JOURNAL OF RESEARCH of the National Bureau of Standards
Vol. 87, No. 1, January–February 1982

A Note on the Behavior of Least Squares Regression Estimates When Both Variables Are Subject to Error

Clifford Spiegelman*

National Bureau of Standards, Washington, DC 20234

April 19, 1981

For the errors in variables model $X = U + V$, $Y = \beta f(U) + W$, sufficient conditions are given for the L.S. limiting estimate of β to satisfy $P(\hat{\beta}/\beta < 1) = 1$ or $P(\hat{\beta}/\beta > 1) = 1$ as the sample size tends to infinity.

Key words: Errors in variable; structural; functional; regression; large sample, convex.

The problem of linear regression when both variables are subject to error is known to be difficult, see Madansky [1],[1] and Moran [2]. In particular under general conditions there is no consistent estimator for β in the model (1.1), (1.2) below based upon only the first two moments of X and Y. Let

$$X = U + V \tag{1.1}$$

$$Y = \beta f(U) + W, \quad \text{where} \tag{1.2}$$

U, V, and W are unobservable independent random variables with $EV = EW = 0$. In addition β is an unknown constant and f is a given function. We suppose that EX^2 and EY^2 are finite. This is known as the structural form of the errors in variables problem. Since there is a great deal of confusion in the literature between the case when U is a random variable and when U is not (the functional case), only the structural case is dealt with directly. Parallel results for the functional case can be obtained in a straightforward manner. These results will, however, restrict the values that a sequence of constants $U_1, \ldots, U_n$ can take.

The least squares estimate of β is, of course,

$$\hat{\beta} = \sum Y_i f(X_i) \Big/ \sum (f(X_i))^2 \tag{2}$$

where the observable random pairs (X_i, Y_i) $i = 1, \ldots, n$, are independent and have the same joint distribution as X and Y, see (1.1) and (1.2). It is well known that when $\beta \neq 0$, and $f(X) = X$ that

$$P\left(\frac{\hat{\beta}}{\beta} < 1\right) \to 1 \quad \text{as} \quad n \to \infty. \tag{3}$$

(The least squares estimate is biased toward zero.)

It is also known (see Kendall and Stuart [3]) that for $f(X) = X^k$, $(k = 1, 2, \ldots)$, and X and Y are normal that result (3) holds. However until now general conditions under which (3) holds were not available. We give sufficient conditions under which either (3) (Theorem 1) or the opposite result (Theorem 2)

$$P\left(\frac{\hat{\beta}}{\beta} > 1\right) \to \quad \text{as} \quad n \to \infty \quad \text{holds.} \tag{4}$$

While the author has not found these two results in the literature he believes that they may be well known by somebody.

* Center for Applied Mathematics, National Measurement Laboratory.

[1] Figures in brackets indicate literature references at the end of this paper.

THEOREM 1. *If* $f(X)^2$ *is convex and not constant a.s. then* (3) *holds for* $\beta \neq 0$.

PROOF:

$$\frac{\hat{\beta}}{\beta} = \frac{\Sigma f(U_i)f(X_i)}{\Sigma f(X_i)^2} + \frac{1}{\beta}\frac{\Sigma W_i f(X_i)}{\Sigma f(X_i)^2} = \frac{\Sigma f(U_i)f(X_i)}{\Sigma f(X_i)^2} + O_p(n^{-1/2}) \tag{5}$$

Since $f^2(X)$ is convex it follows by Jensen's inequality that

$$Ef(X)^2 \geqslant E(f(E[X|U]))^2 = Ef(U)^2 \tag{6}$$

From another application of Jensen's inequality we have

$$Ef(X)^2 > E(E[f(X)|U])^2 \tag{7}$$

In addition notice that

$$\begin{aligned} Ef(U)f(X) &= E(f(U)E[f(X)|U)] \\ \leqslant E(|f(U)||E[f(X)|U]|) &\leqslant (E(f(U)^2)E(E[f(X)|U])^2)^{1/2} \end{aligned}$$

by the Cauchy-Swartz inequality. Therefore by (6) and (7)

$$(Ef(X)^2)^2 \geqslant (Ef(U)f(X))^2$$

which implies that

$$Ef(X)^2 > Ef(U)f(X)$$

The theorem now follows by applying the strong law of large numbers to the terms $\frac{\Sigma f(U_i)f(X_i)}{n}$ and $\frac{\Sigma f(X_i)^2}{n}$ in (4).

If f^2 is not convex a positive β may be overestimated. Theorem 2 provides the necessary support for this statement.

