BIAS IN ESTIMATION OF SERIAL CORRELATION COEFFICIENTS

By A. SREE RAMA SASTRY

Statistical Laboratory, Calcutta
and
Engineering Research Department, Hyderabad (Dn)

Introduction

Statistical time-series may be roughly classified as *evolutive* and *stationary*. In evolutive series, different sections are more or less different as regards the average, the variance, or similar other characteristics. In stationary time-series the fluctuations up and down may seem random or show tendencies towards regularity, but in either case, the general pattern of the fluctuations is, on the whole, the same in different sections of the series. The following are four main types of stationary series.

(1) Random series
(2) Periodic series
(3) Autoregressive series
(4) Moving average series

Each of these types has important applications in statistical analysis.

In the analysis of stationary time series, the sequence of its serial correlation coefficients or its *correlogram* to use the brief term introduced by H. Wold is of fundamental importance. Empirically, the correlogram is the sequence r_k of the correlation coefficient between the given time series x_t and the series obtained by lagging the series x_t by k time-units. Theoretically the correlogram refers to variables ξ_t and ξ_{t+k} of which x_t and x_{t+k} can be regarded as samples, ρ_k being the correlation coefficient between ξ_t and ξ_{t+k}. We shall here confine our attention to such time-series where the theoretical correlogram ρ_k may be interpreted as an empirical correlogram that is formed on the basis of a time-series which is given for an infinite range of time, from $t = -\infty$ to $t = +\infty$. It will be observed that if we confine our attention to moments up to the second order a complete description of the time-series is given by its mean, its variance and its correlogram.

In the empirical correlogram, a bias creeps in because the correction for the mean is made on the basis of a finite number of observations. It is the purpose of the present paper to examine this bias. The analysis covers stationary time-series of the above types (3) and (4). The bias turns out to be rather substantial in a good many cases. Correction formulae for the bias are worked out and illustrated by numerical examples.

SECTION 1

Let $x_1, x_2 ... x_T$ denote a stationary time-series of T observations. Then the k^{th} correlation coefficient is defined as

$$r_k = \frac{\frac{1}{T-k}\sum_{s=1}^{T-k} x_s x_{s+k} - \frac{1}{(T-k)^2}\left\{\sum_{s=1}^{T-k} x_s\right\}\left\{\sum_{s=1}^{T-k} x_{s+k}\right\}}{\left[\left\{\frac{1}{T-k}\sum_{s=1}^{T-k} x_s^2 - \frac{1}{(T-k)^2}\left(\sum_{s=1}^{T-k} x_s\right)^2\right\}\left\{\frac{1}{T-k}\sum_{s=1}^{T-k} x_{s+k}^2 - \frac{1}{(T-k)^2}\left(\sum_{s=1}^{T-k} x_{s+k}\right)^2\right\}\right]^{\frac{1}{2}}} \quad ... \quad (1)$$

If the theoretical mean value of the stationary time-series is known already, then,

$$r_k = \frac{\sum_{s=1}^{T-k} x'_s x'_{s+k}}{\left[\left(\sum_{s=1}^{T-k} x'^2_s\right) \Big/ \left(\sum_{s=1}^{T-k} x'^2_{s+k}\right)\right]^{\frac{1}{2}}} \quad ... \quad (2)$$

where x'_s and x'_{s+k} denote the deviations of x_s and x_{s+k} from the theoretical mean of the series.

In this case

$$E(r_k) = E\left[\frac{\sum_{s=1}^{T-k} x'_s x'_{s+k}}{\left\{\left(\sum_{s=1}^{T-k} x'^2_s\right)\left(\sum_{s=1}^{T-k} x'^2_{s+k}\right)\right\}^{\frac{1}{2}}}\right] \quad ... \quad (3)$$

$$\simeq \frac{E\left(\sum_{s=1}^{T-k} x'_s x'_{s+k}\right)}{\left\{E\left(\sum_{s=1}^{T-k} x'^2_s\right) . E\left(\sum_{s=1}^{T-k} x'^2_{s+k}\right)\right\}^{\frac{1}{2}}}$$

or

$$\simeq \frac{E\left(\sum_{s=1}^{T-k} x'_s x'_{s+k}\right)}{E\left(\sum_{s=1}^{T-k} x'^2_s\right)} = \rho_k$$

in large samples.

We propose here to arrive at the expected values of serial correlations when sample sizes are small with the assumption that

$$E\left(\frac{A}{\sqrt{BC}}\right) = \frac{E(A)}{\sqrt{E(B).E(C)}} \quad ... \quad (4)$$

and thus obtain the bias due to the smallness of the sample size.

Let ξ_t be a stochastic variable of stationary time-series.

Suppose
$$\xi_t = b_0\eta_t + b_1\eta_{t-1} + b_2\eta_{t-2} + ... + b_p\eta_{t-p} \quad ... \quad (5)$$
where $b_0, b_1, b_2 ... b_p$ are constants and $\eta_t, \eta_{t-1}, ... \eta_{t-p}$ are random deviates with zero mean and unit standard deviation.

Here p may be finite or infinite. In the latter case suppose $\sum_{j=0}^{T-1} \sum_{h=0}^{\infty} b_h b_{h+j}$ be convergent.

Then

$$E(\xi_t) = E(b_0\eta_t + b_1\eta_{t-1} + \ldots + b_p\eta_{t-p}) = 0 \qquad \ldots \quad (6)$$

$$V(\xi_t) = E(b_0\eta_t + b_1\eta_{t-1} + \cdots + b_p\eta_{t-p})^2 = \sum_{h=0}^{p} b_h^2 = \mu_2(0) \text{ say (since } V(\eta) = 1) \ldots \quad (7)$$

Similarly

$$\text{cov}\,(\xi_t \xi_{t+k}) = b_0 b_k + b_1 b_{k+1} + \ldots + b_{p-k} b_p = \sum_{h=0}^{p-k} b_h b_{h+k} = \mu_2(k) \text{ say} \qquad \ldots \quad (8)$$

Thus in large samples from (3)

$$E(r_k) = \rho_k = \frac{\sum_{h=0}^{p-k} b_h b_{h+k}}{\sum_{h=0}^{p} b_h^2} = \frac{\mu_2(k)}{\mu_2(0)} \qquad \ldots \quad (9)$$

