{UV} = n^{-1/2}\sum_{i=1}^{n}U_iV_i.
$$

Using (A.1) and (A.2) in $b = s_{xy}/s_x^2$, we have

$$
\begin{aligned}
b =\;& [\sigma_{XY}+n^{-1/2}\{W_{XY}+W_{YV}+W_{XU}+W_{UV}-\mu_X(W_U+W_Y)\\
&-\mu_Y(W_V+W_X)-n^{-1/2}(W_V+W_X)(W_U+W_Y)\}]\;/\\
&[\sigma_X^2+n^{-1/2}\{W_{X^2}+2W_{XV}+W_{V^2}-2\mu_X(W_V+W_X)\\
&-n^{-1/2}(W_V+W_X)^2\}].
\end{aligned}
$$

Whence, using (2.2), we have

$$
b(\mu_X-\overline{x}) = \frac{\sigma_{XY}}{\sigma_X^2}\left[1+\frac{1}{n^{1/2}\sigma_{XY}}\{W_{XY}+W_{YV}+W_{XU}+W_{UV}\right.
$$

$$
\begin{aligned}
& -\mu_X(W_U + W_Y) - \mu_Y(W_V + W_X) \\
& -n^{-1/2}(W_V + W_X)(W_U + W_Y)\}\Big] \\
& \left[1 - \frac{1}{n^{1/2}\sigma_X^2}\{W_{X^2} + 2W_{XV} + W_{V^2} - 2\mu_X(W_V + W_X)\right. \\
& \left. -n^{-1/2}(W_V + W_X)^2 + \cdots\}\right] \left\{-n^{-1/2}(W_V + W_X)\right\} \\
= \;& \frac{\sigma_{XY}}{\sigma_X^2}\left[-n^{-1/2}(W_V + W_X) - \frac{1}{n\sigma_{XY}}(W_V + W_X)\right. \\
& \{W_{XY} + W_{YV} + W_{XU} + W_{UV} - \mu_X(W_U + W_Y) \\
& -\mu_Y(W_V + W_X)\} + \frac{1}{n\sigma_X^2}(W_V + W_X)\{W_{X^2} + 2W_{XV} \\
& \left. +W_{V^2} - 2\mu_X(W_V + W_X)\} + O_p\left(n^{-3/2}\right)\right]. \qquad \text{(A.3)}
\end{aligned}
$$

Acknowledgements

The authors wish to thank the referee for the valuable suggestions.

(Received February, 2000. Revised August, 2001.)

References

Cochran, W. G. (1963). *Sampling Techniques*, 2.ed. New Delhi: Wiley Eastern Private Limited.

Shalabh (1997). Ratio method of estimation in the presence of measurement errors. *Journal of the Indian Society of Agricultural Statistics,* **L**, 150–155.

Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S. and Ashok, C. (1984) *Sampling Theory of Surveys with Applications.* Ames: Iowa State University Press.