COWLES COMMISSION DISCUSSION PAPER: STATISTICS 381

also

STATISTICAL RESEARCH CENTER - 30629G17

NOTE: Cowles Commission Discussion Papers are preliminary materials circulated privately to stimulate private discussion and are not ready for critical comment or appraisal in publications. References in publications to Discussion Papers (other than mere acknowledgement by a writer that he has had access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

A Comment on Consistent Estimation In an Econometric Shock Model

Leo A. Goodman [1]

September 14, 1953

The following problem was discussed by Jacob Wolfowitz in a very interesting address before the Cowles Commission for Research in Economics and the Committee on Statistics, The University of Chicago.

The random variables x_i are observed for $i = 0, 1, 2, 3, \ldots, n$. These variables satisfy the equations

$$x_i = u_i + au_{i-1} \quad (i = 1, 2, 3, \ldots, n), \tag{1}$$

where the u_i are unobserved independent identically distributed random variables and a is an unknown constant, $|a| < 1$. The problem is that of estimating a from the observed values of x_i.

The following illustration of the model considered herein was brought to my attention by Professor George Tolley of the Economics Department, The University

1/ Prepared in connection with research sponsored by the Office of Naval Research and carried out at the Statistical Research Center, The University of Chicago. I am indebted to Professor Jacob Wolfowitz for bringing this problem to my attention.

of Chicago: It is believed that the yield per acre of grain during a given year depends on (weather and other) conditions during that year and also on (weather and other) conditions the preceding year. Writing x_i for the grain yield per acre in year i and u_i for the (weather and other) conditions in year i, the unknown constant a in equation (1) determines "how much" grain yield per acre in a given year depends on (weather and other) conditions in the preceding year. By studying the data on annual grain yields $x_o, x_1, x_2, \ldots,$ the unknown constant a can be estimated consistently.

Another illustration of the model considered herein is the following: Let x_i and u_i be the expenditure and income in year i, respectively. Equation (1) states that expenditure in a given year depends on income in that year and also on income in the preceding year. The unknown constant a determines "how much" expenditure in a given year depends on income in the preceding year. By studying the expenditure sequence $x_o, x_1, x_2, \ldots, x_n$, the unknown constant a can be estimated consistently.

Wolfowitz has suggested a very general method of estimation which might be used for the particular problem of obtaining a super-consistent estimator of a. In order to use this method absolutely nothing need be assumed about the distribution of the random variables u_i. An advantage of the method suggested by Wolfowitz is its extreme generality. However, computation of the estimate of a by this method is not a simple matter.

In order to obtain a simpler estimator of a, the generality of the problem will be somewhat restricted. We shall assume that the random variables u_i have a finite mean and variance. We do not assume that the mean and variances are known nor that the distribution of u_i is of any specified form. Of course, the assumption of finite mean and variance is not a serious restriction for the illustrations we have discussed. Under these assumptions we shall present a a simple estimator of a.

Theorem: Let $\sum_{i=0}^{m} x_i / (M + 1) = \bar{x}$, $x_i - \bar{x} = y_i$,

$\sum_{i=1}^{m} y_i \, y_{i-1} / \sum_{i=0}^{m} y_i^2 = z$. Then the statistic $\hat{a} = 2z / (1 + \sqrt{1 - 4z^2})$

is a consistent estimator of a.

Proof: We easily see that $E\left\{x_i\right\} = E(1 + a)$, $\sigma^2 \left\{x_i\right\} = \sigma^2 (1 + a^2)$,

$E\left\{x_i \, x_{i-1}\right\} = (1 + a)^2 E^2 + a\sigma^2$, where E and σ^2 are the mean and variance of u_i respectively. Also, x_i and x_j are independent when $j \neq i \pm 1$.

The sequence of products $x_0 \, x_1$, $x_3 x_4$, $x_6 x_7, \ldots$, $x_{3i} \, x_{3i+1}$, ... is a sequence of independent identically distributed random variables with mean $(1 + a)^2 E^2 + a\sigma^2$. Hence, by Kintchine's Theorem (see [1], p. 253) the variable $\sum_{i=0}^{m} x_{3i} \, x_{3i+1} / m$ converges in probability to $(1 + a)^2 E^2 + a\sigma^2$. The products $x_1 x_2$, $x_4 x_5$, $x_7 x_8$, ..., x_{3i+1}, x_{3i+2}, ... are independent identically distributed random variables and, hence,

$\sum_{i=0}^{m} x_{3i+1} \, x_{3i+2} / m$ also converges to $(1 + a)^2 E^2 \; a\sigma^2$. Also

$\sum_{i=1}^{m} x_{3i-1} \, x_{3i} / m$ converges to $(1 + a)^2 E^2 + a\sigma^2$. Hence,

$\sum_{i=1}^{m} x_i \, x_{i-1} / m$ also converges to $(1 + a)^2 E^2 + a\sigma^2$ (see [1], p. 254).

By reasoning similar to that used in the preceding paragraph (first consider three sequences since the x_i are not independent of each other), $\bar{x}$ converges to $E(1 + a)$ and, hence, $(\bar{x})^2$ converges to $E^2 (1 + a)^2$. Therefore,

$\sum_{0=1}^{m} x_i x_{i-1} / m - (\bar{x})^2$ converges to $a\sigma^2$. Whence $\sum_{i=1}^{m} y_i y_{i-1} / m$ converges to $a\sigma^2$.

Again by a similar reasoning, $\sum_{i=0}^{m} y_i^2 / m$ converges to $\sigma^2 (1 + a^2)$.

Hence $z = \sum_{i=1}^{m} y_i y_{i-1} / \sum_{i=0}^{m} y_i^2$ converges to $a / (1 + a^2)$, and $za^2 - a + z$ converges to zero. Hence,

$$\hat{a} = (1 - \sqrt{1 - 4z^2}) / 2z = 2z / (1 + \sqrt{1 - 4z^2})$$ converges to a.

Q.E.D.

REFERENCE

[1] Cramér, Harald, Mathematical Methods of Statistics, Princeton, Princeton University Press, 1946.