ZASTOSOWANIA MATEMATYKI
VII (1964)

Z. PAWŁOWSKI (Katowice)

ON A SPECIAL CASE WHEN THE VARIANCE OF A LEAST SQUARES ESTIMATOR DOES NOT TEND TO ZERO

The purpose of the present paper is to show a rather interesting case when the estimator of the regression coefficient of a regression function estimated by the classical least squares method from time series data is not consistent. This case, although a relatively simple one, has apparently been overlooked so far in the statistical literature.

Let us assume the simplest form of the regression function

$$E(Y_t|x_t) = ax_t \tag{1}$$

which can be equivalently written as

$$Y_t = ax_t + \xi_t. \tag{2}$$

In the formulas (1) and (2) a denotes the unknown constant parameter, the x_t's are fixed real numbers(1) such that

$$0 < m \leqslant x_t \leqslant M < +\infty \quad \text{for all natural } t \tag{3}$$

and ξ_t is a random component with mathematical expectation equal to zero. We assume further that random variables $\xi_1, \xi_2, \ldots, \xi_n$ are mutually independent.

The numerical value a of the classical least squares estimator A of the parameter a is obtained by minimizing the sum of squares of the observed residuals, i.e. by finding such a which minimizes the expression

$$\sum_{t=1}^{n} e_t^2 \doteq \sum_{t=1}^{n} (y_t - ax_t)^2$$

where the e_t's are the residuals and the y_t's are the observed values of Y_t.

---

(1) Equivalently we may consider the x_t's as fixed realisations of random variables conditioning thus our analysis by the condition that $X_1 = x_1$, $X_2 = x_2$, $\ldots$, $X_n = x_n$. For equivalence of the two approaches, see for instance Johnston [2].

As is easily seen, the classical least squares estimator of the parameter is

$$(4)\qquad A = \frac{\sum_{t=1}^{n} Y_t x_t}{\sum_{t=1}^{n} x_t^2}.$$

It is easy to show that under the assumptions listed above the estimator A is unbiased. Its variance, however, may not tend to zero when the number n of (time series) observations increases to infinity. In order to show it let us find the variance of the estimator A

$$D^2(A) = D^2\left(\frac{\sum_{t=1}^{n} Y_t x_t}{\sum_{t=1}^{n} x_t^2}\right) = \frac{D^2\left(\sum_{t=1}^{n} Y_t x_t\right)}{\left(\sum_{t=1}^{n} x_t^2\right)^2}.$$

By the assumption, the random variables ξ_t $(t = 1, 2, \ldots)$ are independent and — as can be seen from (2) — the same applies to the variables Y_t. Hence

$$(5)\qquad D^2(A) = \frac{\sum_{t=1}^{n} x_t^2 D^2(Y_t)}{\left(\sum_{t=1}^{n} x_t^2\right)^2} = \frac{\sum_{t=1}^{n} x_t^2 D^2(\xi_t)}{\left(\sum_{t=1}^{n} x_t^2\right)^2}.$$

In the particular case when for every t there is $D^2(\xi_t) = \sigma^2$ we get the usual formula

$$(6)\qquad D^2(A) = \frac{\sigma^2}{\sum_{t=1}^{n} x_t^2}.$$

It is known that (6) tends to zero when the number of observations increases to infinity in such a way that the variance of the x_t's is positive and remains (approximately) constant.

The situation when $D^2(\xi_t) = \sigma^2$ for all t may not, however, occur in practice. It is possible to find a model (see reference [3]) when, due to the growing impact of other, besides x_t, factors on Y_t, the variance of the random components is an increasing function of time.

Without going into mathematical details of this model ([2]) an example will be given of the situation in which one can reasonably expect the

(²) The mathematical formulation can be found in [3].

variance of the random component to increase in time steadily. Let us suppose that we wish to estimate from statistical time series data the demand function for radio sets, say, and that the chosen form of this function is that introduced first by Stone and Rowe [4] and which has been largely used since by many authors.

Stone and Rowe assume the demand for a given consumer durable good to be dependent on such factors as the price of this commodity, the average level of consumers' income, the level of stock of this commodity already being used by the consumers and on the rate of decay of durable items. The demand function is dynamic in the sense that it relates quantities referring to different periods of time.

On the other hand, due to the fast technical progress, the variety of consumer durable goods on the market is increasing. This process means that not only new variations of already known commodities are introduced on the market but also totally new types of consumer durable goods are supplied on the market. This means, however, that the possibilities of choice and substitution are steadily increasing, i.e. new factors begin to influence the decisions of consumers and their inpact can be supposed to be ever stronger and stronger. As all these newly appearing factors are not explicitly accounted for in the demand function their effect will be revealed by the size of the random components whose variation will tend to increase. This, however, means that $D^2(\xi_t)$ is no longer constant but that it is a function of time, instead.

