Some Comments on Sampling Experiments by Orcutt and Cochrane

John Gurland

April 18, 1950

1. Introduction.

These comments refer mainly to a paper by Guy H. Orcutt and Donald Cochrane [1] which appears in the September 1949 issue of the Journal of the American Statistical Association. By considering a simple system of two linear stochastic equations, the authors investigate the following phenomena manifest in samples of size 20.

(i) The bias in maximum likelihood, limited information, and least-squares estimates of the structural parameters

(ii) The loss of efficiency when limited information estimates are used in place of maximum likelihood estimates

(iii) How autocorrelation of disturbances causes bias and loss of efficiency in the foregoing estimates of the structural parameters.

This paper by Orcutt and Cochrane [1] (which will be referred to as O.C.) is an extension of a paper by Cochrane and Orcutt [2] (which will be referred to as C.O.) As the C.O. paper has close connections with the O.C. paper, and as its results suggest wide implications, it seems advisable to allot some discussion to it.

The results of these two papers should sound a warning as to how serious are the unsolved problems of small sample bias and inefficiency of estimates which are known to be asymptotically optimal; also the problem of how to cope with the autocorrelation of the disturbances.

2. The C.O. Paper.

(a) Notation

Let

(1) $$y_t = a_o + \sum_{i=1}^{k} a_i x_{it} + u_t \qquad (t = 1, 2, \ldots, n)$$

represent a linear relation in which y is the endogenous variable, $x_1, x_2, \ldots, x_k$ are the exogenous variables, u is the random disturbance, such that

(2) $$x_{i\tau} \text{ stochastically independent of } u_t \quad \text{for } t, \tau = 1, 2, \ldots, n \quad i = 1, 2, \ldots, k$$

Suppose u_t is generated by the Markoff scheme

(3) $$u_t - \beta_1 u_{t-1} - \beta_2 u_{t-2} \cdots - \beta_m u_{t-m} = \varepsilon_t$$

where

(3') $$E\, \varepsilon_t \, \varepsilon_\tau = 0 \qquad \text{for } t \neq$$

and

$$E\, \varepsilon_t^2 = \sigma^2 \qquad \text{independent of } t.$$

The β's are referred to as autoregressive coefficients of u_t.

(b) Least-squares estimation of $a_o, a_1, \ldots, a_k$.

If $\beta_1 = \beta_2 = \ldots = \beta_m = 0$, then the least-squares estimates of the a's will be best linear unbiassed [3]. Otherwise they will not, in general, have this optimal property. The C.O. sampling study indicates how badly biassed and/or inefficient these estimates may be when $\beta_1 \neq 0$ and $\beta_2 = \beta_3 = \ldots = \beta_m = 0$.

Consider the least-squares estimates of $a_o', a_1, a_2, \ldots, a_k$ in the linear relation

(4) $$y_t' = a_o' + \Sigma a_i x_{it}' + \varepsilon_t$$

where

(5) $$\begin{aligned} y_t' &= y_t - \beta_1 y_{t-1} \cdots &- \beta_m y_{t-m} \\ x_{it}' &= x_{it} - \beta_1 x_{i,t-1} \cdots &- \beta_m x_{i,t-m} \\ a_o' &= a_o (1 - \beta_1 - \beta_2 \cdots &- \beta_m) \end{aligned}$$

These estimates, which are linear combinations of y_1', ..., y_n' are, in fact, best linear unbiassed with respect to the whole class of linear combinations of $y_1, y_2, \ldots, y_n$. This optimal property is remarkable because the variance of these estimates, which is minimum with respect to the class of unbiassed linear combinations of $y_1', y_2', \ldots, y_n'$, is also minimum with respect to the larger class of unbiassed linear combination of $y_1, y_2, \ldots, y_n$.

The property of best linear unbiassed with respect to the class of linear combinations of $y_1, y_2, \ldots, y_n$ is proved for the case $m = k = 1$ in the Appendix of the C.O. paper. The general case is considered in a paper by Champernowne [4].

(c) Von Neumann's statistic $\frac{\delta^2}{s^2}$ used for testing autocorrelation.

