XV. LEAST-SQUARES BIAS IN TIME SERIES[1]

BY LEONID HURWICZ

1.1. In this papèr it is shown that there exist cases where the least-squares and the maximum-likelihood estimates of the regression and structural coefficients are biased[2] for any finite-sized sample[3] drawn from a population defined by a noncircular stochastic difference-equation system. This bias is evaluated for certain special cases. It is found that for very small samples the bias may amount to as much as 25 per cent of the true value of the parameter[4], while for medium sized samples (say of 20 observations) the bias is still almost 10 per cent (the numerical value of the expectation of the estimate always being below the true parameter value[5]). The relative bias seems to tend to zero, although rather slowly, as the damping of the system becomes weaker.

1.2. The initial objective of this paper was to prove that the least-squares method yields biased estimates of regression coeffi-

---

[1]Part of the work on this paper was done in 1945-46 during the author's tenure of the Guggenheim Memorial Fellowship. Some of the problems considered arose in connection with the author's research at the Institute of Meteorology at the University of Chicago in 1944.

[2]I.e., the mathematical expectations of the estimates are not identically equal to the true parameter values. *Bias* is defined as the difference between the expectation of the estimate and the true parameter value. *Relative bias* is the ratio of this difference to the true parameter. The quantity N_T appearing in the formulae below is the relative bias plus 1 in samples of T observations; hence N_T equals the ratio of the expectation of the estimate to the true parameter, so that when N tends to 1 the relative bias tends to zero.

[3]Containing more than two observations.

[4]Highest relative bias, found in samples of four observations, is $26\,{}^2\!/_3$ per cent.

[5]I.e., the bias, where known, is always negative for positive parameter values and vice versa.

cients in autoregressive[1] noncircular time series. It was generally realized that the usual proofs of the Markoff Theorem [David and Neyman] were not valid for this case[2] but the author does not know of any proof of the actual existence of a bias[3].

It was possible to obtain explicit formulae for the bias in very small samples (three or four observations). This was done for two types of assumptions with regard to the initial value: (1) *fixed initial-value* case – where the initial observation may be regarded as a fixed variate (the initial value chosen being zero) see section *3.1.2* below; (2) *stochastic initial-value* case – where the initial value is a stochastic variable whose distribution is the marginal distribution of the later observations, see section *2.4.2* below. The author regards the latter assumption as being more realistic, but since the *exact* equivalence of least-squares and maximum-likelihood criteria of estimation of the regression coefficients applies only to the former (asymptotically to both), it is the existence of bias in the fixed initial-value case that constitutes a proof of *bias of the maximum-likelihood* criterion as well. Moreover, the case treated is one where the regression coefficient is identically equal to the corresponding structural coefficient. Hence, the *bias* of least-squares and maximum-likelihood methods exists in *structural* as well as *predictive* estimation.

Once the *existence* of the bias had been shown it was of interest to investigate its *magnitude* for samples of various sizes. The first three terms of a series expansion for the bias in a sample of arbitrary size were obtained; see equations (4.6) to (4.10) below.

Because of the equivalence of the structural and regression coefficients in the cases treated, the results described may be regarded as a first (and very modest) step in the small-sample theory of the maximum-likelihood estimates of the structural coefficients in (noncircular) stochastic difference-equation systems.

The results thus far obtained indicate the importance of the

---

[1] I.e., generated by a system of stochastic difference equations, some or all of which contain *lagged* values of *endogenous* variables; autoregressive time series are to be distinguished from those composed *additively* of a given function of time (often a polynomial or a Fourier series) and a stochastic ("error") term.

[2] They *are* valid in the *lagless* case, as shown in [VI].

[3] In this note we are only concerned with "stable" systems, i.e., those satisfying Assumption IV_2 in [Mann and Wald, p. 192]; for the case here treated this assumption is equivalent to postulating $|\alpha| < 1$ in (2.1) below.

bias. There is urgent need for more intensive study of the small-sample properties of estimates in autoregressive time series.

2.1. Let a sample be given consisting of T observations on a stochastic variable X_t. It will be assumed that the joint cumulative distribution function of $(X_1, \ldots, X_T)$, to be written as $F(X_1, \ldots, X_T)$, has the following two properties:

$$\mathcal{E}(X_t \mid X_{t-1}) = \alpha X_{t-1}, \qquad t = 2, 3, \ldots, T, \tag{2.1}$$

$$\sigma^2(X_t \mid X_{t-1}) = \text{const.} \tag{2.2}$$

Thus the regression of any observation on its predecessor is linear[1] and homoscedastic.