Suppose now that $n \to \infty$. If $D^2(\xi_t)$ is increasing with time fast enough the variance $D^2(A)$ may not tend to zero. Let us assume, for instance, that

$$D^2(\xi_t) = \beta t \sigma^2 \tag{7}$$

where $\beta > 0$. Substituting (7) into (5) we get

$$D^2(A) = \frac{\beta\sigma^2 \sum_{t=1}^{n} t x_t^2}{\left(\sum_{t=1}^{n} x_t^2\right)^2}. \tag{8}$$

We shall show now that the variance (8) does not tend to zero when $n \to \infty$. Using the assumptions expressed by the inequalities (3) we can write

$$\frac{\sum_{t=1}^{n} t x_t^2}{\left(\sum_{t=1}^{n} x_t^2\right)^2} \geqslant \frac{m^2 \sum_{t=1}^{n} t}{(nM^2)^2} = \frac{m^2 n(n-1)}{2M^4 n^2} > 0. \tag{9}$$

Obviously

$$\lim_{n\to\infty} \frac{m^2 n(n-1)}{2M^4 n^2} = \frac{m^2}{2M^4} > 0$$

and, because of (8) and (9), the variance of the estimator A does not tend to zero.

The practical implications of this result are that, when estimating a regression function by time series data, one should pay attention to the behaviour of the random components ξ_t. If there is ground for conclusion that $D^2(\xi_t)$ increases with time linearly of faster(³) it may be worth while to cut the time series used for the estimation of the parameters. The use of too long series may deteriorate the efficiency of the estimation.

An alternative method of approach may be that of using another estimation technique, namely the so-called generalized least squares (see reference [1]) in which, instead of assuming $D^2(\xi_t)$ to be constant, the observations are weighed in some way by the elements of the matrix of the variances and covariances of the random components ξ_t. In the case of the relation (2) the generalized least squares estimator of the parametr α is

$$A^* = \frac{\sum_{t=1}^{n} \frac{Y_t x_t}{\sigma_t^2}}{\sum_{t=1}^{n} \frac{x_t^2}{\sigma_t^2}} \tag{10}$$

and its variance can be shown to be

$$D^2(A^*) = \frac{1}{\sum_{t=1}^{n} \frac{x_t^2}{\sigma_t^2}} \tag{11}$$

where $\sigma_t^2 = D^2(\xi_t)$.

Now it is easy to show that even if the relation (7) holds true, i.e. when the variance of the random components increases linearly in time, $D^2(A^*)$ is still convergent to zero when $n \to \infty$.

In order to prove it let us substitute (7) into (11)

$$D^2(A^*) = \frac{1}{\frac{1}{\beta\sigma^2} \sum_{t=1}^{n} \frac{x_t^2}{t}}.$$

(³) Appropriate tests can be found, for instance, in [3].

Because of (3), there is obviously

$$\sum_{t=1}^{n} \frac{x_t^2}{t} \geqslant m^2 \sum_{t=1}^{n} \frac{1}{t}$$

and since the series $\sum_{t=1}^{\infty} 1/t$ diverges the same applies to the series $\sum_{t=1}^{\infty} x_t^2/t$ and, hence, $D^2(A^*)$ tends to zero when $n \to \infty$. This shows the advantage of using the method of generalized least squares as compared with the classical method of least squares in the simple case considered here.

By using a similar argument it can be shown that the estimator A has the same property also in the more general case when there is

$$D^2(\xi_t) = \beta t^p \sigma^2, \quad \beta > 0 \tag{12}$$

where[4] $0 < p \leqslant 1$. Since in this case the series $\sum_{t=1}^{\infty} 1/t^p$ diverges the same must apply to the series $\sum_{t=1}^{\infty} x_t^2/t^p$ and the variance of the estimator A^* is seen to tend to zero when $n \to \infty$.

On the other hand, if the variance of the random component increases in time faster than linearly then even Aitken's generalized method of least squares does not provide an estimator whose variance tends to zero when $n \to \infty$. More specifically, we shall assume now the parameter p in the relation (12) to be greater than one. Under this assumption we shall show that

$$\lim_{n \to \infty} D^2(A^*) = c \tag{13}$$

where $0 < c < +\infty$, i.e. we shall show that the variance of the generalized least squares estimator does not tend to zero when the sample size increases indefinitely.