When the sample size is small and we only know that the expectation of the mean is zero we have to use equation (1) and obtain $E\,(r_k)$.
Thus we have

$$E(r_k) = \frac{\left(1-\frac{1}{T-k}\right)\mu_2(k) - \frac{1}{(T-k)^2}\sum_{i=1}^{T-k-1}(T-k-i)\{\mu_2(k+i)+\mu_2(|k-i|)\}}{\left(1-\frac{1}{T-k}\right)\mu_2(0) - \frac{2}{(T-k)^2}\sum_{i=1}^{T-k-1}(T-k-i)\mu_2(i)} \qquad \ldots \quad (10)$$

Bias, g_k, due to the smallness of the sample size therefore is (10)−(9) i.e.

$$g_k = -\frac{\sum_{i=1}^{T-k-1}(T-k-i)\{\mu_2(k+i)+\mu_2(|k-i|)-2\mu_2(k)\mu_2(i)\}}{(T-k-1)(T-k)\mu_2(0) - 2\sum_{i=1}^{T-k-1}(T-k-i)\mu_2(i)} \qquad \ldots \quad (11)$$

Instead of defining r_k for small samples as in equation (1) if we define

$$r_k = \frac{\frac{1}{T-k}\sum_{s=1}^{T-k} x_s x_{s+k} - \frac{1}{T^2}\left(\sum_{s=1}^{T} x_s\right)^2}{\left\{\frac{1}{T}\sum_{s=1}^{T} x_s^2 - \frac{1}{T^2}\left(\sum_{s=1}^{T} x_s\right)^2\right\}} \qquad \ldots \quad (12)$$

then,

$$E(r_k) = \frac{\mu_2(k) - \frac{1}{T}\mu_2(0) - \frac{2}{T^2}\sum_{i=1}^{T-1}(T-i)\mu_2(i)}{\left(1-\frac{1}{T}\right)\mu_2(0) - \frac{2}{T^2}\sum_{i=1}^{T-1}(T-i)\mu_2(i)} \qquad \ldots \quad (13)$$

Thus in this case

$$g_k = -\frac{(1-\rho_k)\left\{T+2\sum_{i=1}^{T-1}(T-i)\rho_i\right\}}{(T-1)T-2\sum_{i=1}^{T-1}(T-i)\rho_i} \qquad \dots \quad (14)$$

where ρ_s is as defined in equation (9).

The bias resulting from the equation (14) is less than the same obtained from (11) for some serial correlations in the beginning. But as k increases the bias given by equation (14) becomes greater than the same from equation (11) and finally tends to its maximum. Bias given by equation (11) however decreases after a certain k and finally becomes zero.

These conclusions are amply borne out by the results calculated from the two formulae *viz* (11) and (14) for the following models (vide Table 1).

$$\begin{aligned}
&\text{(A)} \quad \xi_t = \eta_t+\eta_{t-1} \\
&\text{(B)} \quad \xi_t = \eta_t+\tfrac{1}{2}\eta_{t-1} \\
&\text{(C)} \quad \xi_t = \eta_t+\eta_{t-1}+\eta_{t-2}+\eta_{t-3} \\
&\text{(D)} \quad \xi_t = .8\xi_{t-1}+\eta_t \\
&\text{(E)} \quad \xi_t = .7\xi_{t-1}-.0125\,\xi_{t-2}+\eta_t
\end{aligned} \qquad \dots \quad (15)$$

Thus the definition of r_k as per equation (1) is more acceptable than equation (12).

SECTION 2

Calculation of expectation of r_k or g_k involves the calculation of $\mu_2(s)$ for various values of s. $\mu_2(s)$ has already been defined vide relation (8).

Substituting the values of $\mu_2(s)$ in equation (10) we see that in the case of a stationary time series representable by a moving average scheme (*i.e.* when p in (5) is finite) for $k \leqslant p < T-2k+1$

$$\begin{aligned}
E(r_k) = \Bigg[&\sum_{h=0}^{p-k} b_h b_{h+k} - \frac{1}{T-k}\left\{\sum_{j=0}^{p}\sum_{h=0}^{p-j} b_h b_{h+j} + \sum_{j=1}^{p}\sum_{h=0}^{p-j} b_h b_{h+j}\right\} \\
&+\frac{1}{(T-k)^2}\left\{\sum_{j=0}^{p}\sum_{h=0}^{p-j}|k-j|\,b_h b_{h+j} + \sum_{j=1}^{p}\sum_{h=0}^{p-j}(k+j) b_h b_{h+j}\right\}\Bigg] \Bigg/ \Bigg[\left\{1-\frac{1}{(T-k)}\right\}\sum_{h=0}^{p} b_h^2 \\
&-\frac{2}{(T-k)^2}\sum_{j=1}^{p}\sum_{h=0}^{p-j}(T-k-j) b_h b_{h+j}\Bigg] \qquad \dots \quad (16)
\end{aligned}$$

and

$$\begin{aligned}
g_k = \Bigg[&\frac{1}{T-k}\left\{\sum_{j=0}^{k-1}\sum_{h=0}^{p-j} b_h b_{h+j} + \sum_{j=k+1}^{p}\sum_{h=0}^{p-j} b_h b_{h+j} + \sum_{j=1}^{p}\sum_{h=0}^{p-j} b_h b_{h+j}\right\} \\
&-\frac{1}{(T-k)^2}\Bigg[\sum_{j=0}^{p}\sum_{h=0}^{p-j}|k-j|\,b_h b_{h+j} + \sum_{j=1}^{p}\sum_{h=0}^{p-j}(k+j) b_h b_{h+j} \\
&+2\left\{\sum_{h=0}^{p-k} b_h b_{h+k}\right\}\left\{\sum_{j=1}^{p}\sum_{h=0}^{p-j}(T-k-j) b_h b_{h+j}\right\}\Bigg]\Bigg] \Bigg/ \Bigg[\left(1-\frac{1}{T-k}\right)\sum_{h=0}^{p} b_h^2 \\
&-\frac{2}{(T-k)^2}\sum_{j=1}^{p}\sum_{h=0}^{p-j}(T-k-j) b_h b_{h+j}\Bigg] \qquad \dots \quad (17)
\end{aligned}$$

For $p < k \leqslant \frac{T-1}{2}$, $E(r_k)$ will be same as (16) with the first term in the numerator vanishing and g_k will be equal to $E(r_k)$. Further equation (17) will then be identical with $E(r_k)$.