The least-squares estimate $\tilde{a}_T$ of α is in this case given by

$$\tilde{a}_T = \frac{\sum_{t=2}^{T} X_t X_{t-1}}{\sum_{t=2}^{T} X_{t-1}^2}. \tag{2.3}$$

2.2. An estimate a_T of α will be said to be (conditionally) *unbiased with regard to a family* $\mathfrak{F}$ *of cumulative distribution functions* F *if all* F *in* $\mathfrak{F}$ *satisfy* (2.1) *and* (2.2) *and if*

$$\mathcal{E}(a_T) \equiv \alpha \quad \text{for all } \alpha, \quad \text{for all } F \in \mathfrak{F}, \quad \text{and for all } T. \tag{2.4}$$

An estimate conditionally unbiased with regard to *all* cumulative distribution functions is said to be *absolutely unbiased*.[2]

Clearly, if a family $\mathfrak{F}_0$ can be found with regard to which a_T is not unbiased, a_T cannot be absolutely unbiased.

2.3. In this note we show that there exists a family[3] $\mathfrak{F}_0$, in

[1] The condition (2.1) is stronger than linearity, but for the purposes of this paper no loss of generality is involved.

[2] These definitions are special cases of those given in [VI].

[3] This family, defined by (2.1) and (2.2) with the additional assumption of normality, is divided into two "branches" depending on whether the

fact one of considerable practical importance, with regard to which the least-squares estimate $\tilde{a}_T$ of α is not unbiased. This shows that $\tilde{a}_T$ is not absolutely unbiased. It is not known whether there exists a family, say $\mathfrak{F}_1$, such that $\tilde{a}_T$ is unbiased with regard to $\mathfrak{F}_1$.

However, an unbiased estimate of α does exist. Thus in the sample $(X_0; X_1, \ldots, X_T)$ where X_0 is fixed and different from zero, the ratio X_1/X_0 is an unbiased, though in general very inefficient, estimate of α. For, assuming that

$$(2.5) \qquad X_t = \alpha X_{t-1} + u_t, \qquad t = 1, 2, \ldots, T,$$

where the u's have zero means, we have

$$(2.6) \qquad \mathcal{E}\left(\frac{X_1}{X_0}\right) = \mathcal{E}\left(\frac{\alpha X_0 + u_1}{X_0}\right) = \alpha + \frac{1}{X_0}\,\mathcal{E}(u_1) = \alpha.$$

It may be remarked that for $T = 2$, the ratio X_1/X_0 is a least-squares estimate. In fact, $T = 2$ is the only known case among finite-sized samples where the least-squares estimate $\tilde{a}_T$ is unbiased.

On the other hand, in the sample $(X_1, \ldots, X_T)$, where X_1 is stochastic and the likelihood function is given by (2.9), the expectation $\mathcal{E}(X_2/X_1)$ exists only in the Cauchy principal-value sense. In that sense, however, X_2/X_1 is an unbiased estimate of α, since it has the Cauchy distribution with a mode at zero. In fact, with the u's independent of each other, the mean

$$\frac{1}{T-1}\sum_{t=2}^{T}\frac{X_t}{X_{t-1}}$$

would also be unbiased in the Cauchy sense, although no more efficient than X_2/X_1. One might conjecture that the median of the ratios X_t/X_{t-1}, $t = 2, \ldots, T$, would be a more efficient estimate of α and perhaps an unbiased one.

2.4.1. In the following sections of the paper the proof of existence of the bias will be given and its magnitude evaluated for first-order stochastic difference equations with the initial value

initial value is assumed stochastic or fixed. The bias exists in both cases. The respective likelihood functions are given by (2.9) and (3.15).

stochastic and fixed.

For samples of three and four observations the size of relative bias[1] is given for all $|\alpha| < 1$ in Table 1 and Figure 1; these are based on the formulae (3.12), (3.19), and (3.37). The derivations and comments are to be found in section *3* of this paper. The case of fixed initial value (chosen as zero) is worked out for a sample of three observations only.

For larger samples, as well as the small ones, the limiting value of bias[1] as $|\alpha| \to 0$ is given in Table 2 and Figure 2; these are based on (4.4). The latter formula is valid for both the stochastic and fixed (zero) initial-value case.

Equations (4.7) to (4.10) give a more general result, viz., the first three terms of the Maclaurin expansion of the relative bias. The derivations are given in section *4*. Sections *4.2* and *4.3* contain some conjectures with regard to the nature of the approximation provided by the expansion and with regard to the nature of relative biås in stochastic difference equations in general, as well as some suggestions for further research.

2.4.2. Define a stochastic process by

$$(2.7) \qquad X_t = \alpha X_{t-1} + u_t, \qquad t = 2, 3, \ldots, T,$$

where the u's are independently normally distributed with zero means and unit variances, and where $|\alpha| < 1$. Hence, if the process is stationary[2],

$$(2.8) \qquad \mathcal{E}(X_t) = 0, \qquad \mathcal{E}(X_t^2) = \frac{1}{1-\alpha^2}, \qquad t = 1, 2, \ldots, T.$$

Given a sample of size T, its likelihood function will be

$$(2.9) \quad (1-\alpha^2)^{1/2}(2\pi)^{-T/2}\exp\{-\frac{1}{2}[(1-\alpha^2)X_1^2 + \sum_{t=2}^{T}(X_t - \alpha X_{t-1})^2]\},$$

which is equivalent to equation (9) in [Koopmans, 1942]. It can be

[1]The formulae, tables, and figures give the values of $N_T \equiv \mathcal{E}(\tilde{a}_T)/\alpha$, i.e., the relative bias plus 1.

