Economics Bulletin

Volume 32, Issue 3

On the limit of the variation of the explanatory variable in simple linear regression model

Umberto Triacca
Dipartimento di Ingegneria e Scienze dell''Informazione e Matematica, University of L''Aquila

Abstract

The simple linear regression model tries to explain the observed values of the dependent variable in terms of those of the explanatory variable. In particular, this note considers the assumption concerning the mean square deviation of the explanatory variable. It is showed that it is not a neutral assumption because it excludes some relevant data generator processes.

The author wishes to thank the referee for helpful comments.
Citation: Umberto Triacca, (2012) "On the limit of the variation of the explanatory variable in simple linear regression model", *Economics Bulletin*, Vol. 32 No. 3 pp. 1927-1932.
Contact: Umberto Triacca - umberto.triacca@ec.univaq.it.
Submitted: March 19, 2012. **Published:** July 12, 2012.

1 Introduction

The simple linear regression model attempts to model the relationship between two variables. The model has five assumptions about the way in which the observations are generated, the so-called data generator process (DGP). In traditional elementary statistics courses, the teaching of simple linear regression model has a prominent role. This note is devoted to filling a gap present in most of the statistics and econometrics textbooks used in these courses, concerning the discussion of the assumption about the limit of the variation of the explanatory variable. We show that this assumption excludes some relevant DGPs. Finally, we think that the discussion about this assumption has pedagogical benefits. Indeed, it concerns the important notion of consistency and shows how a mathematical result (the Stolz-Cesàro Theorem) can be utilized advantageously throughout statistics and econometrics.

2 The classical assumptions

Consider a set of observed data

$$\{(x_t, y_t)\,;\ t = 1, ..., T\}\,.$$

The simple linear regression model can be summarized by the following:

1. $y_t = \alpha + \beta x_t + u_t \ \ \forall t$
2. $E(u_t) = 0 \ \ \forall t$
3. $E(u_t^2) = \sigma^2 \ \ \forall t$
4. $E(u_t u_s) = 0 \ \ t \neq s$
5. $\frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x}_T)^2 \neq 0$ and $\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x}_T)^2 = Q$, with $0 < Q < \infty$ and $\bar{x}_T = \frac{1}{T}\sum_{t=1}^{T} x_t$

There are thus three parameters to be estimated in the model, namely α, β and σ. Comments on these assumptions can be found in many textbooks. See, for example, Kennedy (2008) and Baltagi (2002). We focus on the part of assumption 5 concerning the mean square deviation of explanatory variable x, defined by:

$$\mathrm{MSD}(x)_T = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x}_T)^2.$$

First, we note that it is required that there is some variation in the explanatory variable x in the sample. In fact, the condition $\text{MSD}(x)_T \neq 0$ requires that we have at least two distinct values for x. Second, we have $\lim_{T\to\infty}\text{MSD}(x)_T = Q$, with $0 < Q < \infty$. This means that it is required that the variation in the regressor x does not become 0 and that it cannot grow without bound as the sample size T approaches infinity. It is important to note that the condition $0 < Q < \infty$ plays a central role in the estimators consistency proof.
Consider, for example, the ordinary least square (OLS) estimator of the unknown parameter β, that is

$$\hat{\beta}_{OLS} = \frac{\sum_{t=1}^{T}(x_t - \bar{x}_T)(y_t - \bar{y}_T)}{\sum_{t=1}^{T}(x_t - \bar{x}_T)^2}.$$

It is well-known that $\hat{\beta}_{OLS}$ is an unbiased estimator for β, thus in order to prove that $\hat{\beta}_{OLS}$ is a consistent estimator for β, we have to show that its variance tends to zero as T tends to infinity. We have that

$$\lim_{T\to\infty}\text{Var}(\hat{\beta}_{OLS}) = \lim_{T\to\infty}\frac{\sigma^2/T}{\text{MSD}(x)_T} = 0$$

It is clear that the second equality follows from the fact that

$$\lim_{T\to\infty}\frac{\sigma^2}{T} = 0$$

and $\text{MSD}(x)_T$ has finite limit Q different from 0.