Let $z_1, z_2, \ldots, z_n$ be n random variables with the probability density function

(6) $$\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(z_i - \rho z_{i-1})^2}$$

where

$$z_o = 0$$

Let

(7) $$\delta^2 = \frac{1}{n-1}\sum_{1}^{n-1}(z_{i+1} - z_i)^2$$

(8) $$s^2 = \frac{1}{n}\sum_{1}^{n}(z_i - \bar{z})^2$$

where

$$\bar{z} = \frac{1}{n}\sum_{1}^{n} z_i$$

Then $\frac{\delta^2}{s^2}$ may be used as a statistic to test the hypothesis $\rho = 0$. Its distribution, on the assumption that $\rho = 0$, is given by von Neumann [5]. Other statistics could also be used to test the same hypothesis. For instance, the statistic

(9)
$$\frac{z_1 z_2 + z_2 z_3 + \dots + z_{n-1} z_n}{z_1^2 + z_2^2 + \dots + z_n^2}$$

which Koopmans [6] employs in the case of a stationary stochastic process ($|\rho| < 1$), where the z's have the probability density

(10)
$$\frac{\sqrt{1-\rho^2}}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}\left[z_1^2 + z_n^2 - 2\rho(z_1 z_2 + \dots + z_{n-1} z_n) + (1+\rho^2)(z_2^2 + \dots + z_{n-1}^2)\right]}$$

might also be used here to test the hypothesis $\rho = 0$.

In the C.O. paper the von Neumann statistic is used to test the disturbances for "randomness." (The authors use the term "randomness" in the sense of serial independence.) They point out, however, that the applicability of this statistic is questionable in the present instance, since the disturbances are not observable and the computed residuals are used instead.

Thus, let $\hat{a}_0, \hat{a}_1, \dots, \hat{a}_k$ be the least-squares estimates of $a_0, a_1, \dots, a_k$ respectively.

Let

(11)
$$r_t = y_t - \hat{a}_0 - \sum_{i=1}^{k} \hat{a}_i x_{it}$$

(12)
$$\delta^2 = \frac{1}{n-1} \cdot \sum_{1}^{n-1} (r_{t+1} - r_t)^2$$

(13)
$$s^2 = \frac{1}{n} \cdot \Sigma(r_t - \bar{r})^2$$

where $\bar{r} = \frac{1}{n} \sum_{1}^{n} r_t$

As the authors emphasize, a test based on the statistic $\frac{\underline{\delta}^2}{\underline{s}^2}$, using the tables of the distribution of $\frac{\delta^2}{s^2}$ given by Hart and von Neumann [7] may frequently fail to detect non-zero values of β_1 in (3), (where $\beta_2, \beta_3, \ldots, \beta_m$ are taken to be zero.); and the test appears to be the worse, the larger the values of k in (1). They base their conclusions on empirical evidence, some of which is given in Table I, on page 39 of the C.O. paper. The table summarizes the values of $\frac{\underline{\delta}^2}{\underline{s}^2}$, used in testing the hypothesis $\rho = 0$, on the assumption the disturbances have probability density (6) or (10).

The distribution of $\frac{\delta^2}{s^2}$, when $\rho = 0$, is symmetrical around $\frac{2n}{n-1}$ (= 2.11 for n = 20). The aforementioned table indicates how the hypothesis $\rho = 0$ is rejected less frequently as the number of parameters increase. The 2½% and 5% critical values of $\frac{\delta^2}{s^2}$, on the lower tail of the distribution are indicated in the last two rows of the table, as well as the number of equations for which the hypothesis $\rho = 0$ is rejected. If one is willing to accept the authors' opinion that the disturbances of the economic series in question are all positively autocorrelated, then the evidence of the above table could be considered quite suggestive of a "bias toward randomness" in using the statistic $\frac{\underline{\delta}^2}{\underline{s}^2}$ along with the tables of the distribution of $\frac{\delta^2}{s^2}$. Further evidence which is more convincing will be mentioned presently.

(d) Description of the sampling experiment.

Let ε be a random variable which assumes each of the discrete values -49, -48, ..., -1, 0, 1, 2, ..., 48, 49 with equal probability $\frac{1}{99}$. By referring to tables of random numbers [8], 60 series of ε_t were constructed, each one 20 items in length

$(t = 1, 2, \ldots, 20)$. This type of series is referred to as

(14) D: $x_{t+1} = \varepsilon_{t+1}$ $(t = 0, 1, \ldots, 19)$

Using the 60 series of type D, it is now possible to construct the following types of series, each type comprising 60 series.