[2]This specifies the stochastic initial-value case. For fixed initial-value case see below, section *3.1.2*.

seen that

$$\mathcal{E}(X_t \mid X_{t-1}) = \alpha X_{t-1}, \qquad t = 2, 3, \ldots, T, \tag{2.10}$$

so that (2.1) is satisfied. (2.2) is also satisfied with a unit variance.

3.1.1. To show the existence of bias it will suffice to find a T for which (2.4) does not hold. The following proof demonstrates that (2.4) does not hold for $T = 3$.

The proof consists in finding the expectation of $\tilde{a}_T$ for $T = 3$, where

$$\tilde{a}_3 = \frac{X_1 X_2 + X_2 X_3}{X_1^2 + X_2^2}. \tag{3.1}$$

Following the procedure used in [Williams] and in [Dixon], we write

$$\tilde{a}_3 = \frac{A_3}{B_3}, \qquad A_3 = X_1 X_2 + X_2 X_3, \qquad B_3 = X_1^2 + X_2^2, \tag{3.2}$$

then find the characteristic function $\varphi_3(t_1, t_2)$ of A_3 and B_3, and, finally, obtain the expectation of $\tilde{a}_3$ from

$$\mathcal{E}(\tilde{a}_3) = \int_{-\infty}^{0} \left[\frac{\partial \varphi_3(t_1, t_2)}{\partial t_1} \right]_{t_1 = 0} dt_2 . \tag{3.3}$$

We have

$$(1 - \alpha^2)\, \varphi_3^{-2}(t_1, t_2) = \begin{vmatrix} y & b & 0 \\ b & z & b \\ 0 & b & 1 \end{vmatrix} = \begin{vmatrix} y & f & 0 \\ 1 & z & f \\ 0 & 1 & 1 \end{vmatrix}, \tag{3.4}$$

where

$$\begin{aligned} y &= -2t_2 + 1, \\ f &= b^2 = (t_1 + \alpha)^2, \\ z &= y + \alpha^2 = -2t_2 + 1 + \alpha^2. \end{aligned} \tag{3.5}$$

Denoting the determinant in (3.4) by $C^{(3)}$, we have

$$\mathcal{E}(\tilde{a}_3) = -\frac{\sqrt{1-\alpha^2}}{2} \int_{-\infty}^{0} \frac{1}{C_0^{(3)}\sqrt{C_0^{(3)}}} \left.\frac{\partial C^{(3)}}{\partial t_1}\right|_{t_1=0} dt_2 , \tag{3.6}$$

where

$$C_0^{(3)} \equiv C^{(3)}\Big|_{t_1=0} = 4\left(t_2^2 - t_2 + \frac{1-\alpha^2}{4}\right) \tag{3.7}$$

and

$$\left.\frac{\partial C^{(3)}}{\partial t_1}\right|_{t_1=0} = 4\alpha(t_2 - 1). \tag{3.8}$$

The integral in (3.6) may now be evaluated [Pierce, formula 200] and we obtain

$$\mathcal{E}(\tilde{a}_3) = \frac{\alpha}{2}\left(1 + \frac{1-\sqrt{1-\alpha^2}}{\alpha^2}\right). \tag{3.9}$$

It can be seen that

$$\begin{aligned} \mathcal{E}(\tilde{a}_3) &\to \frac{3}{4}\alpha \quad \text{for } |\alpha| \to 0, \\ \mathcal{E}(\tilde{a}_3) &\to \alpha \quad \text{for } |\alpha| \to 1. \end{aligned} \tag{3.10}$$

Thus $\tilde{a}_3$ is a biased estimate of α.

Writing $\beta = \alpha^2$ and

$$N_T \equiv N_T(\beta) = \frac{\mathcal{E}(\tilde{a}_T)}{\alpha} , \tag{3.11}$$

we may state the above results as

$$N_3(\beta) = \frac{1}{2}\left(1 + \frac{1-\sqrt{1-\beta}}{\beta}\right) \tag{3.12}$$

with $N_3(\beta)$ varying from[1] $N_3(0) = 0.75$ to $N_3(1) = 1$. The values of

[1] $N_T(0) \equiv \lim_{\beta\to 0} N_T(\beta)$; $N_T(1) \equiv \lim_{\beta\to 1} N_T(\beta)$.

