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 priori 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.) 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

\[ Y_i = \alpha + \beta X_i \]

\[ y_i = y_i + u_i \]

\[ x_i = X_i + v_i \]

\[ X_i = m_i + w_i \quad (i = 1, 2, ..., n) \]

where \( \alpha \) is the intercept term, \( \beta \) 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, \( u_1 \)'s, \( u_2 \)'s and \( w_1 \)'s are distributed independently of each other; \( u_1, u_2, \ldots, u_n \) are identically and independently distributed with mean 0 and variance \( \sigma_u^2 \) along with third and fourth central moments \( \gamma_{1u}, \sigma_u^3 \) and \( (\gamma_{2u} + 3)\sigma_u^4 \) respectively; \( v_1, v_2, \ldots, v_n \) are also identically and independently distributed with mean 0 and variance \( \sigma_v^2 \) 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 \( w_1, w_2, \ldots, w_n \) are also identically and independently distributed with mean 0 and variance \( \sigma_w^2 \) 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 \( \bar{m} = \frac{1}{n} \sum_{i=1}^{n} m_i \), \( \bar{S}_{w} = \frac{1}{n} \sum_{i=1}^{n} (m_i - \bar{m})^2 \) tends to finite quantity as \( n \) becomes large.

When \( X_i \) is measured without error, we know that the least squares estimator of \( \beta \) 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 \( \sigma_v^2 \) is utilized to purge the contamination in \( b \) due to measurement error in \( x_i \)'s and thus yielding the immaculate estimator of \( \beta \) to be

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

which is now consistent Madansky\(^2\), Schneeweiss\(^3\) and Fuller\(^4\) (see ch. 1 for more details).

In the light of our knowledge

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

and

\[
(2.5) \quad p \lim_{n \to \infty} s_{xx} = s_{mm} + \sigma_w^2 + \sigma_v^2
\]
A Generalized Classes of Consistent Estimators

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

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

(2.6)

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

$$p \lim_{n \to \infty} f(s_{xx}) = s_{mm} + \sigma_w^2.$$  

(2.7)

we get a generalized consistent estimator

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

(2.8)

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

$$b_f - \beta = \frac{s_{xy}}{f(s_{xx})} - \beta$$

$$= \left[ s_{xx} - \beta f(s_{xx}) \right] \left[ f(s_{xx}) \right]^{-1}$$

so that

$$p \lim_{n \to \infty} (b_f - \beta) = \left[ p \lim_{n \to \infty} (s_{xx}) - \beta p \lim_{n \to \infty} f(s_{xx}) \right] \left[ p \lim_{n \to \infty} f(s_{xx}) \right]^{-1}$$

$$= \left[ \beta (s_{mm} + \sigma_w^2) - \beta (s_{mm} + \sigma_w^2) \right] \left[ s_{mm} + \sigma_w^2 \right]^{-1}$$

{from (2.4) and (2.7)}

(2.9)

$$= 0$$

showing that $b_f$ is consistent for $\beta$.

3. Some Remarks

(a) The estimators

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

and by Srivastava and Shalabh$^5$

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

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 $s_{xx} = s_{xx} - \left(1 - \frac{g^n}{n}\right) \sigma_v^2$ satisfying the condition (2.7), that is,
\[ p \lim_{n \to \infty} (s_{xx} - \sigma_v^2) = p \lim_{n \to \infty} \{ (s_{xx} - \frac{g^*}{n}) \sigma_v^2 \} = s_{mm} + \sigma_v^2. \]

(b) For the \( g \)-class estimator

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

and by Srivastava and Shalabh\(^5\) with \( g \) being the characterizing scalar,

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

and

\[ p \lim_{n \to \infty} f(s_{xx}) = p \lim_{n \to \infty} (s_{xx} - g \sigma_v^2) = s_{mm} + \sigma_w^2 + \sigma_v^2 - p \lim_{n \to \infty} g \sigma_v^2 = s_{mm} + \sigma_w^2 - p \lim_{n \to \infty} (g - 1) \sigma_v^2, \]

(3.2)

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

\[ p \lim_{n \to \infty} (g - 1) = 0 \]

(3.3)

showing that the result of by Srivastava and Shalabh\(^5\) 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

\[
(3.4) \quad \text{(i)} \quad b_{f_1} = \frac{s_{xy}}{s_{xx} - (1 - \frac{g^*}{n}) \sigma_v^2}
\]

and

\[
(3.5) \quad \text{(ii)} \quad b_{f_2} = \frac{s_{xy}(1 + \frac{k_1}{n})}{s_{xx} - (1 - \frac{g^*}{n}) \sigma_v^2}
\]

having

\[ f(s_{xx}) = s_{xx} - (1 - \frac{g^*}{n}) \sigma_v^2 \quad \text{and} \quad \frac{(1 + \frac{k_1}{n})^2}{s_{xx} - (1 - \frac{g^*}{n}) \sigma_v^2} \]

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


2. A. Madansky: The fitting of straight lines when both variables are subject to errors, J. Amer. Statist. Assoc., 54 (1959) 173-205.