For $k > \frac{T-1}{2}$ both $E(r_k)$ and g_k will be zero. Table 2, gives the $E(r_k)$ for the models in (15) for samples of sizes ranging from 5 to 1,000. From these it is clear that in the case of models A and B a sample of size 200 terms shows a bias of the order of 0.01 in their serial correlations. Samples from model C shows this bias in their serial correlations even when their sizes exceed 300. Thus sample sizes which are considered to be very large when x's are independent, appear to be small in the analysis of time series. So formulae or tests based on large sample theory should be critically viewed or corrected for the bias therein before their application to practical problems.

Another important factorthat (17) and Table 2 (Models A—C) bring out clearly is that the bias, g_k is a function of the size of the sample, the number of the serial correlation *i.e.* k of the kth serial correlation and $b_0, b_1 \ldots b_p$. Thus in samples of size 50 and below, estimation of $b_0, b_1, \ldots b_p$ becomes more complex.

Graph 1 shows how g_k changes with the sample sizes when k and $b_0, b_1, \ldots b_p$ are kept constant. Graph 2 shows how g_k changes with k when the size of the samples and $b_0, b_1 \ldots b_p$ are kept constant.

SECTION 3

When p in (5) is infinite, we obtain a general class of stationary series which as a special case includes the autoregressive series. The autoregressive series which were first studied by G. U. Yule are defined implicitly by.

$$\xi_t + a_1\xi_{t-1} + a_2\xi_{t-2} + \cdots + a_h\xi_{t-h} = \eta_t \qquad \ldots \quad (18)$$

where $a_1, a_2 \ldots a_h$ are constants. As shown by H. Wold the a_i's should be such that the equation

$$z^h + a_1 z^{h-1} + \ldots + a_h = 0 \qquad \ldots \quad (19)$$

has the modulus of each of its real and imaginary roots less than unity; otherwise the series will not be stationary but evolutive.

It is known that a_i's of (18) and b_i's of (5) are related by the set of relations.

$$\left.\begin{array}{l} b_0 = 1 \\ a_1 + b_1 = 0 \\ a_2 + a_1 b_1 + b_2 = 0 \\ \ldots \quad \ldots \\ a_h + a_{h-1} b_1 + a_{h-2} b_2 + \ldots + a_1 b_{h-1} + b_h = 0 \\ \ldots \quad \ldots \quad \ldots \quad \ldots \\ a_h b_s + a_{h-1} b_{s+1} + a_{h-2} b_{s+2} + \ldots + a_1 b_{s+h-1} + b_{s+h} = 0 \end{array}\right\} \qquad \ldots \quad (20)$$

Thus

$$b_k = \sum \frac{(-1)^r\ a_1^{p_1} a_2^{p_2} \cdots a_h^{p_h} (p_1+p_2+\cdots+p_h)!}{p_1! p_2! \ldots p_h!}$$

where $p_1+p_2+\ldots+p_h = r$

and $p_1+2p_2+3p_3+\ldots+hp_h = k.$

Now if $\theta_1, \theta_2, \ldots \theta_h$ are the real or imaginary roots of the equation (19) with $|\theta_i| < 1$

$$a_r = \Sigma(-1)^r\ \theta_{i_1}\, \theta_{i_2}\, \theta_{i_3} \cdots \theta_{i_m} \text{ where } \theta_{i_1}, \theta_{i_2} \ldots \theta_{i_m} = 1, 2 \ldots h. \qquad \ldots \quad (22)$$

Thus b_k expressed in θ's becomes

$$b_k = \Sigma\, \theta_1^{p_1} \cdot \theta_2^{p_2} \ldots \theta_h^{p_h} \text{ where } p_1+p_2+\cdots+p_h = k \qquad \ldots \quad (23)$$

or

$$b_k = \frac{A(0,1,2\ldots, h-2, h-1+k)}{A(0,1,2\ldots, h-1)} \qquad \ldots \quad (24)$$

where $A(0,1,2\ldots, h-2, h-1+s)$ is the alternant

$$\begin{vmatrix} 1 & 1 & 1 & \ldots & 1 \\ \theta_1 & \theta_2 & \theta_3 & \ldots & \theta_h \\ \theta_1^2 & \theta_2^2 & \theta_3^2 & \ldots & \theta_h^2 \\ \ldots & \ldots & \ldots & & \ldots \\ \theta_1^{h-2} & \theta_2^{h-2} & \theta_3^{h-2} & \cdots & \theta_h^{h-2} \\ \theta_1^{h-1+s} & \theta_2^{h-1+s} & \theta_3^{h-1+s} & \cdots & \theta_h^{h-1+s} \end{vmatrix}$$

Therefore

$$\mu_2(j) = \sum_{m=0}^{\infty} b_m\, b_{m+j} = \sum_{r=1}^{h} A_r\, \theta_r^j \qquad \ldots \quad (25)$$

where

$$A_r = \sum_{s=1}^{h} \left\{ \frac{(\theta_r^{h-1})(\theta_s^{h-1})}{(1-\theta_r \theta_s)} \right\} \qquad \ldots \quad (26)$$

and (θ_q^{h-1}) is the inverse of the element θ_q^{h-1} in the alternant $(0,1,2,\ldots h-1)$ of $\theta_1, \theta_2 \ldots \theta_h$.