$N_3(\beta)$ are shown below in Table 1 and plotted in Figure 1.

The convergence of N_3 to 1 as $|\alpha|$ tends to 1 is very slow. The relative bias is still 12½ per cent (i.e., $N_3 = 0.875$) for $|\alpha| = 0.94$, so that it takes 94 per cent of the range of $|\alpha|$ to remove one-half of the relative bias!

3.1.2.[1] By a similar procedure we can evaluate the bias for the case where the initial value is a fixed variate, here chosen as zero. Let

$$X_0 = \text{fixed}, \tag{3.13}$$

$$X_t = \alpha X_{t-1} + u_t, \quad t = 1, 2, \ldots, T, \tag{3.14}$$

where α and the u's have the same properties as before. T now denotes the number of *stochastic* observations. The likelihood function becomes

$$(2\pi)^{-T/2} \exp\{-\frac{1}{2} \sum_{t=1}^{T} (X_t - \alpha X_{t-1})^2\}. \tag{3.15}$$

Then the *least-squares*[2] estimate $\tilde{a}_T^*$ of α, which in this case [unlike that specified by (2.9)] is also a *maximum-likelihood* estimate, is given by

$$\tilde{a}_T^* = \frac{\sum_{t=1}^{T} X_t X_{t-1}}{\sum_{t=1}^{T} X_{t-1}^2}. \tag{3.16}$$

Now consider the special case where

$$X_0 = 0. \tag{3.17}$$

Here we find that $\tilde{a}_T^*$ as defined in (3.16) equals $\tilde{a}_T$ as defined in (2.3).

---

[1] T. W. Anderson has made helpful suggestions in connection with the problem treated in this section.

[2] The asterisk in $\tilde{a}_T^*$ indicates that $\tilde{a}_T^*$ is the least-squares estimate of α for the sample $(X_0, X_1, \ldots, X_T)$ while $\tilde{a}_T$ is the least-squares estimate for the sample $(X_1, \ldots, X_T)$.

Denote by $\mathcal{E}^*(\tilde{a}_T)$ the expectation of $\tilde{a}_T$ evaluated on the basis of (3.15) with $X_0 = 0$; also, write

$$N_T^* \equiv N_T^*(\beta) \equiv \frac{\mathcal{E}^*(\tilde{a}_T)}{\alpha} . \tag{3.18}$$

Then we find that

$$N_3^*(\beta) = \frac{3+\beta}{4+\beta} . \tag{3.19}$$

Thus $N_3^*(0) = 0.75 = N_3(0)$;[1] but while $N_3(1) = 1$, we have here $N_3^*(1) = 4/5$: there is a bias even at $|\alpha| = 1$!

The case where $X_0 \neq 0$ has not been treated, but one might conjecture that, for given T and β, $N_T^*(\beta) \to 1$ as X_0 becomes numerically large.

3.2.1. When $T = 4, 5$ the integrals to be evaluated are of the elliptic type. For $T \geq 6$ they are hyperelliptic. Only for the case of $T = 4$ has the bias been evaluated in closed form.

To perform the integration in the elliptic case it is necessary to factor the T-rowed determinant $C_0^{(T)}$ where[2] [cf. (3.4) above]

$$C^{(T)} = (1-\beta)\,\varphi^{-2}(t_1, t_2) = \begin{vmatrix} y & f & 0 & \cdots & 0 & 0 & 0 \\ 1 & z & f & \cdots & 0 & 0 & 0 \\ 0 & 1 & z & \cdots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & 1 & z & f \\ 0 & 0 & 0 & \cdots & 0 & 1 & 1 \end{vmatrix}_{(T)} \tag{3.20}$$

in a sample of T observations and

$$C_0^{(T)} = C^{(T)}\Big|_{t_1=0} . \tag{3.21}$$

We may write

---

[1] This is an example of a more general phenomenon: $N_T^*(0) = N_T(0)$ for all T, cf. (4.4).

[2] For the definitions of y, z, f, see (3.5).

$$C_0^{(T)} = \begin{vmatrix} z-\beta & \beta & 0 & \dots & 0 & 0 & 0 \\ 1 & z & \beta & \dots & 0 & 0 & 0 \\ 0 & 1 & z & \dots & 0 & 0 & 0 \\ . & . & . & \dots & . & . & . \\ 0 & 0 & 0 & \dots & 1 & z & \beta \\ 0 & 0 & 0 & \dots & 0 & 1 & 1 \end{vmatrix}_{(T)} . \tag{3.22}$$