In order to prove that when $p > 1$ the relation (13) holds true let us note that because of (3) and (11) the following inequalities are true

$$D^2(A^*) \geqslant \frac{1}{\sum_{t=1}^{n} \frac{M}{\sigma_t^2}} = \frac{1}{\frac{M}{\beta\sigma^2} \sum_{t=1}^{n} \frac{1}{t^p}},$$

$$D^2(A^*) \leqslant \frac{1}{\sum_{t=1}^{n} \frac{m}{\sigma_t^2}} = \frac{1}{\frac{m}{\beta\sigma^2} \sum_{t=1}^{n} \frac{1}{t^p}}.$$

(4) We omit here the case $p < 0$ since with t taking only natural values the relation (12) would imply a reduction of $D^2(\xi_t)$ in time. In this case, as is easily seen, both A and A^* have the variances which tend to zero as $n \to \infty$.

But, since $p > 1$, the series $\sum_{t=1}^{\infty} 1/t^p$ is convergent. Let its sum be denoted by S, where obviously $S > 0$. Hence

$$\frac{\beta\sigma}{MS} \leqslant \lim_{n\to\infty} D^2(A^*) \leqslant \frac{\beta\sigma^2}{mS}. \tag{14}$$

Since $\beta\sigma^2/MS$ is positive and $\beta\sigma^2/mS$ is finite, the relation (13) is proved.

Similar results can be obtained for more general types of estimated relations, namely for regression functions with several independent variables. The method of approach is essentially the same but the algebra involved is of course much more complicated.

It is perhaps worthwhile to point out the fact that the estimator A discussed in this note is an interesting example of an unbiased estimator which may not be consistent.

References

[1] A. C. Aitken, *On least squares and linear combination of observations*, Proc. Royal Soc., Edinburgh 1935.

[2] J. Johnston, *Econometric Methods*, New York 1963.

[3] Z. Pawłowski, *Modele ekonometryczne a regresja pierwszego i drugiego rodzaju*, Przegląd Statystyczny 2 (1962).

[4] R. Stone and D. A. Rowe, *The market demand for durable goods*, Econometrica (1957).

Received on 18. 11. 1963

Z. PAWŁOWSKI (Katowice)

O SPECJALNYM PRZYPADKU, GDY WARIANCJA OTRZYMANEGO ZA POMOCĄ METODY NAJMNIEJSZYCH KWADRATÓW ESTYMATORA NIE DĄŻY DO ZERA PRZY $n\to\infty$

STRESZCZENIE

W pracy omawia się prosty przypadek, gdy metoda najmniejszych kwadratów zastosowana do estymacji parametru a równania (2) nie daje estymatora, którego wariancja dążyłaby do zera przy $n\to\infty$. W równaniu (2) wielkości x_t traktuje się jako nielosowe, a wielkości ξ_t są składnikami losowymi o nadziejach matematycznych równych zeru. Pokazuje się, że jeśli wariancje składników losowych spełniają relację (7), to jest rosną liniowo w czasie, wtedy estymator A określony równaniem (4) jest wprawdzie nieobciążony, ale jego wariancja nie dąży do zera gdy liczba obserwacji wzrasta nieograniczenie. W tego rodzaju przypadkach lepiej stosować jest uogólnioną metodę najmniejszych kwadratów Aitkena. Odpowiedni estymator A^* parametru a ma wtedy postać (10) i jego wariancja dąży do zera gdy $n\to\infty$, jeżeli tylko wariancja składników losowych nie wzrasta w czasie szybciej niż liniowo.

З. ПАВЛОВСКИ (Катовице)

О СПЕЦИАЛЬНОМ СЛУЧАЕ, В КОТОРОМ ДИСПЕРСИЯ ОЦЕНКИ, ПОЛУЧЕННОЙ С ПОМОЩЬЮ МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ НЕ СТРЕМИТСЯ К НУЛЮ ПРИ $n \to \infty$

РЕЗЮМЕ

В статье рассматривается простой случай, в котором применение метода наименьших квадратов к параметри a в уравнении (2) не приводит к оценке, дисперсия которой стремится к нулю при $n \to \infty$. В уравнении (2) величины x_t считаются неслучайными, а величины ξ_t являются случайными слагаемыми с математическими ожиданиями равными нулю. Показано, что если дисперсии случайных слагаемых удовлетворяют условию (7), это значит линейно возрастают со временем, тогда оценка A определяемая уравнением (4) хотя и не является смещенной, но ее дисперсия не стремится к нулю когда число наблюдений неограничено увеличивается. В таких случаях лучше применять обобщенный метод наименьших квадратов Айткена. Соответствующая оценка A^* параметра a имеет тогда вид (10) и ее дисперсия стремится к нулю при $n \to \infty$ если только дисперсия случайных слагаемых возрастает со временем не быстрее чем линейно.