3 The result

In this section we show that the assumption 5 is not a neutral assumption because it excludes some relevant DGPs (these DGPs do not satisfy the assumption 5). In particular, because the condition $\text{MSD}(x)_T \to \infty$ is not realistic even though there are plenty of common ways to make it happen, we will discuss cases where $\text{MSD}(x)_T \to 0$.
DGP 1. We consider a set of observed data

$$\{(x_t, y_t)\,;\ t = 1, ..., T\}$$

and we suppose that the assumptions 1-4 hold, and that

$$x_t = x_0 e^{-tk} \ \text{ for } t = 1, ..., T$$

where x_0 is an initial quantity, k is a positive constant and t is the time. This formula represents the notion of exponential decay in some quantity through time. There are a number of contexts in which this notion has an important role. For example, capital goods decline in efficiency at a constant exponential rate.

Here, we show that the assumption 5 excludes this DGP. In order to do this we will use the following well-known result.

Stolz-Cesàro Theorem. Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of real numbers. If

$$\lim_{n\to\infty} a_n = \delta < \infty,$$

then

$$\lim_{n\to\infty} \frac{a_1 + ... + a_n}{n} = \delta.$$

The idea is that since most of the numbers in $\{a_n\}_{n=1}^{\infty}$ are eventually close to δ, then $\frac{a_1+...+a_n}{n}$, which is the average of the first n terms must also be close to δ: as n gets large, the first terms become increasingly less important. For a proof of the Stolz-Cesàro Theorem, see Knopp (1954, p. 72).

Now, we observe that

$$\lim_{T\to\infty} x_T = \lim_{T\to\infty}(x_0 e^{-kT}) = 0.$$

From the Stolz-Cesàro Theorem it follows that

$$\lim_{T\to\infty} \bar{x}_T = 0,$$

and hence

$$\lim_{T\to\infty}(x_T - \bar{x}_T)^2 = 0.$$

Thus, using once again the Stolz-Cesàro Theorem, we have that

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x}_T)^2 = 0$$

We can conclude that the DGP 1 does not satisfy the assumption 5.

DGP 2. We consider a set of observed data

$$\{(x_t, y_t)\,;\ t = 1, ..., T\}\,.$$

and we suppose that the assumptions 1-4 hold, and that:

1. there exists at least a pair (x_t, x_s) such that $x_t \neq x_s$
2. there exists a $k \in \mathbb{N}$ such that $x_t = \delta$ for $t = k + 1, ..., T$.

We observe that:

$$\bar{x}_T = \frac{\sum_{t=1}^{k} x_t + \sum_{t=k+1}^{T} x_t}{T} = \delta + \frac{k}{T}(\bar{x}_k - \delta)$$

where $\bar{x}_k = \sum_{t=1}^{k} x_t$. Thus

$$\sum_{t=1}^{T}(x_t - \bar{x}_T)^2 = \sum_{t=1}^{k}(x_t - \delta - \frac{k}{T}(\bar{x}_k - \delta))^2 + \frac{k^2(T-k)}{T^2}(\bar{x}_k - \delta)^2,$$

and hence

$$\lim_{T\to\infty} \sum_{t=1}^{T}(x_t - \bar{x}_T)^2 = \sum_{t=1}^{k}(x_t - \delta)^2 < \infty.$$

It follows that:

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x}_T)^2 = 0.$$

Also the DGP 2 does not satisfy the assumption 5.

It is important to note that DGP 1 and DGP 2 are particular cases of the following DGP:

$$\{(x_t, y_t)\,;\ t = 1, ..., T\}$$

where x_t is a sequence such that $x_t \to \nu$, with $\nu \in \mathbb{R}$. We observe that all DGP's of this kind do not satisfy the assumption 5. Consider a sequence with $x_t \to \nu$. Then, $(x_t - \nu)^2 \to 0$ but also $\bar{x}_T \to \nu$ (by Stolz-Cesàro Theorem) and $(\bar{x}_T - \nu)^2 \to 0$. Moreover, $\frac{1}{T}\sum_{t=1}^{T}(x_t - \nu)^2 \to 0$ (by Stolz-Cesàro Theorem), so $[\frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x}_T)^2 + (\bar{x}_T - \nu)^2] \to 0$, hence $\frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x}_T)^2 \to 0$. Thus the assumption 5 does not hold.

4 Conclusion

In this note we have re-considered one of the assumption of the classical linear regression model about the variability of the explanatory variable and we have shown that it is not an innocuous assumption, since it excludes some important data generating processes. Moreover, we have shown how the Stolz-Cesàro Theorem can be used to demonstrate how certain data generating processes violate this assumption.

5 References

Baltagi, B.H. (2002), *Econometrics*, Springer: Berlin.

Kennedy, P. (2008), *A Guide to Econometrics Malden*, Blackwell Publishing: MA.
Knopp, K. (1954), *Theory and Application of Infinite Series*, Blackie & Son Ltd: London