For $\mu = 2, 3, \dots$ we have

$$\begin{aligned} C_0^{(2\mu+1)} &= R_\mu^2 - \beta R_{\mu-1}^2 = (R_\mu - \alpha R_{\mu-1})(R_\mu + \alpha R_{\mu-1}) , \\ C_0^{(2\mu)} &= R_{\mu-1}(z R_{\mu-1} - 2\beta R_{\mu-2}) , \end{aligned} \tag{3.23}$$

where

$$R_\mu = \begin{vmatrix} z & \beta & 0 & \dots & 0 & 0 & 0 \\ 1 & z & \beta & \dots & 0 & 0 & 0 \\ 0 & 1 & z & \dots & 0 & 0 & 0 \\ . & . & . & \dots & . & . & . \\ 0 & 0 & 0 & \dots\ 1 & 1 & z & \beta \\ 0 & 0 & 0 & \dots & 0 & 1 & 1 \end{vmatrix}_{(\mu+1)}$$

(3.24)

$$= \begin{vmatrix} z-\beta & \beta & 0 & \dots & 0 & 0 & 0 \\ 1 & z & \beta & \dots & 0 & 0 & 0 \\ 0 & 1 & z & \dots & 0 & 0 & 0 \\ . & . & . & \dots & . & . & . \\ 0 & 0 & 0 & \dots & 1 & z & \beta \\ 0 & 0 & 0 & \dots & 0 & 1 & z \end{vmatrix}_{(\mu)}$$

with

$$R_0 = 1, \qquad R_1 = \begin{vmatrix} z & \beta \\ 1 & 1 \end{vmatrix}. \tag{3.25}$$

Equations (3.23) make it possible to find by radicals the roots of the determinant $C_0^{(T)}$ for $T \leq 9$. For example,

$$\begin{aligned} C_0^{(4)} &= (z-\beta)[z(z-\beta) - 2\beta], \\ C_0^{(5)} &= \{[z(z-\beta) - \beta - \alpha(z-\beta)]\}\{[z(z-\beta) - \beta] + \alpha(z-\beta)\}, \\ C_0^{(6)} &= [z(z-\beta) - \beta]\{z[z(z-\beta) - \beta] - 2\beta(z-\beta)\}, \end{aligned} \tag{3.26}$$

etc.

3.2.2. Now in order to obtain $N_4(\beta)$ we follow a procedure similar to that for $T = 3$. We have

$$\mathcal{E}(\tilde{a}_4) = -\frac{1}{2}\sqrt{1-\beta}\int_{-\infty}^{0}[C_0^{(4)}]^{-3/2}\left.\frac{\partial C^{(4)}}{\partial t_1}\right|_{t_1=0} dt_2, \tag{3.27}$$

where[1]

$$C^{(4)} = \begin{vmatrix} y & f & 0 & 0 \\ 1 & z & f & 0 \\ 0 & 1 & z & f \\ 0 & 0 & 1 & 1 \end{vmatrix}, \tag{3.28}$$

so that

$$C_0^{(4)} = y(y^2 + \beta y - 2\beta), \tag{3.29}$$

$$\left.\frac{\partial C^{(4)}}{\partial t_1}\right|_0 = -2\alpha[(y^2 + \beta y - 2\beta) + (2y + \beta)]. \tag{3.30}$$

Hence, substituting these values into (3.27) and dividing both

[1]The symbols used here are defined in (3.5).

sides by α we have

$$(3.31)\quad N_4(\beta) = \sqrt{1-\beta}\int_{-\infty}^{0}\frac{(y^2+\beta y-2\beta)+(2y+\beta)}{[y(y^2+\beta y-2\beta)]^{3/2}}\,dt_2 .$$

Splitting into partial fractions and making the substitutions

$$(3.32)\quad y = -2t_2 + 1 ,$$

$$(3.33)\quad w^2 = \frac{y-y_1}{y}, \qquad y_1 = \frac{1}{2}(\sqrt{\beta^2+8\beta}-\beta) ,$$

and, finally,

$$(3.34)\quad u = \mathrm{sn}^{-1} w ,$$

we obtain

$$(3.35)\quad \frac{4\sqrt{\beta^2+8\beta}}{\sqrt{1-\beta}}N_4(\beta) = \frac{1}{y_1^2}\int \mathrm{cs}^2 u\,du + \frac{1}{2y_1}\int \mathrm{cn}^2 u\,du + \frac{1}{y_3(y_1-y_3)}\int \mathrm{cd}^2 u\,du ,$$

$$(3.36)\quad y_1,\ y_3 = \frac{1}{2}(\pm\sqrt{\beta^2+8\beta}-\beta) ,$$

where Glaisher's notation is used ($\mathrm{cs}\,u \equiv \mathrm{cn}\,u/\mathrm{sn}\,u$, $\mathrm{cd}\,u \equiv \mathrm{cn}\,u/\mathrm{dn}\,u$); the upper limit of integration is $\mathrm{sn}^{-1}(1)$ and the lower limit $\mathrm{sn}^{-1}[(1-y_1)^{1/2}]$.