In large samples we therefore have from (9)

$$E(r_k) = \frac{\mu_2(k)}{\mu_2(0)} = \rho_k = \frac{\sum_{r=1}^{h} A_r \theta_r^k}{\sum_{r=1}^{h} A_r} \qquad \ldots \quad (27)$$

Further substituting the value of $\mu_2(j)$ in equation (10) we have in small samples for $k \leqslant \frac{T-1}{2}$

$$E(r_k) = \frac{\sum_{r=1}^{h} \frac{A_r}{(1-\theta_r)^2} \left\{(T-k)^2(1-\theta_r)^2\theta_r^k-(T-2k)(1-\theta_r^2)+2\theta_r^{k+1}-\theta_r^{T-2k+1}-\theta_r^{T+1}\right\}}{\sum_{r=1}^{h} \frac{A_r}{(1-\theta_r)^2} \{(T-k+1)(T-k)(1-\theta_r)^2-2(T-k)(1-\theta_r)+2\theta_r-2\theta_r^{T-k+1}\}} \quad \dots \quad (28)$$

For $k > \frac{T-1}{2}$

$$E(r_k) = \frac{\sum_{r=1}^{h} \frac{A_r}{(1-\theta_r)^2} \{(T-k)^2(1-\theta_r)^2\theta_r^k+2\theta_r^{k+1}-\theta_r^{T+1}-\theta_r^{2k-T+1}\}}{\sum_{r=1}^{h} \frac{A_r}{(1-\theta_r)^2} \{(T-k+1)(T-k)(1-\theta_r)^2-2(T-k)(1-\theta_r)+2\theta_r-2\theta_r^{T-k+1}\}} \quad \dots \quad (29)$$

Thus bias in $E(r_k)$ due to the size of the sample can be obtained by subtracting the values given by (27) from (28) or (20) as the case may be.

In particular, if in equation (18) $h = 1$ *i.e.* if we have a Markoff series

$$\xi_t - a\xi_{t-1} = \eta_t$$

then equation (27) reduces to

$$E(r_k) = \rho_k = a^k \quad \dots \quad (30)$$

In small samples for $k \leqslant \frac{T-1}{2}$

$$E(r_k) = \frac{(T-k)^2(\rho_k-2\rho_{k+1}+\rho_{k+2})-(T-2k)(1-\rho_2)+(2\rho_{k+1}-\rho_{T-2k+1}-\rho_{T+1})}{(T-k+1)(T-k)(\rho_0-2\rho_1+\rho_2)-2(T-k)(1-\rho_1)+2(\rho_1-\rho_{T-k+1})} \quad \dots \quad (31)$$

and for $k > \frac{T-1}{2}$

$$E(r_k) = \frac{(T-k)^2(\rho_k-2\rho_{k+1}+\rho_{k+2})+2\rho_{k+1}-\rho_{2k-T+1}-\rho_{T+1}}{(T-k+1)(T-k)(\rho_0-2\rho_1+\rho_2)-2(T-k)(1-\rho_1)+2(\rho_1-\rho_{T-k+1})} \quad \dots \quad (32)$$

where ρ_s is as given in (29) for large samples.

Thus for $k \leqslant \frac{T-1}{2}$

$$g_k = -\frac{(1-\rho_2)(T-2k)-\rho_k(T-k)\}+\rho_{T-2k+1}(1-\rho_{2k})}{(T-k+1)(T-k)(\rho_0-2\rho_1+\rho_2)-2(T-k)(1-\rho_1)+2(\rho_1-\rho_{T-k+1})} \quad \dots \quad (33)$$

and for $k > \frac{T-1}{2}$

$$g_k = -\frac{\rho_{2k-T+1}(1-\rho_{2(T-k)})-\rho_k(1-\rho_2)(T-k)}{(T-k+1)(T-k)(\rho_0-2\rho_1+\rho_2)-2(T-k)(1-\rho_1)+2(\rho_1-\rho_{T-k+1})} \quad \dots \quad (34)$$

Table 3 gives the bias due to sample sizes of 35 and 15 respectively for $a = .1, .2, \dots .9$. From these it is clear that for a moderate size of the sample namely 35 which we generally meet in practical work the bias is very high and increases with 'a'. Graph 3 for example, shows how the bias changes with 'a' and k of r_k. Thus in all practical problems requiring the use of Markoff series the bias has to be taken into account. Further as the bias is a function of 'a', estimation of 'a' becomes more complex.

Table 2 (Model D) gives the bias in $E(r_k)$ when the sample sizes are 5 to 1000 for $a = .8$ and $k = 1, 2..., 12$. From this it will be clear that bias even in samples of size 500 is of the order of 0.01.

When we have an autoregressive series of the type of Yule's equation viz.

$$\xi_t + a_1\xi_{t-1} + a_2\xi_{t-2} = \eta_t \qquad \dots \quad (35)$$

where

$$z^2 + a_1 z + a_2 = 0 \qquad \dots \quad (36)$$

has imaginary roots $\theta e^{i\phi}$ and $\theta e^{-i\phi}$ and $|\theta| < 1$, A_s of the equation (26) are

$$\left.\begin{aligned} A_1 &= \frac{(e^{2i\phi}-1)}{4\sin^2\phi\,(1-\theta^2)\,(1-\theta^2 e^{2i\phi})} \\ \text{and} \quad A_2 &= \frac{(e^{-2i\phi}-1)}{4\sin^2\phi\,(1-\theta^2)(1-\theta^2 e^{-2i\phi})} \end{aligned}\right\} \qquad \dots \quad (37)$$

Thus in large samples i.e. from (27)

$$E(r_k) = \rho_k = \theta^k\{\cos k\phi + \frac{1-\theta^2}{1+\theta^2}\cot\phi \sin k\phi\} \qquad \dots \quad (38)$$

a well known result.

In small samples i.e. from equations (10) and (25), for $k \leqslant \frac{T-1}{2}$,

$$\begin{aligned} E(r_k) = [&\rho_k(T-k)^2(1-\rho_1^2)(1+\theta^2)^2-(T-2k)(1-\rho_1^2)(1-\theta^4)+\rho_1(1+\theta^4) \\ &-2\theta^2+\rho_{k+1}+\theta^4\rho_{k-1}-2\theta^2\rho_k-\rho_{T-2k+1}-\theta^4\rho_{T-2k-1}+2\theta^2\rho_{T-2k}-\rho_{T+1} \\ &-\theta^4\rho_{T-1}+2\theta^2\rho_T]\,/\,[(T-k)^2(1-\rho_1)^2(1+\theta^2)^2-(T-k)(1-\rho_1^2)(1-\theta^4)+2\rho_1(1+\theta^4) \\ &-2\theta^2-\rho_{T-k+1}-\theta^4\rho_{T-k-1}+2\theta^2\rho_{T-k}] \end{aligned} \qquad \dots \quad (39)$$

and for $k > \frac{T-1}{2}$,

$$\begin{aligned} E(r_k) = [&\rho_k(T-k)^2(1-\rho_1)^2(1+\theta^2)^2+2\{\rho_{k+1}+\theta^4\rho_{k-1}-2\theta^2\rho_k\}-\{\rho_{T+1}+\theta^4\rho_{T-1}-2\theta^2\rho_T \\ &+\rho_{2k-T+1}+\theta^4\rho_{2k-T-1}+2\theta^2\rho_{2k-T}\}]\,/\,[(T-k)^2(1-\rho_1)^2(1+\theta^2)^2 \\ &-(T-k)(1-\rho_1^2)(1-\theta^4)+2\{\rho_1(1+\theta^4)-2\theta^2-\rho_{T-k+1}-\theta^4\rho_{T-k-1}+2\theta^2\rho_{T-k}\}] \end{aligned} \qquad \dots \quad (40)$$

where ρ_s is the $E(r_s)$ in large samples.