(15) A: $x_{t+1} = x_t + 0.3(x_t - x_{t-1}) + \varepsilon_{t+1}$ $x_0 = x_{-1} = 0$

(16) B: $x_{t+1} = x_t + \varepsilon_{t+1}$ $x_0 = 0$

(17) C: $x_{t+1} = 0.3\, x_t + \varepsilon_{t+1}$ $x_0 = 0$

(18) E: $x_{t+1} = \varepsilon_{t+1} - \varepsilon_t$ 19 terms

All the above series were used to investigate linear regression with autocorrelated disturbances. Consider the general form of relationship

(19) $$X_1 = k + b_{12.3t}\, X_2 + b_{13.2t}\, X_3 + b_{1t.23}\, t + u$$

where X_2, X_3, u are independently constructed series, all possessing the same autoregressive structure A, B, C, D or E, where t represents a linear trend and the true parameter values are

(20) $$k = 0; \quad b_{12.3t} = 2; \quad b_{13.2t} = 1; \quad b_{1t.23} = 0$$

The results given in Table II, C.O. page 46, can best be described by a typical example. Apply the 60 series of type B, three at a time, to the equation

(21) $$X_1 = 2X_2 + X_3 + u$$

where X_2, X_3, u are given the values assumed by the constructed series. (Note that the same type, B, is used for each of the three variables). Inserting the values X_1, X_2, X_3 so obtained, into (19), least-squares estimates of the parameters in (19) are then obtained. This procedure is carried out 20 times until the 60 series of type B are used up. The results for this case are given in line 6 (equation no. 6) of Table II. The main headings of the table are self-explanatory. The heading "correlation coefficient" signifies the usual multiple correlation coefficient.

In other parts of Table II blank spaces occur under certain parameters. For instance, in line 1 there are blanks under b_{13} and b_{1t}. This means that

(22) $$X_1 = 2X_2 + u$$

is the true relationship, and the least-squares estimates of the parameters in

(23) $$X_1 = k + b_{12.3t} X_2 + u$$

are obtained. In this case, the series for X_1 and X_2 are obtained from (22) by using constructed series of type A for X_2 and u.

There are only three cases in which the trend term t is fitted by least-squares, namely equations 2, 4, 6 of the table. Of course, the true value of b_{1t} is 0, (see (20)), so the trend term is a bogus variable, when it is fitted by least-squares. It is interesting to see how the introduction of the trend-term affects the least-squares estimates of the other parameters in (19) or in the corresponding variations such as (21).

(e) Remarks on table II of C.O.

If the mean values of the estimates of the parameters listed in the table are normally distributed, then the t - test, with 19 degrees of freedom, may be applied to test whether the bias is significant. On the basis of this test, there is no bias in the table which can be judged as statistically significant at the 5% level.

On the other hand, the standard deviation of these estimates reveals unruly behavior in several cases. The standard error of the mean of estimates of k is larger than the numerical mean value in most cases, being largest for types A and B and smallest for types D and E. The variance of the estimates of b_{12} and b_{13} is also largest for types A and B. In the three cases where the trend was fitted, the standard error of the mean of the estimates of b_{1t} exceeds the numerical mean value of the estimates.

The mean of the estimated multiple correlation coefficient behaves curiously. Equations 2 and 6 reveal a large estimated multiple correlation coefficient with comparatively small variance in spite of the lack of precision in estimating k and b_{1t}. It is also interesting that this variance in equation 6 is approximately as small as in equation 10, where the disturbance is of type D, (serially independent), and the bogus variable t is not present.

In Table II there is also evidence that the residuals of the regression are "biassed toward randomness" when the statistic $\frac{\delta^2}{s^2}$ is used. This is strikingly illustrated in equation 6, on comparing the mean of the actual error series with the mean of estimated residuals. It is also illustrated in equations 5, 4, 2.

(f) Further investigation of $\frac{\delta^2}{s^2}$.

Table III, p. 44, C.O. refers to the cases where the disturbances and explanatory variables all have the same autoregressive structure of type B. As in Table I, the hypothesis $\rho = 0$ is tested, but the situation here is more convincing since the autoregressive structure of the disturbances is known. The table shows fewer rejections of the hypothesis $\rho = 0$ as the number of parameters is increased, confirming the "biassedness toward randomness" in the use of $\frac{\delta^2}{s^2}$, with tables of the distribution of $\frac{\delta^2}{s^2}$.

(g) Further remarks.

There are further interesting tables in C.O. which are not discussed here, since they merely corroborate the main themes that auto-correlation of the disturbances may seriously impair the efficiency of the least-squares estimates, and the statistic $\frac{\delta^2}{s^2}$ is misleading in the presence of auto-correlation. Table VI, p. 47, should be mentioned, however, as it indicates that the results, when different types of autoregressive structure are used in the same equation, are compatible with the aforementioned themes.

As regards the autoregressive transformation indicated by (3), the least-squares procedure applied to the transformed variables would remove the autocorrelation of the disturbances; but the main difficulty is that in practice the true values of the β's are not known. Cochrane and Orcutt offer no definite solution to this problem, but make two suggestions.