With the help of formulae given by [Whittaker and Watson, 22.72, Ex. 3], we perform the integration, thus obtaining, after simplification,

$$(3.37)\quad N_4(\beta) = \frac{1}{2\beta}\left[(1+\beta) - \frac{\sqrt{1-\beta}}{4\sqrt{\beta^2+8\beta}}\,\Delta F\right] ,$$

where

(3.38) $$\Delta F = F_{90°}(\theta) - F_{\varphi}(\theta).$$

$F_{\varphi}(\theta)$ is the incomplete elliptic integral of the first kind with modular angle θ and amplitude φ; $F_{90°}(\theta)$ is the corresponding complete integral. The angles θ and φ are given by

(3.39) $$\sin\varphi = \sqrt{1-y_1}\,, \qquad \sin\theta = k\,, \quad 2k^2 = 1 + \frac{\beta}{\sqrt{\beta^2+8\beta}}\,.$$

3.2.3. It is easily seen that

(3.40) $$N_4(1) \equiv \lim_{\beta\to 1} N_4(\beta) = 1\,.$$

From this and the behavior of $N_3(1)$, one might conjecture that $N_T(1) = 1$ for all T.

It will be shown later in (4.4) that $N_4(0) \equiv \lim_{\beta\to 0} N_4(\beta) = 11/15 < N_3(0)$. As can be seen in Table 1 and Figure 1, $N_4(\beta)$ is a monotonic function of β. Whether this is the property of $N_T(\beta)$ for all T is not known.

3.2.4. The numerical values[1] computed from (3.37) are given in Table 1 and plotted in Figure 1. It is of interest to note that $N_3(\beta) > N_4(\beta)$ for all β. It will be seen later [from (4.4)] that $N_T(0)$ has a minimum for $T = 4$ (if T is an integer).[2] The convergence of $N_4(\beta)$ to 1 as $\beta \to 1$ is very slow, as was also the case for $N_3(\beta)$. To reduce the relative bias to one-half of its maximal value (so that $N_4 = 0.866$) we must have $|\alpha| = 0.95$. Whether the situation is quite as serious for larger samples is not known.

4.1. By expanding the integrand of (3.27), with T replacing

[1]Obtained with the aid of Miss Estelle Mass.

[2]For $T \geq 2$ and real, but not necessarily an integer, the minimum is at $T = 2 + \sqrt{3}$; $N_{2+\sqrt{3}}(0) = \sqrt{3} - 1 = 0.7321$.

TABLE 1.

Ratio of Estimate Expectation to True Parameter Value for Samples of 3 and 4 Observations

α	$\beta\equiv\alpha^2$	$N_3 \equiv \mathcal{E}(\tilde{a}_3)/\alpha$	$N_3^* \equiv \mathcal{E}^*(\tilde{a}_3)/\alpha$	$N_4 \equiv \mathcal{E}(\tilde{a}_4)/\alpha$
0	0	0.7500	0.7500	0.7333
0.1	0.01	0.7506	0.7506	0.7340
0.2	0.04	0.7526	0.7525	0.7359
0.3	0.09	0.7559	0.7555	0.7388
0.4	0.16	0.7609	0.7596	0.7434
0.5	0.25	0.7679	0.7647	0.7501
0.6	0.36	0.7778	0.7706	0.7595
0.7	0.49	0.7918	0.7773	0.7730
0.8	0.64	0.8125	0.7845	0.7938
0.9	0.81	0.8482	0.7921	0.8307
0.95	0.9025	0.8810	0.7960	0.8656
0.99	0.9801	0.9382	0.7992	0.9289
1.00	1.0000	1.0000	0.8000	1.0000

N and N^* refer to stochastic and fixed (zero) initial-value cases, respectively. The subscript indicates the number of (stochastic) observations in the sample.

TABLE 2.

The Limit (for Small Parameter Values) of the Ratio of Estimate Expectation to True Parameter Value: Stochastic Initial-Value Case

Sample Size T	$N_T(0) \equiv \lim_{\alpha\to 0} \mathcal{E}(\tilde{a}_T)/\alpha$	Sample Size T	$N_T(0) \equiv \lim_{\lvert\alpha\rvert\to 0} \mathcal{E}(\tilde{a}_T)/\alpha$
22	1.0000	13	0.8690
3	0.7500	14	0.8769
4	0.7333	15	0.8839
5	0.7500	16	0.8902
6	0.7714	17	0.8958
7	0.7917	18	0.9009
8	0.8095	19	0.9056
9	0.8250	20	0.9098
10	0.8384	50	0.9616
11	0.8500	100	0.9804
12	0.8601	500	0.9960