Thus for $k \leqslant \frac{T-1}{2}$,

$$\begin{aligned} g_k = -[&(1-\rho_1^2)(1-\theta^4)\{(T-2k)-\rho_k(T-k)\}-\{\rho_1(1+\theta^4)-2\theta^2\}\{1-2\rho_k\} \\ &+\{\rho_{T+1}+\rho_{T-2k+1}-\rho_{k+1}-2\rho_k\rho_{T-k+1}\}+\theta^4\{\rho_{T-1}+\rho_{T-2k-1}-\rho_{k-1}-2\rho_k\rho_{T-k-1}\} \\ &-2\theta^2\{\rho_T+\rho_{T-2k}-\rho_k-2\rho_k\rho_{T-k}\}]\,/\,[(T-k)^2(1-\rho_1)^2(1+\theta^2)^2-(T-k)(1-\rho_1^2)(1-\theta^4) \\ &+2(\rho_1-\rho_{T-k+1})+2\theta^4(\rho_1-\rho_{T-k-1})-4\theta^2(1-\rho_{T-k})] \end{aligned} \qquad \dots \quad (41)$$

and for $k > \frac{T-1}{2}$,

$$\begin{aligned} g_k = -[&\rho_k(T-k)(1-\rho_1^2)(1-\theta^4)+2\{\rho_{k+1}+\theta^4\rho_{k-1}-\rho_1\rho_k(1+\theta^4)\}-\{\rho_{T+1}+\rho_{2k-T+1} \\ &-2\rho_k\rho_{T-k+1}\}-\theta^4\{\rho_{T-1}+\rho_{2k-T-1}-2\rho_k\rho_{T-k-1}\} \\ &+2\theta^2\{\rho_T+\rho_{2k-T}-2\rho_k\rho_{T-k}\}]\,/\,[(T-k)^2(1-\rho_1)^2(1+\theta^2)^2-(T-k)(1-\rho_1^2)(1-\theta^4) \\ &+2\{\rho_1(1+\theta^4)-2\theta^2-\rho_{T-k+1}-\theta^4\rho_{T-k-1}+2\theta^2\rho_{T-k}\}] \end{aligned} \qquad \dots \quad (42)$$

Table 2 also gives the bias in $E(r_k)$ for model E for sample sizes 5 to 1000 when $k = 1, 2, \ldots 12$. From this it is seen that the bias is much less than in the case of the other models. This appears to be a special property of the autoregressive equations with imaginary roots and fairly big periods. Graph 4 shows the relation between the bias and the sample size for various values of k.

When the roots of the equation (36) are real and say are equal to p and q each being numerically less than unity then

$$\left.\begin{aligned} A_1 &= \frac{p}{(p-q)(1-pq)(1-p^2)} \\ A_2 &= \frac{q}{(q-p)(1-pq)(1-q^2)} \end{aligned}\right\} \quad \ldots \quad (43)$$

Thus in large samples we have from (27)

$$E(r_k) = \rho_k = \frac{p^{k+1}(1-q^2)-q^{k+1}(1-p^2)}{(1+pq)(p-q)} \quad \ldots \quad (44)$$

a well known result.

In small samples we have for $k \leqslant \frac{T-1}{2}$

$$\begin{aligned} E(r_k) = [&(T-k)^2(1-p)^2(1-q)^2(1+pq)(p-q)\rho_k \\ &-(T-2k)(1-pq)(p-q)(1-p^2)(1-q^2) \\ &+2p^{k+1}-p^{T-2k+1}-p^{T+1}-2q^{k+1} \\ &+q^{T-2k+1}+q^{T+1}] \,/\, [(T-k+1)(T-k)(1-p)^2(1-q)^2(1+pq)(p-q) \\ &-2(T-k)(1-pq)(p-q)(1-p)(1-q)+2p-2p^{T-k+1}-2q+2q^{T-k+1}] \quad \ldots \quad (45) \end{aligned}$$

and for $k > \frac{T-1}{2}$

$$\begin{aligned} E(r_k) = [&(T-k)^2(1-p)^2(1-q)^2(1+pq)(p-q)\rho_k \\ &+2p^{k+1}-p^{2k-T+1}-p^{T+1}-2q^{k+1} \\ &+q^{2k-T+1}+q^{T+1}] \,/\, [(T-k+1)(T-k)(1-p)^2(1-q)^2(1+pq)(p-q) \\ &-2(T-k)(1-pq)(p-q)(1-p)(1-q)+2p-2p^{T-k+1}-2q+2q^{T-k+1}] \quad \ldots \quad (46) \end{aligned}$$

where ρ_k is as given in (44).

Bias in $E(r_k)$ due to the sample size may in this case be calculated for $k \leqslant \frac{T-1}{2}$ by subtracting (44) from (45) and for $k > \frac{T-1}{2}$ by subtracting (44) from (46).

In conclusion it may be said that as samples of even size 200 give biased values, it is often necessary to make proper corrections for bias in the treatment of stationary time series.

The author wishes to thank Prof. Herman Wold and Dr. C. R. Rao for their guidance.

REFERENCE

WOLD, HERMAN (1938): A study in the analysis of stationary time series. *Dissertation, Stockholm.* Almqvist & Wicksells, Uppsala.