Suggestion one is to find the β's by the following iterative process. First, obtain the least-squares estimates of the regression parameters and compute the residuals. From these residuals, compute least-squares estimates of the autoregressive parameters, then use these to make an autoregressive transformation of the form (3). Re-estimate the regression parameters, using this transformation and put these revised estimates back in the original equation to get a new series of residuals. Then repeat the whole process on these residuals, and terminate the iterations when no further adjustments are necessary. The authors do not recommend the above process because the residuals will be "biassed toward randomness", hence the transformation may not effectively correct the lack of serial independence.

Suggestion two is try different values of the β's in the autoregressive transformation (3), and accept as suitable those values which render the autocorrelations of the series of computed residuals sufficiently close to the expected values of the autocorrelations of "random" (serially independent) series of the same length. In the case $\beta_2 = \beta_3 = .. = \beta_m = 0$, and $\beta_1 \neq 0$, they suggest $\beta_1 = 1$ as a first try; and conjecture that in many economic relations this will render the disturbances approximately random. Although the authors prefer the second of the two suggested methods, the reviewer believes the second method may be as unreliable as the first, because observed residuals can be misleading approximations to the true disturbances.

3. The O.C. Paper.

(a) Object of the paper.

As stated in the introduction, this paper investigates by means of a sampling experiment, the bias and loss of efficiency in estimating the structural parameters of a system of two linear stochastic equations. Not only is the question of autocorrelation of the disturbances considered, but also the behavior of maximum likelihood and least-squares estimates when the disturbances are non-autocorrelated.

(b) The models used in the experiment.

To reduce computations to a minimum, two simple models were adopted for the study. They are defined as follows

Model I

(24) $$x_t = a_o + a_1 y_t + (u_{1t} + u_{2t})$$

(25) $$y_t = b_o + b_1 x_{t-1} + u_{3t}$$

where the error terms u_{it} ($i = 1, 2, 3$) are generated by the auto-regressive equations

(26) $$u_{it} = u_{i,t-1} + \varepsilon_{it}$$

and ε_{it} denote series of independent disturbances. The endogenous variables in the two equations of Model I are x_t and y_t. The values of the parameters are taken to be

(27) $$\begin{cases} a_o = b_o = 0 \\ a_1 = 1.0 \\ b_1 = 1.4 \end{cases}$$

These particular values were selected so that x_t and y_t would have the same autoregressive structure as claimed by Orcutt [9] for the series used in Tinbergen's [10] model of the economic system of the United States. Thus, from (24) and (25) it would follow that

(28) $$x_t = x_{t-1} + 0.4(x_{t-1} - x_{t-2}) + \eta_{1t}$$

(29) $$y_t = y_{t-1} + 0.4(y_{t-1} - y_{t-2}) + \eta_{2t}$$

where η_{1t}, η_{2t} are linear combinations of the ε_{it}.

Model II

The only difference between this model and Model I is that u_{3t} of (25) is replaced by u_{2t}. Thus, the error terms of the two equations will be correlated, in contradistinction to Model I. Prime notation will be used to designate the equations of Model II as follows.

(30) $$x'_t = a_o + a_1 y'_t + (u_{1t} + u_{2t}).$$

(31) $$y'_t = b_o + b_1 x'_{t-1} + u_{2t}.$$

(c) Identification of the parameters in Models I and II.

In the sampling experiment to be described presently, the parameters a_o, b_o, a_1, b_1 will be estimated by diverse methods. The authors state (O.C. p. 362) that both models are exactly identified. This statement would be correct if it is assumed that the error terms u_{it} $(i = 1, 2, 3)$ are serially independent and if in Model I the independence of $u_1 + u_2$ from u_3 is not known a priori. For, if this independence were known a priori, then Model I would be over-identified. This can be seen by applying Theorem (139) of Koopmans [11].

As the error terms are, in fact, autocorrelated in virtue of (26), it becomes necessary to consider in some detail the question of identifiability in these models. The lagged endogenous variable, x_{t-1}, is not independent of the error terms, in both models; consequently the aforementioned theorem of Koopmans cannot be applied directly.

Consider, for the moment, the two equations (24), (25) of Model I, where the error terms u_{it} satisfy (26). By taking first differences, equations (24) and (25) become

(32) $$X_t = A_o + A_1 Y_t + (\mathcal{E}_{1t} + \mathcal{E}_{2t})$$

(33) $$Y_t = B_o + B_1 X_{t-1} + \mathcal{E}_{3t}$$

where

(34) $$\begin{cases} X_t = x_t - x_{t-1} \\ Y_t = y_t - y_{t-1} \end{cases}$$

and

(35) $$A_o = a_o, \quad B_o = b_o, \quad A_1 = a_1, \quad B_1 = b_1$$

in view of (27).