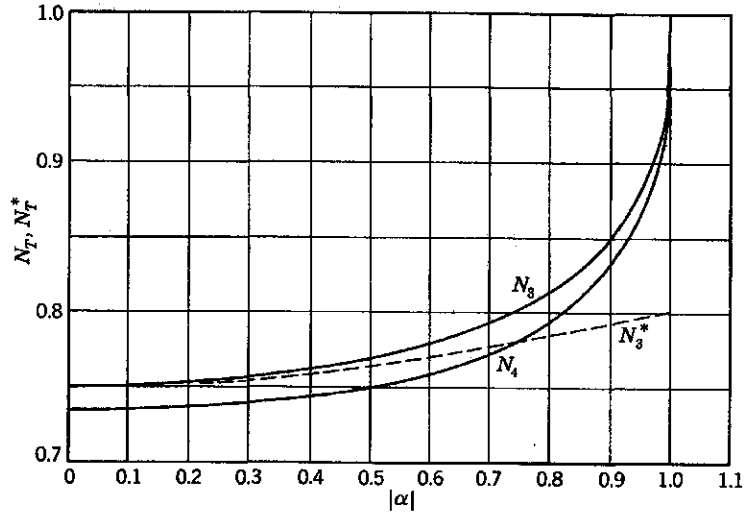

FIGURE 1. *Ratio of Estimate Expectation to True Parameter Value for Samples of 3 and 4 Observations. (For explanation of symbols see Table 1.)*

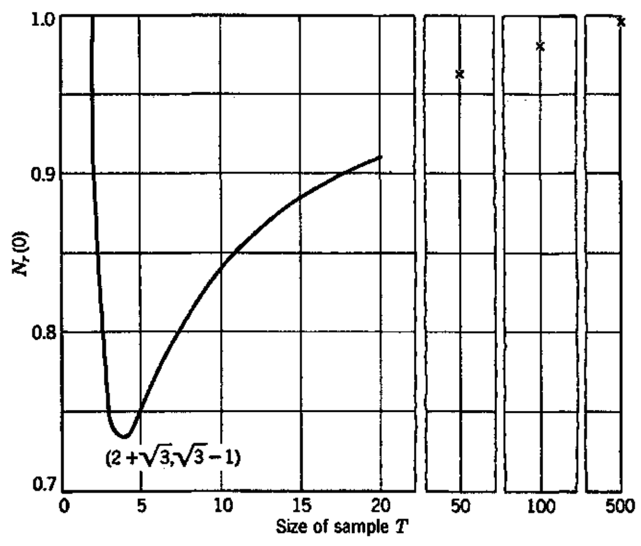

FIGURE 2. *The Limit (for Small Parameter Values) of the Ratio of Estimate Expectation to True Parameter Value: Stochastic Initial-Value Case. (For explanation of symbols see Table 2.)*

the affix 4, in a Maclaurin series and integrating termwise, it is possible to obtain an expansion[1] for $N_T(\beta)$ in powers of β.

It will now be shown how the first term of the expansion is obtained.[2] We first find

$$\lim_{\beta\to 0} C_0^{(T)} = y^{T-1}, \tag{4.1}$$

$$\lim_{\beta\to 0} \frac{1}{\alpha}\frac{\partial C^{(T)}}{\partial t_1}\bigg|_{t_1=0} = -2y^{T-3}\left[y + (T-2)\right]_0 . \tag{4.2}$$

Hence

$$N_T(0) \equiv \lim_{\beta\to 0} \frac{\mathcal{E}(\tilde{a}_T)}{\alpha} = \int_{-\infty}^{0} \frac{y^{T-3}[y + (T-2)]}{(y^{T-1})^{3/2}}\, dt_2 , \tag{4.3}$$

and, upon evaluation, this yields[3]

$$N_T(0) = \frac{T^2 - 2T + 3}{(T-1)(T+1)} . \tag{4.4}$$

Hence the bias exists for all finite-sized samples except $T = 2$. The values of $N_T(0)$ for some T are given in Table 2 and plotted in Figure 2.

It will be noted that the relative bias is 9 per cent for a sample of 20 observations and 2 per cent for a sample of 100 observations.

The second and third terms of the expansion of $N_T(\beta)$ in powers of β are obtained by similar methods, although the procedure becomes quite laborious. Writing

---

[1] As before, $N_T(\beta) \equiv \mathcal{E}(\tilde{a}_T)/\alpha$ and $\tilde{a}_T = \sum_{t=2}^{T} X_t X_{t-1} \Big/ \sum_{t=2}^{T} X_{t-1}^2$ is the least-squares estimate of α.

[2] Valuable suggestions in connection with this problem were made by Professor John von Neumann, Institute for Advanced Study.

