

A Generalized class of Consistent Estimators in Measurement Error Model

R. Karan Singh, Rajiv Saksena and R. Dwivedi

Department of Statistics, Lucknow University, Lucknow 226 007

(Received November 12, 2002)

Abstract. It is known that least squares procedure in measurement error model is not only biased but inconsistent also. To overcome the inconsistency of the least square procedure, generalized class of estimators is proposed and its properties are studied in measurement error model

1. Introduction

Incorporation of available additional a priori information plays vital role in overcoming the problem of inconsistency of estimators in regression analysis in the presence of measurement errors. A prior information may be available from various sources like association with experimental data and similar studies, past experience, some extraneous sources or theoretical considerations etc. For an interesting exposition of various forms of a priori information like the knowledge of one or more measurement error variances or their ratio etc., (see chap. 13, Judge *et al.*¹). In our present investigation, a priori knowledge of the variance of measurement error associated with explanatory variable is utilized for the estimation of regression coefficient in a linear regression model. The ultrastructural formulation of the measurement error model providing unified framework for handling two popular cases, viz., functional and structural models, is considered here.

2. The Model and the Generalized Consistent Estimator

Let the linear ultrastructural model be

$$\begin{aligned} Y_i &= \alpha + \beta X_i \\ (2.1) \quad y_i &= Y_i + u_i \\ x_i &= X_i + v_i \\ X_i &= m_i + w_i \quad (i = 1, 2, \dots, n) \end{aligned}$$

where α is the intercept term, β is the slope parameter, y_i and x_i are the observed values corresponding to the true values Y_i and X_i of the variable to be explained and the

explanatory variable respectively, u_i and v_i are the measurement errors associated with the variable to be explained (study variable) and the explanatory variables respectively. m_i is mean of X_i and w_i is the random error component associated with the explanatory variable.

Further, u_i 's, v_i 's and w_i 's are distributed independently of each other; $u_1, u_2, \dots, u_n$ are identically and independently distributed with mean 0 and variance σ_u^2 along with third and fourth central moments $\gamma_{1u} \sigma_u^3$ and $(\gamma_{2u} + 3)\sigma_u^4$ respectively; $v_1, v_2, \dots, v_n$ are also identically and independently distributed with mean 0 and variance σ_v^2 along with third and fourth central moments $\gamma_{1v} \sigma_v^3$ and $(\gamma_{2v} + 3)\sigma_v^4$ respectively, and the random error components $w_1, w_2, \dots, w_n$ are also identically and independently distributed with mean 0 and variance σ_w^2 along with third and fourth central moments $\gamma_{1w} \sigma_w^3$ and $(\gamma_{2w} + 3)\sigma_w^4$ respectively. For m_i being mean of X_i and $\bar{m} = \frac{1}{n} \sum_{i=1}^n m_i$, $S_{mm} = \frac{1}{n} \sum_{i=1}^n (m_i - \bar{m})^2$ tends to finite quantity as n becomes large.

When X_i is measured without error, we know that the least squares estimator of β is given by

$$(2.2) \quad b = \frac{s_{xy}}{s_{xx}}$$

where

$$s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$s_{xx} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{and} \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

The undesirable property of the least squares estimators b in measurement error model is that it is not only biased but inconsistent also. By using some kind of prior information about the distribution of error in x_i 's, we can remove the inconsistency of b . For example, prior information regarding the variance σ_v^2 is utilized to purge the contamination in b due to measurement error in x_i 's and thus yielding the immaculate estimator of β to be

$$(2.3) \quad b_1 = \frac{s_{xy}}{s_{xx} - \sigma_v^2}$$

which is now consistent Madansky², Schneeweiss³ and Fuller⁴ (see ch. 1 for more details).

In the light of our knowledge

$$(2.4) \quad p \lim_{n \rightarrow \infty} s_{xy} = \beta (s_{mm} + \sigma_w^2)$$

and

$$(2.5) \quad p \lim_{n \rightarrow \infty} s_{xx} = s_{mm} + \sigma_w^2 + \sigma_v^2$$

the reason why the estimator b_1 in (2.3) is consistent is that the denominator $s_{xx} - \sigma_v^2$ in b_1 satisfies the condition

$$(2.6) \quad p \lim_{n \rightarrow \infty} (s_{xx} - \sigma_v^2) = s_{mm} + \sigma_w^2.$$

Thus, replacing denominator of b_1 in (2.3) by a function $f(s_{xx})$ of s_{xx} satisfying the condition

$$(2.7) \quad p \lim_{n \rightarrow \infty} f(s_{xx}) = s_{mm} + \sigma_w^2,$$

we get a generalized consistent estimator

$$(2.8) \quad b_f = \frac{s_{xy}}{f(s_{xx})}$$

having b_1 and some more in the literature as its special cases.