Table 1. Bias in $E(r_k)$ resulting from formulae (11) and (14)

k	size of sample	model A by (11)	model A by (14)	model B by (11)	model B by (14)	model C by (11)	model C by (14)	model D by (11)	model D by (14)	model E by (11)	model E by (14)
1	35	−0.0298	−0.0298	−.0322	−.0322	−.0310	−.0310	−.0587	−.0579	−.0104	−.0135
2	35	−0.0605	−0.0597	−.0542	−.0535	−.0624	−.0620	−.1069	−.1042	−.0254	−.0313
3	35	−0.0604	−0.0597	−.0540	−.0535	−.0946	−.0920	−.1457	−.1413	−.0281	−.0354
4	35	−0.0600	−0.0597	−.0530	−.0535	−.1283	−.1239	−.1784	−.1709	−.0217	−.0274
5	35	−0.0595	−0.0597	−.0531	−.0535	−.1274	−.1239	−.2042	−.1946	−.0139	−.0103
6	35	−0.0587	−0.0597	−.0524	−.0535	−.1260	−.1239	−.2237	−.2136	−.0130	−.0185
7	35	−0.0570	−0.0597	−.0515	−.0535	−.1241	−.1239	−.2383	−.2288	−.0108	−.0229
8	35	−0.0562	−0.0597	−.0502	−.0535	−.1214	−.1239	−.2483	−.2408	−.0200	−.0265
9	35	−0.0544	−0.0597	−.0486	−.0535	−.1179	−.1239	−.2532	−.2506	−.0194	−.0203
10	35	−0.0521	−0.0597	−.0465	−.0535	−.1132	−.1239	−.2536	−.2583	−.0162	−.0240
11	35	−0.0491	−0.0597	−.0439	−.0535	−.1072	−.1239	−.2490	−.2646	−.0138	−.0225
12	35	−0.0455	−0.0597	−.0405	−.0535	−.0995	−.1239	−.2390	−.2696	−.0130	−.0228
								max =	−.2895	as $k \to T$ =	−.0239
1	15	−0.0740	−0.0740	−.0795	−.0702	−.0810	−.0800	−.1533	−.1499	−.0262	−.0250
2	15	−0.1528	−0.1480	−.1352	−.1318	−.1639	−.1618	−.2790	−.2697	−.0700	−.0591
3	15	−0.1488	−0.1480	−.1315	−.1318	−.2500	−.2427	−.3804	−.3659	−.0753	−.0670
4	15	−0.1400	−0.1480	−.1235	−.1318	−.3415	−.3235	−.4432	−.4426	−.0564	−.0510
5	15	−0.1235	−0.1480	−.1087	−.1318	−.3077	−.3235	−.4677	−.5040	−.0355	−.0305
6	15	−0.0938	−0.1480	−.0823	−.1318	−.2400	−.3235	−.4376	−.5532	−.0213	−.0350
7	15	−0.0408	−0.1480	−.0357	−.1318	−.1081	−.3235	−.3037	−.6025	−.0114	−.0434
8	15	0.0000	−0.1480	.0000	−.1318	.0000	−.3235	−.2486	−.0239	−.0604	−.0501
								max =	−.7407	as $k \to T$ =	−.0452

TABLE 2. $E(r_k)$ FOR SAMPLE SIZE T

T	$k =$ 1	2	3	4	5	6	7	8	9	10	11	12
Model A												
5	.2222	−.5000	.0000	.0000	.0000							
10	.3828	−.2440	−.2222	−.1000	.0000							
15	.4260	−.1528	−.1488	−.1400	−.1235	−.0938	−.0408	.0000				
20	.4460	−.1107	−.1094	−.1067	−.1020	−.0947	−.0833	−.0861	−.0400	.0000		
30	.4649	−.0713	−.0710	−.0704	−.0694	−.0684	−.0661	−.0635	−.0600	−.0554	−.0494	−.0415
40	.4740	−.0526	−.0525	−.0522	−.0519	−.0514	−.0508	−.0499	−.0489	−.0470	−.0459	−.0439
50	.4794	−.0416	−.0415	−.0416	−.0413	−.0411	−.0408	−.0405	−.0400	−.0394	−.0488	−.0380
75	.4864	−.0274	−.0274	−.0273	−.0273	−.0272	−.0272	−.0271	−.0270	−.0269	−.0267	−.0265
100	.4898	−.0204	−.0204	−.0204	−.0204	−.0204	−.0203	−.0203	−.0202	−.0202	−.0201	−.0201
150	.4933	−.0135	−.0135	−.0135	−.0135	−.0135	−.0135	−.0135	−.0135	−.0134	−.0134	−.0134
200	.4950	−.0101	−.0101	−.0101	−.0101	−.0101	−.0101	−.0101	−.0101	−.0101	−.0101	−.0101
300	.4966	−.0067	−.0067	−.0067	−.0067	−.0067	−.0067	−.0067	−.0067	−.0067	−.0067	−.0067
500	.4980	−.0040	−.0040	−.0040	−.0040	−.0040	−.0040	−.0040	−.0040	−.0040	−.0040	−.0040
1000	.4990	−.0020	−.0020	−.0020	−.0020	−.0020	−.0020	−.0020	−.0020	−.0020	−.0020	−.0020
∞	.5000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
Model B												
5	.1042	−.4091	.0000	.0000	.0000							
10	.2744	−.2143	−.1935	−.1385	.0000							
15	.3204	−.1352	−.1316	−.1235	−.1087	−.0823	−.0357					
20	.3419	−.0985	−.0912	−.0947	−.0905	−.0839	−.0738	−.0584	−.0352			
30	.3622	−.0637	−.0634	−.0634	−.0620	−.0607	−.0590	−.0566	−.0535	−.0403	−.0440	−.0369
40	.3720	−.0471	−.0470	−.0468	−.0464	−.0460	−.0454	−.0447	−.0437	−.0425	−.0410	−.0392
50	.3778	−.0373	−.0373	−.0372	−.0370	−.0368	−.0366	−.0362	−.0358	−.0353	−.0347	−.0340
75	.3853	−.0246	−.0246	−.0245	−.0245	−.0245	−.0244	−.0243	−.0242	−.0241	−.0240	−.0238
100	.3890	−.0183	−.0183	−.0183	−.0183	−.0183	−.0183	−.0182	−.0182	−.0181	−.0181	−.0180
150	.3927	−.0121	−.0121	−.0121	−.0121	−.0121	−.0121	−.0121	−.0121	−.0121	−.0121	−.0121
200	.3946	−.0091	−.0091	−.0091	−.0091	−.0091	−.0091	−.0091	−.0091	−.0091	−.0091	−.0091
300	.3964	−.0060	−.0060	−.0060	−.0060	−.0060	−.0060	−.0060	−.0060	−.0060	−.0060	−.0060
500	.3978	−.0036	−.0036	−.0036	−.0036	−.0036	−.0036	−.0036	−.0036	−.0036	−.0036	−.0036
1000	.3989	−.0018	−.0018	−.0018	−.0018	−.0018	−.0018	−.0018	−.0018	−.0018	−.0018	−.0018
∞	.4000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000