[3] The same formula holds for $N_T^*(0)$, i.e., for the fixed initial-value case with $X_0 = 0$; see section *3.1.2* above.

$$(4.5) \qquad C_{ij}^{(T)} = \left[\frac{\partial^j}{\partial \beta^j}\left(\frac{\partial^i C^{(T)}}{\partial f^i}\right)_{f=\beta}\right]_{\beta=0} ,$$

we have

$$(4.6) \qquad \begin{aligned} C_{00}^{(T)} &= y^{T-1} , \\ C_{01}^{(T)} &= y^{T-3}[(T-3)y - (T-2)] , \\ C_{02}^{(T)} &= (T-3)(T-4)\, y^{T-5}(y^2 - 2y + 1) , \\ C_{10}^{(T)} &= -\, y^{T-3}[y + (T-2)] , \\ C_{11}^{(T)} &= -\,(T-3)\, y^{T-5}[y^2 + (T-5)y - (T-4)] , \\ C_{12}^{(T)} &= -\,(T-4)\, y^{T-7}[(T-3)y^3 + (T-4)(T-7)y^2 \\ &\qquad - (T-5)(2T-11)y + (T-5)(T-6)] , \end{aligned}$$

where the expression for $C_{12}^{(T)}$ is valid only for $T > 3$. With the help of (4.6) we find that, for $T > 3$, the first three coefficients of the expansion

$$(4.7) \qquad N_T(\beta) = N_T(0) + N_T'(0)\,\beta + \frac{1}{2} N_T''(0)\,\beta^2 + \cdots$$

are given by

$$(4.8) \qquad N_T(0) = \frac{T^2 - 2T + 3}{(T-1)(T+1)} ,$$

$$(4.9) \qquad N_T'(0) = \frac{2(T^2 - 8T + 21)}{(T-1)(T+1)(T+3)(T+5)} ,$$

$$(4.10) \qquad N_T''(0) = \frac{4(T^4 + 24T^3 + 98T^2 - 264T - 99)}{(T-1)(T+1)(T+3)(T+5)(T+7)(T+9)} .$$

Equations (4.8) and (4.9) are valid for $T \geq 2$ and $T \geq 3$, respectively, equation (4.10) for $T > 3$ only.

4.2. It is quite evident that the first three terms of the expansion for $N_T(\beta)$ do not give a very good approximation for values of β near 1. For example, $N_4(1) = 1$, while the first three terms of the expansion give only 0.823. For $|\alpha| = 0.5$, however, while[1] $N_4(0.25) = 0.7501$, the first three terms of the expansion give 0.7497 which is correct to the third decimal digit. If the same phenomenon should exist for $T > 4$, which remains to be proved, the first three terms of the expansion could not only be used successfully for values of $|\alpha|$ below 0.5 but also for values of $|\alpha|$ somewhat (though not too much) above 0.5.

Moreover, it would seem that these three expansion terms will always give values lower than the true $N_T(\beta)$. It would be desirable to examine the correctness of this conjecture and also to obtain an upper bound for $N_T(\beta)$ in terms of the expansion. It seems possible that such a bound is given by the expression

$$(4.11) \qquad 1 - (1-\beta)N_T'(0) - (1-\beta^2)\frac{1}{2}N_T''(0) \ .$$

However, even if proved correct, this would only be useful for values of β in the neighborhood of 1. For example, for $T = 4$ and $|\alpha| = 0.9$, the upper bound would be given by 0.978, while the correct value is only 0.831. It may be observed that the first three terms of the expansion give a much better approximation, namely 0.801, despite the high value of $|\alpha|$.

4.3. The following more general propositions concerning first-order stochastic linear difference equations still await proof:

$$(4.12) \qquad N_T(\beta) \leq 1 \quad \text{for } \beta < 1 \text{ and all } T,$$

$$(4.13) \qquad \frac{\partial N_T(\beta)}{\partial \beta} > 0 \quad \text{for } \beta < 1 \text{ and all } T,$$

and hence

$$(4.14) \qquad \lim_{\beta \to 1} N_T(\beta) = 1 \quad \text{for all } T.$$

[1] It will be remembered that the argument of $N_T(\ )$ is β, not $|\alpha|$.

It would also be desirable to investigate analogous problems first for higher-order difference equations and then for equation systems.[1] It would be of interest to see whether the following two propositions are generally true: (1) that a system becomes more strongly damped when estimate expectations are substituted for the structural parameters, and (2) that the relative bias is lower in systems with stronger damping and tends to zero as the characteristic roots of the system approach 1 in absolute value.

These investigations should be carried out for both the stochastic and fixed-variate initial-value case, without the fixed initial value being necessarily zero.

Finally, higher moments of the sampling distributions of the estimates should be investigated. It is probable that these problems will have to be studied with the help of more powerful tools, especially that of the approximate sampling distributions,[2] provided the upper bound of the approximation error can be determined.

---

[1] At this point one would have to distinguish between the bias of the (maximum-likelihood) estimates of the structural coefficients and the bias of the (least-squares equivalent to maximum-likelihood for the fixed-variate initial-value case) estimates of the regression coefficients. The distributions of the conditional variances and of the disturbance covariance matrix also deserve attention in small samples.

[2] See [Koopmans, 1942], [Dixon], and [Leipnik].