Since $\mathcal{E}_{3t}$ is independent of X_{t-1}, the theorem on identification is now applicable and shows in fact that Model I is over-identified, under this transformation, if the independence of $\mathcal{E}_{1t} + \mathcal{E}_{2t}$ and $\mathcal{E}_{3t}$ is assumed known. If $\mathcal{E}_{3t}$ is replaced by

ε_{2t}, and the first differences of (30) and (31) are taken, namely:

(36) $$X_t' = A_o + A_1 Y_t' + (\varepsilon_{1t} + \varepsilon_{2t})$$

(37) $$Y_t' = B_o + B_1 X_{t-1}' + \varepsilon_{2t}$$

the theorem on identification is again applicable, on account of the independence of ε_{2t} and X_{t-1}. Thus, Model II is exactly identified under this transformation, as the disturbances in (36) and (37) are now correlated.

It should be pointed out, here, that if the transformed variables of (34) are not used in the estimation of the parameters a_o, a_1, b_o, b_1, but rather the original x_t and y_t, then both Models I and II will be overidentified. Thus, (36) and (37) would now be written in the form

(38) $$x_t - a_1 y_t - x_{t-1} - a_1 y_{t-1} + 0 = \varepsilon_{1t} + \varepsilon_{2t}$$

(39) $$0 + y_t - b_1 x_{t-1} - y_{t-1} + b_1 x_{t-2} = \varepsilon_{2t}$$

and since the disturbances are independent of the predetermined variables, x_{t-1}, x_{t-2}, y_{t-1}, the system is clearly seen to be overidentified. The corresponding system of Model I would, a fortiori, be overidentified.

Further comments will be made in (e) on the estimation procedure employed by the authors. It is advisable, at this point, to describe the sampling experiment which they carried out.

(d) Description of the sampling experiment

Let ε be a random variable which assumes each of the values -4, -3, ..., 0, 1, ..., 4 with equal probability $\frac{1}{9}$. By referring to tables of random sampling numbers [8], 60 series of ε_t were constructed, each one 20 items in length. These 60 series were divided into 3 groups ($i = 1, 2, 3$), each group consisting of 20 series of $\mathcal{E}_{it}$. Inserting the values of $\mathcal{E}_{it}$ of these constructed series into (26), with initial values $u_{io} = 0$, the three series of u_{it} were generated, giving in all, 20 groups of 3 series in each group.

The error terms u_{it} were then introduced in the equations of Models I and II to generate the series for x_t and y_t, taking $x_o = 0$, and referring to (27).

(e) Methods of estimating the coefficients a_o, a_1, b_o, b_1.

To describe the methods of estimation and explain the results in Table I (O.C. p. 365), it is necessary to introduce the esoteric notation of the authors, who seem to use regression terminology indiscriminately whether they estimate a reduced form equation or a structural equation not in the reduced form.

Denote equations (24)and (25) of Model I by I and II respectively; and equations (30) and (31) of Model II by IV and V respectively. Let III and VI represent the following equations.

(40) $$\text{III:} \quad x_t = \rho_o + \rho_1 x'_{t-1} + (u_{1t} + u_{2t} + a\, u_{3t})$$

(41) $$\text{IV:} \quad x'_t = \rho_o + \rho_1 x'_{t-1} + [u_{1t} + (1 + a_1)\, u_{2t}]$$

where

(42) $$\rho_o = a_o + a_1 b_o$$

(43) $$\rho_1 = a_1 b_1$$

Thus, II and III constitute the reduced system of Model I; V and VI constitute the reduced system of Model II.

The following equations are designated by IA and IVA

(44) $$\text{IA:} \quad x_t = (\rho_o - \frac{\rho_1}{b_1} b_o) + \frac{\rho_1}{b_1} y_t + (u_{1t} + u_{2t})$$

(45) $$\text{IVA:} \quad x_t' = (\rho_o - \frac{\rho_1}{b_1} b_o) + \frac{\rho_1}{b_1} y_t' + u_{2t}.$$

Obviously IA and IVA are the same as I and IV but the authors use the notation IA and IVA in the special circumstance that the estimates of a_o and a_1 in I and IV are obtained from the estimates of ρ_o, ρ_1, b_o, b_1 in the reduced systems II, III and V, VI respectively.