TABLE 2. (*Contd.*)

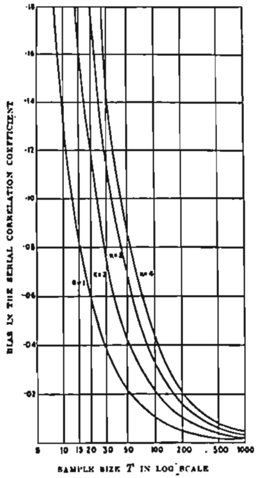
T	$k =$ 1	2	3	4	5	6	7	8	9	10	11	12
Model C												
5	.4000	.0000	+.2083	.0000								
10	.6150	.2207	−.1422	−.4706	.0000							
15	.0090	.3361	.0000	−.3415	−.3077	−.2400	−.1081	.0000				
20	.6922	.3833	.0719	−.2437	−.2353	−.2207	−.1967	−.1584	−.0076	.0000		
30	.7134	.4261	.1378	−.1525	−.1509	−.1485	−.1448	−.1397	−.1326	−.1231	−.1103	−.0934
40	.7232	.4426	.1682	−.1106	−.1101	−.1093	−.1081	−.1065	−.1045	−.1019	−.0986	−.0945
50	.7288	.4575	.1857	−.0867	−.0865	−.0861	−.0856	−.0840	−.0841	−.0830	−.0818	−.0802
75	.7362	.4722	.2081	−.0563	−.0562	−.0561	−.0560	−.0558	−.0557	−.0554	−.0551	−.0548
100	.7397	.4794	.2190	−.0416	−.0416	−.0416	−.0415	−.0415	−.0414	−.0413	−.0412	−.0411
150	.7432	.4864	.2295	−.0274	−.0274	−.0274	−.0274	−.0273	−.0273	−.0273	−.0273	−.0272
200	.7440	.4898	.2347	−.0204	−.0204	−.0204	−.0204	−.0204	−.0204	−.0204	−.0204	−.0203
300	.7466	.4933	.2399	−.0135	−.0135	−.0135	−.0135	−.0135	−.0135	−.0135	−.0135	−.0135
500	.7480	.4960	.2440	−.0081	−.0081	−.0081	−.0081	−.0081	−.0081	−.0081	−.0081	−.0081
1000	.7490	.4980	.2470	−.0040	−.0040	−.0040	−.0040	−.0040	−.0040	−.0040	−.0040	−.0040
∞	.7500	.5000	.2500	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
Model D												
5	.2787	−.1206	−.0018	.0000								
10	.5571	.2045	−.0354	−.1581	−.1240	−.0737	−.0413	.0000	.0000			
15	.6467	.3601	.1316	−.0330	−.1400	−.1705	−.1540	−.0808	−.0643	−.0400	−.0244	−.0124
20	.6896	.4380	.2355	.0770	−.0435	−.1267	−.1742	−.1878	−.1609	−.1137	−.0701	−.0504
30	.7303	.5128	.3377	.1971	.0853	−.0022	−.0698	−.1198	−.1538	−.1731	−.1784	−.1701
40	.7404	.5470	.3861	.2550	.1519	.0602	.0032	−.0484	−.0884	−.1184	−.1399	−.1537
50	.7004	.5681	.4137	.2898	.1906	.1115	.0481	−.0024	−.0422	−.0734	−.0075	−.1157
75	.7744	.5036	.4488	.3327	.2397	.1655	.1059	.0593	.0208	−.0100	−.0341	−.0533
100	.7811	.6058	.4655	.3530	.2630	.1912	.1335	.0874	.0505	.0211	−.0026	−.0213
150	.7876	.6176	.4816	.3720	.2854	.2159	.1600	.1154	.0797	.0511	.0282	.0100
200	.7008	.6234	.4894	.3821	.2962	.2270	.1729	.1289	.0938	.0657	.0432	.0252
300	.7939	.0290	.4971	.3914	.3060	.2396	.1854	.1422	.1076	.0709	.0578	.0400
500	.7964	.0334	.5031	.3987	.3152	.2488	.1953	.1526	.1184	.0911	.0692	.0517
1000	.7982	.6367	.5076	.4042	.3214	.2556	.2025	.1602	.1264	.0993	.0776	.0603
∞	.8000	.0400	.5120	.4096	.3277	.2621	.2097	.1678	.1342	.1074	.0859	.0687

TABLE 2. (Contd.)

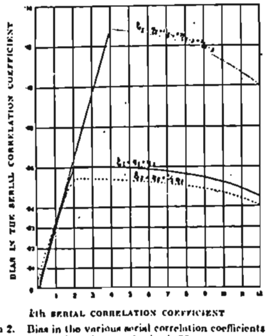
T	$k =$ 1	2	3	4	5	6	7	8	9	10	11	12
Model E												
5	.2671	−.9396	+.0734									
10	.3903	−.4318	−.6151	−.2440	.2269	.2016	.0147	+.0127				
15	.4079	−.3786	−.5571	−.2047	.1659	.2035	.0288	−.1700	−.1383	+.0020	.0817	.0920
20	.4152	−.3573	−.5347	−.1887	.1685	.2053	.0156	−.1410	−.1259	−.0028	.0632	.0478
30	.4219	−.3389	−.5161	−.1740	.1753	.2100	.0208	−.1328	−.1233	−.0214	.0452	.0324
40	.4251	−.3305	−.5062	−.1671	.1791	.2133	.0254	−.1272	−.1184	−.0184	.0465	.0314
50	.4270	−.3258	−.5010	−.1631	.1816	.2155	.0279	−.1236	−.1151	−.0159	.0485	.0330
75	.4294	−.3198	−.4944	−.1580	.1848	.2185	.0323	−.1189	−.1105	−.0120	.0518	.0363
100	.4306	−.3169	−.4912	−.1555	.1865	.2201	.0343	−.1163	−.1082	−.0100	.0537	.0381
150	.4318	−.3140	−.4880	−.1531	.1881	.2217	.0363	−.1142	−.1058	−.0079	.0556	.0400
200	.4324	−.3127	−.4865	−.1518	.1889	.2224	.0372	−.1130	−.1047	−.0069	.0565	.0410
300	.4329	−.3113	−.4849	−.1507	.1898	.2232	.0382	−.1119	−.1036	−.0058	.0575	.0420
500	.4334	−.3102	−.4837	−.1497	.1904	.2239	.0390	−.1110	−.1026	−.0050	.0583	.0427
1000	.4337	−.3094	−.4028	−.1490	.1909	.2243	.0396	−.1103	−.1020	−.0044	.0588	.0433
∞	.4341	−.3086	−.4819	−.1483	.1914	.2248	.0402	−.1096	−.1013	−.0038	.0594	.0439

