[4] A. Wald: "The fitting of straight lines if both variables are subject to error," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 284 ff.
[5] T. Haavelmo: "The probability approach in econometrics," *Econometrica*, Vol. 12 (1944), Supplement.
[6] T. Koopmans: *Linear Regression Analysis in Economic Time Series*, Haarlem, 1937.
[7] G. Tintner: "An application of the variate difference method to multiple regression," *Econometrica*, Vol. 12 (1944), pp. 97 ff.
[8] H. Hotelling: "Simplified calculation of principal components," *Psychometrica*, Vol. 1 (1936), pp. 27 ff.
[9] T. W. Anderson and M. A. Girshick: "Some extensions of the Wishart distribution," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 354 ff.

# NOTE ON THE DISTRIBUTION OF THE SERIAL CORRELATION COEFFICIENT[1]

By William G. Madow

*Bureau of the Census*

The distribution of the serial correlation coefficient when $\rho = 0$ has been previously obtained.[2] The purpose of this note is to derive the distribution of the serial correlation coefficient, using the circular definition, when $\rho \neq 0$.

Let us assume that the random variables $x_1, \cdots, x_N$ have a joint normal distribution[3] $p(x_1, \cdots, x_N \mid A, B, \mu)$ where

$$\log p(x_1, \cdots, x_N \mid A, B, \mu) = \log K_1 - \tfrac{1}{2}\left[A \sum_i (x_i - \mu)^2 + 2B \sum_i (x_i - \mu)(x_{i+L} - \mu)\right]$$

the term in the bracket is positive definite, $K_1$ is independent of the $x_i$ and if $i + L > N$ then $x_{i+L} = x_{i+L-N}$. It is then clear that $\bar{x}$, $V_N$, and ${}_LC_N$, where $\bar{x}$ is the arithmetic mean, $V_N = \sum_i (x_i - \bar{x})^2$ and

$${}_LC_N = \sum_i (x_i - \bar{x})(x_{i+L} - \bar{x})$$

are sufficient statistics with respect to the estimation of $\mu$, $A$, and $B$.

Let $V_N \, {}_LR_N = {}_LC_N$ define ${}_LR_N$, the serial correlation coefficient. Then if

[1] Presented at a meeting of the Cowles Commission for Economic Research in Chicago, January 31, 1945.

[2] See R. L. Anderson, "*Distribution of the serial correlation coefficient*", pp. 1–13 and T. Koopmans, "Serial correlation and quadratic forms in normal variables", pp. 14–33, *Annals of Math. Stat.*, Vol. XIII, No. 1, March, 1942.

[3] The expression $p(\xi_1, \cdots, \xi_m \mid \theta_1, \cdots, \theta_q)$ means the probability density or the distribution of the random variables $\xi_1, \cdots, \xi_m$ for the given values of the parameters $\theta_1, \cdots, \theta_q$. When used as an index of summation or multiplication, the letter $i$ will assume all values from 1 through $N$.