The notation O., F.D., S.D. refers to autoregressive transformations carried out on the equations before estimation of the parameters. O. means the original equation. F.D. stands for the first-difference transformation

(46) $$\begin{cases} X_t = x_t - x_{t-1} \\ Y_t = y_t - y_{t-1} \\ U_t = u_t - u_{t-1} \end{cases}$$

while S.D. stands for the second-difference transformation

(47) $$\begin{cases} W_t = X_t - X_{t-1} \\ Z_t = Y_t - Y_{t-1} \\ V_t = U_t - U_{t-1} \end{cases}$$

The method of least-squares was employed in equations I, II, III, IV, V, VI to estimate a_o, a_1, b_o, b_1, ρ_o, ρ_1. The estimates of a_o and a_1 in IA and IVA were obtained from the above estimates of b_o, b_1, ρ_o, ρ_1 by the limited information method explained by the notation of (44) and (45). The terminology "reduced form estimates" employed in O.C. should really be replaced by "limited information estimates." As may be seen from Koopmans [6], the use of the reduced form is accidental, whereas the limitation on the information used is essential to the method.

Table I (O.C., p. 365) summarizes the mean values of the estimates of the parameters in all the equations, and for each of the transformations O., F.D., S.D. Other relevant information is given in the table under the appropriate headings.

(f) Comments on the estimation procedure.

The estimates obtained under the transformation F.D. are of special interest since the transformed error terms are serially uncorrelated. If the ε_{it} of (26) were normally and independently distributed, with mean 0 and fixed variance, the following facts would obtain

(i) In Model I, the least-squares estimates in I F.D. and II F.D. would be the ordinary maximum-likelihood estimates, since the disturbances in I F.D. are independent of those in II F.D.

(ii) The estimates in IA F.D. would be limited information estimates of a_o, a_1 (ignoring the independence between $\varepsilon_1 + \varepsilon_2$ and ε_3).

(iii) The estimates in IVA F.D. and V F.D. would be ordinary maximum likelihood estimates, due to exact identification in Model II under F.D. (See [12], p. 120, 3.2.1; p. 123, 3.2.4.)

(iv) The least-squares estimates in IV F.D. will not be consistent, since the variable $Y_t = y_t' - y_{t-1}'$ is not independent of $U_{1t} + U_{2t} = u_{1t} + u_{2t} - u_{1,t-1} - u_{2,t-1}$. (This statement holds irrespective of normality of ε_{1t}).

(v) The least-squares estimates in III F.D. and VI F.D. would be maximum likelihood estimates of ρ_o, ρ_1.

As the $\mathcal{E}_{it}$ were, in fact, taken to be rectangularly distributed, the above statements must be qualified to read "quasi maximum likelihood" in place of "maximum likelihood," and "quasi limited information" in place of "limited information." It would be an interesting investigation in itself, either by the brute force method of a sampling experiment, or by theoretical means, to inquire how seriously departure from normality of the $\mathcal{E}_{it}$ would affect the properties of the estimates in question.

The results under O. and S.D. will indicate how badly the quasi maximum likelihood and limited information estimates misbehave in this particular sampling experiment, when it is falsely assumed that the error terms in both models are serially uncorrelated, whereas they are in fact positively autocorrelated (under O.) or negatively autocorrelated (under S.D.)

It should be mentioned here that the authors do not discriminate between quasi-maximum likelihood and maximum-likelihood estimates; nor do they distinguish between O., F.D. and S.D. in comparing the efficiency of maximum-likelihood and limited information estimates. This is especially apparent in the discussion on page 368 O.C. As explained on page 16 supra, it is under F.D., where the error terms are serially uncorrelated, that the least-squares estimates are the same as quasi-maximum likelihood estimates.

(g) Discussion of Table I (O.C. p. 365)

As in the case of C.O. Table II, the t-test, with 19 degrees of freedom is applied, in testing whether the bias in the mean of the estimates is significantly different from 0. Statements about bias, which appear in the discussion below, are made on the basis of this t-test, at the 5% level of significance.

In the case of the original equations (marked O.) the estimates of the constants, a_o, b_o, ρ_o fared badly. The standard error of the mean of these estimates is relatively large, especially for IA O. In III O. and VI O. the mean bias of the estimates of ρ_o is large enough to be judged significant even at the 1/10% level.

The estimates of a_1 and b_1 in the original equations also fared badly. The mean bias for all equations marked O. is significant. The standard errors of the means, however, are much smaller than those of the means of the estimates of a_o and b_o under O.

The results pertaining to F.D. deserve close attention in view of the remarks under (e) and (f). The results of models I and II will be considered separately. Wherever the expression "maximum likelihood estimates" and "limited information estimates" appear, it is to be understood that these are "quasi," since the ε_{it} in (26) are rectangularly distributed.