THEOREM 2. *If* f *has two continuous derivatives and satisfies*

$$\frac{-f(z)f''(z)}{2} > (f'(z))^2 \tag{8}$$

for z *in some interval* I *then there exists distributions for* U, V, *and* W *such that (4) holds.*

PROOF: Following statements in the proof of Theorem 1 it is sufficient to show that there exists distributions for U, V, and W such that

$$Ef(U)f(X) > Ef(X)^2.$$

Since f has two continuous derivatives it follows that

$$\begin{aligned} f(X) &= f(U) + f'(U)V + f''(\Theta_1)\frac{V^2}{2} \\ f(X)^2 &= f(U)^2 + 2f(U)f'(U)V + [f''(\Theta_2)f(\Theta_2) + (f'(\Theta_2))^2]V^2 \end{aligned}$$

where Θ_1 and Θ_2 are points between U and V.

Take u_0 to be in I. Let η be any point such that

$$\inf_{u-\eta \leq z_1, z_2 \leq u+\eta} \frac{f(U_0)f''(z_1)}{2} - f(z_2)f''(z_2) > (f'(z_2))^2. \tag{9}$$

Such a point η exists because f'' is continuous, and (8) holds by hypothesis. Take V to have a two point distribution

$$P(V=\eta) = P(V=-\eta) = 1/2 .$$

Then

$$\begin{aligned} &E(f(u_0)f(X) - f^2(X)) \\ &= f(u_0)\left(f(u_0) + Ef''(\Theta_1)\frac{\eta^2}{4}\right) - \left[f(u_0)^2 + \frac{\eta^2}{2}E(f''(\Theta_2)f(\Theta_2) + f'(\Theta_2)^2)\right] > 0 . \end{aligned}$$

Finally we note that, since this last inequality is strict, U may have uniform distribution in a narrow interval around the chosen point U_0.

EXAMPLE 1: Let $f(X) = X^\alpha$, $X > 0$. Then, if $\alpha \geq 1/2$ the conditions of Theorem 1 are satisfied. On the other hand if $\alpha < 1/3$, the conditions of Theorem 2 are satisfied. This example is important for NBS standards work for concrete strength, see [4]. (However, the functional case is appropriate.) It is also important for background characterization in x-ray spectroscopy.

EXAMPLE 2: The conditions of theorem 2 are not necessary. If $f(X)$ satisfies condition (8) for Z a rational number in the unit interval and arbitrary elsewhere, then the proof of Theorem 2 can be used to construct distributions such that (4) holds.

COMMENT 1: The conditions of Theorems 1 and 2 can be used to check parameters are estimated in the new model.

$$\begin{aligned} &X \text{ as in } 1.1 \\ &Y' = \Sigma a_j h_j(U) + W \quad (\text{for example } h_j(x) = x^{j-1}) \end{aligned} \tag{1.3}$$

by considering orthogonalized h_j's, see Ferguson [5]. Since $(X-\alpha)^2$ is always convex Theorem 1 holds for the slope in the linear case when a constant term is in the model (1.1), (1.3).

COMMENT 2: While the results given here directly relate to the property of being biased toward or away from the origin they do not relate to attenuation of slope. Attenuation requires the extra condition that

$$P\left(\frac{\hat{\beta}}{\beta} > 0\right) \to 1 \quad \text{as} \quad n \to \infty .$$

To see that this extra condition may fail in Theorem 1 take

$$\begin{aligned} &f(x) = x^2 - 2cx, \text{ where } c \text{ is a fixed positive constant,} \\ &P(0 \leq U \leq c) = 1, \text{ and } P(|V| \geq 4c) = 1. \end{aligned}$$

References

[1] Mandansky, A. The Fitting of Straight Lines when Both Variables are Subject to Error, 1959 December; JASA 54(285):173–205; 1959 March; JASA CORR 54(288):812.

[2] Moran, P. Estimating Structural and Functional Relationships. J. Multivariate Anal. 1(2): 232–255; 1971 June.
[3] Kendall, M. and Stuart, A. *The Advanced Theory of Statistics*, 4th Edn., Chapter 29, volume 2, New York: Macmillan; 1979. 748 p.
[4] Carino, N.J. and Lew, H.S. Re-Examination of the Relation Between Splitting Tensile and Compressive Strength of Normal Weight Concrete (Manuscript); 1981.
[5] Ferguson, T.S. *Mathematical Statistics, A Decision Theoretic Approach*. Chapter 5.6, New York: Academic; 1967. 396 p.