On Consistency of Estimators in Simple Linear Regression[1]

Anindya ROY and Thomas I. SEIDMAN

Department of Mathematics and Statistics

University of Maryland

Baltimore, MD 21250

(*anindya@math.umbc.edu*) (*seidman@math.umbc.edu*)

We derive a property of real sequences which can be used to provide a natural sufficient condition for the consistency of the least squares estimators of slope and intercept for a simple linear regression.

KEY WORDS: Least squares estimators.

[1]This has now appeared in Calcutta Statistical Assoc. Bulletin **53**, pp. 261–264, (2003).

I. Introduction

Consider the simple linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + e_i, \quad i = 1, 2, \ldots, n, \tag{1}$$

where $\{x_i\}$ is a fixed regressor sequence and $e_i \sim$i.i.d. normal$\,(0, \sigma^2)$. Such examples are commonly encountered in practice, for example a linear trend model. It is clear that in applying (1) the regressor sequence cannot be completely arbitrary. Even if we were to have error-free data, it would obviously be necessary to have more than one x-value to determine both components of the parameter vector $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ for intercept and slope. This suggests as a natural condition for the statistical analysis that it would be desirable to avoid a situation in which $\{x_i\}$ converges to a limiting value. The point of this note is that this heuristic observation does, indeed, provide a simple sufficient condition for consistency:

Theorem: If $\{\bar{x}_n\}$ does not converge to a finite limit then one has consistency of the ordinary least squares estimators for both β_0 and β_1.

Note that the variance of the least squares estimator, $\hat{\boldsymbol{\beta}}$, of the parameter vector $\boldsymbol{\beta}$ is $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$ where $\boldsymbol{X}$ is the $n \times 2$ matrix of columns of 1 and x_i. Thus, one sufficient condition for the consistency of the whole vector is that $\|(\boldsymbol{X}'\boldsymbol{X})^{-1}\| \to 0$, i.e., that the smallest eigenvalue of $\boldsymbol{X}'\boldsymbol{X}$ diverges to infinity. However, in most common text books (Fuller 1987, Ferguson 1996) the sufficient conditions are stated individually for the intercept and slope estimates in terms of their respective variances, i.e., that

$$\begin{aligned} Var\{\hat{\beta}_0\} &= \sigma^2(n^{-1} + s_n^{-2}\bar{x}_n^2) & \longrightarrow 0, \\ Var\{\hat{\beta}_1\} &= \sigma^2 s_n^{-2} & \longrightarrow 0, \end{aligned} \tag{2}$$

where $\bar{x}_n = n^{-1}\sum_{i=1}^n x_i$ and $s_n^2 = \sum_{i=1}^n (x_i - \bar{x}_n)^2$.

In view of the sufficient conditions (2), our Theorem will be an immediate consequence of the fact, whose proof is our main contribution, that a sufficient condition for (2) is that the sequence $\{x_i\}$ is not convergent.

II. Proof

As noted in the Introduction, the following Lemma provides a natural sufficient condition for consistency of the least squares estimators, proving the Theorem.

Lemma: Let $\{x_i\}_1^\infty$ be an arbitrary sequence of real numbers which does not converge to a finite limit. Then, with $\{\bar{x}_n, s_n^2\}$ as above, $s_n^{-2}\bar{x}_n^2 \longrightarrow 0$.

Proof: Observe that $s_n^2 = \sum_{i=1}^n (x_i - \xi)^2 - n(\xi - \bar{x}_n)^2$ for arbitrary ξ so, taking $\xi = \bar{x}_{n+1}$, we have

$$s_{n+1}^2 - s_n^2 = (x_{n+1} - \bar{x}_{n+1})^2 + n(\bar{x}_{n+1} - \bar{x}_n)^2, \tag{3}$$

showing that $\{s_n^2\}$ is an increasing sequence. We now consider several cases:

1. Suppose the sequence $\{|\bar{x}_n|\}$ is unbounded. Fix n and let N be such that $|\bar{x}_N| \geq 2|\bar{x}_n|$ for all $N \geq n$. Then we will show that $|\bar{x}_k|/s_k \leq 2/\sqrt{n}$ for $k \geq N$. Fix $k \geq N$ and consider two subcases.

(a) Let $|\bar{x}_k| \geq 2|\bar{x}_n|$ so $|\bar{x}_k - \bar{x}_n| \geq |\bar{x}_k|/2$. Then

$$\begin{aligned} s_k^2 &= \sum_{j=1}^{k} |x_j - \bar{x}_k|^2 \quad \geq \quad \sum_{j=1}^{n} |x_j - \bar{x}_k|^2 \\ &= \sum_{j=1}^{k} |x_j - \bar{x}_n|^2 + n|\bar{x}_k - \bar{x}_n|^2 \quad = \quad s_n^2 + n|\bar{x}_k - \bar{x}_n|^2 \\ &\geq n|\bar{x}_k - \bar{x}_n|^2 \quad \geq \quad n|\bar{x}_k^2|/4 \end{aligned}$$

Therefore $|\bar{x}_k|/s_k \leq 2/\sqrt{n}$ for $k \geq N$.

(b) Let $|\bar{x}_k| < 2|\bar{x}_n| \leq |\bar{x}_N|$. Then $|\bar{x}_k|/s_k < |\bar{x}_N|/s_k$. Because s_k is a nondecreasing sequence we have $|\bar{x}_k|/s_k < |\bar{x}_N|/s_N$. The result then follows from the previous case.

2. Now suppose $\{|\bar{x}_n|\}$ is bounded so it is now sufficient to show that $s_n^2 \to \infty$. We again consider two subcases:

(a) Suppose $\{\bar{x}_n\}$ does not converge to a limit and so has at least two convergent subsequences with distinct limit points: $\{\bar{x}_{n_1}\}$ and $\{\bar{x}_{n_2}\}$ converging to limit points a_1 and a_2, respectively, with $a_1 \neq a_2$ so $0 < 2\delta = |a_1 - a_2|$. Given any n we can find $k_1 > n$ and then $k_2 > k_1$ such that $\bar{x}_{k_1} \in \{\bar{x}_{n_1}\}$, $\bar{x}_{k_2} \in \{\bar{x}_{n_2}\}$, and $|\bar{x}_{k_1} - \bar{x}_{k_2}| \geq \delta$. Much as in the argument for part a in case 1, we now have $s_{k_1}^2 \geq s_{k_2}^2 + k_1\delta^2 \geq s_n^2 + n\delta^2$. Since $\{s_k^2\}$ is nondecreasing, this shows that $s_k^2 \to \infty$.

(b) Finally, if $\{\bar{x}_n\}$ is convergent but $\{x_n\}$ is not, then $[x_{n+1} - \bar{x}_n] \not\to 0$, i.e., $|x_{n+1} - \bar{x}_n| \geq \delta$ infinitely often for some $\delta > 0$, so (3) gives $s_n^2 \to \infty$.

III. Summary and Discussion

We have shown that if the sequence of means, $\{|\bar{x}_n|\}$ from a sequence of values for a fixed regressor in a simple linear regression does not converge to a finite limit, then the least squares estimates of the intercept and the slope are consistent. Of course in the case the sequence of means converges to a finite limit, we can still have consistent estimators as long as the sample sum of squares of deviations, s_n^2, diverges. The sequence s_n^2 will not diverge if all but a finitely many values of the explanatory variable behave like a constant. It will be interesting to see if in the multiple regression case such characterizations are possible.

IV. References

Ferguson, T. S. (1996) A Course in Large Sample Theory. Chapman and Hall, London.

Fuller, W. A. (1987) Measurement Error Models. Wiley, NY.