Model I (F.D.)

(i) Bias of maximum-likelihood estimates (equations I, II).

No significant bias evident.

(ii) Bias of limited information estimates (equations IA, II).

Mean of estimates of a_1 in I A significantly biassed.

(iii) Comparison between variance of maximum likelihood estimates and variance of limited information estimates (equations I, I A). The ratio of the latter to the former is given below

a_o	a_1
55.8	8.2 (10.6)

The figure in parenthesis indicates the ratio of variances which are computed assuming the true values of a_o and a_1 are given. (See column 10 of table.)

(iv) Bias of maximum-likelihood estimates in III.

Mean of estimates of ρ_o and ρ_1 significantly biassed. (See remark (iv) under model II (F.D.) below.)

Model II (F.D.)

(i) Bias of maximum-likelihood estimates (equations IV A, V).

The mean bias from a_o and from a_1 is significant in IV A.

The mean bias from b_1 is significant in V.

(ii) Bias of single-equation least squares estimates in IV.

The mean bias from a_1 is highly significant.

(iii) Comparison between variance of maximum-likelihood estimates and variance of single equation least-squares estimates. (Equations IV A, IV.) The ratio of the former to the latter is given below.

a_o	a_1	
7.3	33.4	(3.1)

(iv) Bias of maximum likelihood estimate in VI.

Mean of estimates of ρ_1 is significantly biassed. It may be remarked that the downward bias in the estimate of ρ_1, both in Models I and II, is compatible with results of Hurwicz [13] and Leipnik [14].

(h) Further remarks.

The bias in the statistic $\frac{\delta^2}{s^2}$, which was discussed in C.O. is also considered in O.C.; but it is omitted from the discussion, since the results merely corroborate the conclusion in C. O. It is, perhaps, unfortunate that the $\mathcal{E}_{it}$ of (26) were not constructed from normal populations in the sampling experiment, but the results are highly suggestive, notwithstanding. As the authors indicate, a knowledge of the autocorrelation may be of great help in controlling bias and loss of efficiency. (It is curious from Table I O.C. that the second-difference transformation fared so well, in view of the presence of negative autocorrelation in the error terms for this case.) As yet, no simple and reliable method is available for estimating the autocorrelation in the disturbances of linear relationships. The fact that very little seems to be known about small sample bias of maximum likelihood estimates in general, adds no consolation to the situation.

(i) Additional further remarks.

The following statements appear C. O., p. 371. "Care must be taken when drawing general conclusions from sampling experiments of the type considered in this paper, but the results appear to us to be of a striking and significant nature. They indicate that unless it is possible to specify with some degree of accuracy the inter-correlation between the error terms of a set of relations and unless it is possible to choose approximately the correct autoregressive transformation so as to randomize the error terms, then a certain amount of skepticism is justified concerning the possibility of estimating structural parameters from aggregative time series of only twenty observations when generated by systems analagous to those examined in this paper."

The systems under study in the above paper are particularly simple, and there are relatively few complications present. In practice, the systems of equations are not as simple as these, and there may be present several other types of complications which might conceivably eclipse some of those types in the systems studied by O.C. A list of problems will be stated in (j), but it is advisable to pause at this point and reflect on the merits and demerits of the sampling method in attacking problems associated with the estimation of the structural parameters in a system of stochastic equations.

The sampling method is usually relatively simple as compared with the theoretical method; this is its chief advantage. It also provides a check on theory, and may even suggest further theoretical problems. But opposed to these advantages are certain serious disadvantages.

In the first place, to confine the consumption of time, energy, and expense within reasonable limits, the system under experimentation must be simple and severely limited in its generality. High speed computing devices may alleviate this defect, although the expense may still be considerable.

Secondly, even in a simple experiment it may be difficult to separate what are the precise causes which correspond to certain effects. For instance, a certain estimate in the system investigated by O.C. may be biassed or inefficient. One could ask, is this due mainly to autocorrelation of the disturbances, or to ignoring knowledge about the distribution of the disturbances, or to small sample complications of "asymptotic" estimates.

Finally, the sample randomly selected to investigate the system may itself be a misleading one although the probability associated with obtaining a misleading sample may be small.

The simplicity of the system in O.C. might be a cause for overlooking certain types of complications which could be a serious handicap in the estimation of structural estimates. For this reason, a list of problems is given in (j) which is relevant to the question of structural estimation.

(j) Some problems in estimation of structural parameters in simultaneous equations.

A convenient classification of these problems is according to the following types

(1) Incorrect specification.