TABLE 3: $E(r_k)$ AND g_k FOR VARIOUS VALUES OF 'a' IN $\xi_t = a\,\xi_{t-1}+\eta_t$ AND SAMPLES OF SIZE-35

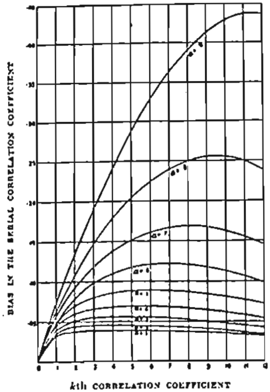
serial corre-lation	$\xi_t=.1\xi_{t-1}+\eta_t$		$\xi_t=.2\xi_{t-1}+\eta_t$		$\xi_t=.3\xi_{t-1}+\eta_t$		$\xi_t=.4\xi_{t-1}+\eta_t$		$\xi_t=.5\xi_{t-1}+\eta_t$		$\xi_t=.6\xi_{t-1}+\eta_t$		$\xi_t=.7\xi_{t-1}+\eta_t$		$\xi_t=.8\xi_{t-1}+\eta_t$		$\xi_t=.9\xi_{t-1}+\eta_t$	
	$E(r_k)$	g_k	$E(r_k)$	g_k	$E(r_k)$	g_k	$E(r_k)$	g_k	$E(r_k)$	g_k	$E(r_k)$	g_k	$E(r_k)$	g_k	$E(r_k)$	g_k	$E(r_k)$	g_k
r_1	+.0075	−.0325	+.1645	−.0355	+.2013	−.0387	+.3580	−.0420	+.4540	−.0454	+.5500	−.0491	+.6407	−.0533	+.7413	−.0587	+.8327	−.0673
r_2	−.0257	−.0357	−.0028	−.0428	+.0394	−.0506	+.1008	−.0592	+.1813	−.0687	+.2807	−.0793	+.3984	−.0916	+.5331	−.1069	+.6807	−.1293
r_3	−.0374	−.0384	−.0361	−.0441	−.0271	−.0541	−.0021	−.0661	+.0445	−.0805	+.1182	−.0978	+.2240	−.1190	+.3665	−.1457	+.5420	−.1861
r_4	−.0383	−.0384	−.0426	−.0442	−.0468	−.0549	−.0430	−.0686	−.0237	−.0862	+.0207	−.1080	+.1017	−.1384	+.2312	−.1784	+.4185	−.2376
r_5	−.0382	−.0382	−.0435	−.0438	−.0524	−.0548	−.0590	−.0692	−.0574	−.0887	−.0373	−.1161	+.0164	−.1517	+.1235	−.2042	+.3066	−.2839
r_6	−.0379	−.0379	−.0431	−.0432	−.0534	−.0541	−.0647	−.0688	−.0736	−.0892	−.0715	−.1182	−.0428	−.1604	+.0384	−.2237	+.2066	−.3248
r_7	−.0375	−.0375	−.0424	−.0424	−.0530	−.0532	−.0662	−.0678	−.0808	−.0886	−.0910	−.1190	−.0829	−.1653	−.0286	−.2383	+.1183	−.3600
r_8	−.0369	−.0369	−.0414	−.0414	−.0519	−.0520	−.0656	−.0663	−.0830	−.0869	−.1013	−.1181	−.1096	−.1671	−.0803	−.2483	−.0412	−.3893
r_9	−.0360	−.0360	−.0400	−.0400	−.0502	−.0502	−.0639	−.0642	−.0825	−.0845	−.1054	−.1155	−.1257	−.1661	−.1190	−.2532	−.0249	−.4123
r_{10}	−.0349	−.0349	−.0383	−.0383	−.0480	−.0480	−.0614	−.0615	−.0801	−.0811	−.1055	−.1115	−.1344	−.1626	−.1402	−.2536	−.0701	−.4278
r_{11}	−.0335	−.0335	−.0361	−.0361	−.0453	−.0453	−.0581	−.0581	−.0762	−.0767	−.1024	−.1060	−.1362	−.1560	−.1631	−.2490	−.1222	−.4360
r_{12}	−.0317	−.0317	−.0333	−.0343	−.0419	−.0410	−.0537	−.0537	−.0709	−.0711	−.0964	−.0986	−.1327	−.1465	−.1703	−.2390	−.1534	−.4358



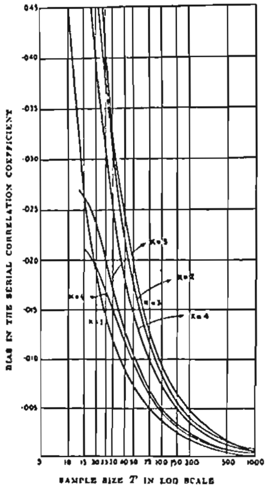
Graph 1. Bias in the kth serial correlation coefficient when the sample size is T and the model is $\xi_t = \eta_t + \eta_{t-1} + \eta_{t-2} + \eta_{t-3}$



Graph 2. Bias in the various serial correlation coefficients when the sample size is 25.



Graph 3. Bias in the serial correlation coefficients for various values of 'a' when the model is $\xi_t = a\xi_{t-1} + \eta_t$ and the sample size is 35.



Graph 4. Bias in the kth serial correlation coefficient when the sample size is T and the model is $\xi_t = .7\,\xi_{t-1} - .0125\,\xi_{t-2} + \eta_t$