The consistency of the generalized estimator b_f representing a class of estimators in measurement error model may be easily proved under the condition (2.7). We have

$$\begin{aligned} b_f - \beta &= \frac{s_{xy}}{f(s_{xx})} - \beta \\ &= [s_{xx} - \beta f(s_{xx})] [f(s_{xx})]^{-1} \end{aligned}$$

so that

$$\begin{aligned} p \lim_{n \rightarrow \infty} (b_f - \beta) &= \left[p \lim_{n \rightarrow \infty} (s_{xy}) - \beta p \lim_{n \rightarrow \infty} f(s_{xx}) \right] \left[p \lim_{n \rightarrow \infty} f(s_{xx}) \right]^{-1} \\ &= \left[\beta(s_{mm} + \sigma_w^2) - \beta(s_{mm} + \sigma_w^2) \right] [s_{mm} + \sigma_w^2]^{-1} \end{aligned}$$

{from (2.4) and (2.7)}

$$(2.9) \quad = 0$$

showing that b_f is consistent for β .

3. Some Remarks

(a) The estimators

$$b_1 = \frac{s_{xy}}{s_{xx} - \sigma_v^2}$$

and by Srivastava and Shalabh⁵

$$b_{g^*} = \frac{s_{xy}}{s_{xx} - (1 - \frac{g^*}{n}) \sigma_v^2}$$

for non-stochastic $g^* (> 0)$ being independent of n , are the special cases of the generalized estimator b_f for $f(s_{xx}) = s_{xx} - \sigma_v^2$ and $s_{xx} - (1 - \frac{g^*}{n}) \sigma_v^2$ satisfying the condition (2.7), that is,

$$\begin{aligned} p \lim_{n \rightarrow \infty} (s_{xx} - \sigma_v^2) &= p \lim_{n \rightarrow \infty} \left\{ (s_{xx} - \frac{g^*}{n}) \sigma_v^2 \right\} \\ &= s_{mm} + \sigma_v^2. \end{aligned}$$

(b) For the g -class estimator

$$(3.1) \quad b_g = \frac{s_{xy}}{s_{xx} - g \sigma_v^2}$$

and by Srivastava and Shalabh⁵ with g being the characterizing scalar,

$$f(s_{xx}) = s_{xx} - g \sigma_v^2$$

and

$$\begin{aligned} p \lim_{n \rightarrow \infty} f(s_{xx}) &= p \lim_{n \rightarrow \infty} (s_{xx} - g \sigma_v^2) \\ &= s_{mm} + \sigma_w^2 + \sigma_v^2 - p \lim_{n \rightarrow \infty} g \sigma_v^2 \\ (3.2) \quad &= s_{mm} + \sigma_w^2 - p \lim_{n \rightarrow \infty} (g - 1) \sigma_v^2, \end{aligned}$$

We see from the general consistency condition (2.7) that the condition of consistency for b_g becomes

$$(3.3) \quad p \lim_{n \rightarrow \infty} (g - 1) = 0$$

showing that the result of by Srivastava and Shalabh⁵ is a special case of the general consistency condition (2.7) of this study.

(c) Some more consistent estimators satisfying the general consistency condition (2.7) from the class b_f may be easily found. For example, for g, k, k_1 and k_2 being the characterizing scalars independent of n , the estimators

$$(3.4) \quad (i) \quad b_{f_1} = \frac{s_{xy}}{s_{xx} - (1 - \frac{g}{n})^k \sigma_v^2}$$

and

$$(3.5) \quad (ii) \quad b_{f_2} = \frac{s_{xy} \{1 + (\frac{k_1}{n})\}^{k_2}}{s_{xx} - (1 - \frac{g}{n})^k \sigma_v^2}$$

having

$$f(s_{xx}) = s_{xx} - (1 - \frac{g}{n})^k \sigma_v^2 \quad \text{and} \quad \frac{(1 + \frac{k_1}{n})^{k_2}}{s_{xx} - (1 - \frac{g}{n})^k \sigma_v^2}$$

respectively and satisfying consistency condition (2.7), belong to the generalized class b_f of consistent estimators.

References

1. G.G. Judge, N. E. Griffiths, R. C. Hill, H. Lutkepohl and T. C. Lee : The theory and practice of Econometrics, second Edition, Wiley, New York, (1985).
2. A. Madansky : The fitting of straight lines when both variables are subject to errors, *J. Amer. Statist. Assoc.*, **54** (1959) 173-205.
3. H. Schneeweiss : Constant estimation of regression with errors in the variables, *Metrika*, **23** (1976) 101-105.
4. W. A. Fuller : Measurement error models. Wiley, New York, (1987).
5. Anil K. Srivastava and Shalabh : A new property of Stein procedure in measurement error model, *Statistics and Probability Letters*, **32** (1997) 231-234.
6. G. R. Dolby : The ultra-structural model: A synthesis of the functional and structural relations, *Biometrika*, **63** (1976) 39-50.
7. K. Hoffman : Improved estimation of distribution parameters, Stein Type Estimators, *Teubner*, Leipzig, (1992).
8. H. Schneeweiss : Note on a linear model with errors in the variables and with trend, *Statist. Papers*, **32** (1991) 261-264.