(2) Incomplete specification.

(3) More general models.

(4) Sampling problems.

These will be subdivided further, as indicated hereunder.

(1) <u>Incorrect specification</u>

(1.1) Autocorrelation of disturbances when they are assumed to be serially uncorrelated.

(1.2) Dependence between the disturbances and the predetermined variables, when independence is assumed.

(1.3) Omission of variables.
This destroys the validity of the assumptions about the disturbances, and may be the cause of (1.1), (1.2)

(1.4) Mislabelling of the endogenous and exogenous variables, especially in an incomplete system. This may also be the cause of (1.1), (1.2).

(1.5) Some of the equations which are assumed to be linear are in fact non-linear. This is also a possible cause of (1.1), (1.2).

(1.6) The disturbances have a non-normal distribution when they are assumed to have a multivariate normal distribution.

(2) <u>Incomplete specification</u>

The problems of this type are essentially problems of incorrect specification, but they differ from (1) in this sense. Under (2) we do not know enough about the model, whereas under (1) we know too much, that is, our presumed knowledge is erroneous. For brevity, only two examples of problems under (2) is given here, as several others may be realized by referring to (1) above.

(2.1) Some restrictions on the structural parameters exist which are not known.

(2.2) Some restrictions on the distribution of the disturbances exist, which are not known.

(3) <u>More general models</u>

The problems stated under (1) and (2) suggest possible ways of generalizing the models used in systems of simultaneous stochastic equations.

(3.1) Autocorrelation of the disturbances. This raises three questions. First the possibility of identifying the parameters to be estimated; second, how to estimate the parameters; third, how to estimate the degree of autocorrelation of the disturbances.

(3.2) Some of the equations may be non-linear. The problem of identification and estimation is raised here also.

(4) Sampling problems

In most problems of structural estimation, the usual methods reduce to some form of maximum likelihood. Even when such estimates are available, the question of the properties of such estimates is raised. In those cases where it is known that these are consistent and/or asymptotically efficient, it would be interesting to know how these estimates behave in small or even medium size samples. The need of such knowledge is rather urgent when it is realized that solutions of problems in terms of asymptotically normal estimates may be quite undesirable for small or medium size samples.

REFERENCES

[1] Guy H. Orcutt and Donald Cochrane, "A sampling study of the merits of autoregressive and reduced form transformations in regression analysis," Journal of the American Statistical Association, Vol. 44, 1949, pp. 356-372.

[2] D. Cochrane and G. H. Orcutt, "Application of least squares regression to relationships containing autocorrelated error terms," Journal of the American Statistical Association, Vol. 44, 1949, pp. 32-61.

[3] F. N. David and J. Neyman, "Extension of the Markoff theorem on least squares," Statistical Research Memoirs, Vol. II, London, 1938, pp. 105-116.

[4] D. G. Champernowne, "Sampling theory applied to autoregressive sequences," Journal of the Royal Statistical Society, Series B, Vol. X, 1948, pp. 204-242.

[5] J. von Neumann, "Distribution of the ratio of the mean square successive difference to the variance," Annals of Mathematical Statistics, Vol. 12, pp. 367-395.

[6] Tjalling Koopmans, "Serial correlation and quadratic forms in normal variables," Annals of Mathematical Statistics, Vol. 13, 1942, pp. 14-33.

[7] B. S. Hart and J. von Neumann, "Tabulation of the probabilities for the ratio of the mean square successive differences to the variance," Annals of Mathematical Statistics, Vol. 13, pp. 207-214.

[8] M. G. Kendall and B. Babington Smith, "Tracts for computers," No. 24, Cambridge University Press, 1939.

[9] G. H. Orcutt, "A study of the autoregressive nature of the time series used for Tinbergen's model of the economic system of the United States 1919-1932," Journal of the Royal Statistical Society, Vol. X, Series B, 1948, pp. 1-53.

[10] J. Tinbergen, "Statistical testing of business cycle theories Vol. II; Business cycles in the United States of America 1919-32," League of Nations, Geneva, 1939.

[11] Tjalling C. Koopmans, "Statistical methods of measuring economic relationships," Cowles Commission Discussion Paper: Statistics No. 310.

[12] T. C. Koopmans (editor), "Statistical inference in dynamic economic models," Cowles Commission Monograph No. 10, New York, John Wiley and Sons.

[13] L. Hurwicz, "Least squares bias in time series," Cowles Commission Monograph No. 10.

[14] R. B. Leipnik, "Distribution of the serial correlation coefficient in a circularly correlated universe," Annals of Mathematical Statistics, Vol. 18, 1947, pp. 80-87.