

Determination of Linear Relations Between Systematic Parts of Variables with Errors of Observation the Variances of Which are Unknown*

R.C. GEARY

Department of Applied Economics, University of Cambridge

Abstract: Given a sufficient number of *instrumental* variables significantly correlated with the *investigational* variables, consistent estimates of the coefficients of the linear relations can be determined (if they exist), without knowledge of the disturbance variances. The estimates are discussed from the viewpoint of probability convergence. In the case of two investigational and one instrumental variable, all three variables distributed on the normal surface, the distribution of the estimate of the coefficient is found exactly for all sample sizes, on certain hypotheses. The distribution function is remarkably simple. The applicability of the theorem to economic time series is discussed by (a) comparing the probability inferences derived from this Model A with those for the simplest stationary time-series model, termed Model B, and (b) by comparing the large-sample variances on several models. It is found that the theory can be used with confidence when the series are not too short and the error variances not too large. The theory is applied to a particular time series, showing that the accuracy of the estimate of the coefficient depends on the correlation between the instrumental variable and the two investigational variables. The theory to which reference is made in Sections II, III, and IV, relating to the two-investigational-variable case, is extended to many variables and tests are given, applicable when samples are not small, for determining the significance of coefficient estimates.

I INTRODUCTION

Since the appearance in 1934 of Ragnar Frisch's well-known book *Statistical Confluence Analysis by Means of Complete Regression Systems*, statisticians have come to recognize that, whatever else they are, the classical regression equations are not functional relations between the

*This paper was first published in *Econometrica*, Vol. 17, No. 1, 1949. It is reproduced here with minor editorial amendments with the kind permission of The Econometrics Society.

variables. Reiersøl (1941, 1945) and Geary (1942, 1943) have shown how relations, *consistent* in the statistical sense, between variables can be derived. Reiersøl used what he termed the *instrumental sets* of variables in contradistinction to the *investigational set*, which are the variables the systematic parts of which enter into the relations; Reiersøl's method is that used in this communication. Geary showed how relations could be established from knowledge of the investigational set alone by having recourse to semi-invariants of power greater than two and of dimension greater than unity.

It is assumed throughout the paper that the variances of the disturbances are not known in advance or cannot be efficiently estimated from the observations. The disturbance variances occur explicitly in several formulae, usually in order that means and variances of estimates of the coefficient of relationship, computed for different mathematical models, may be compared. It will be shown elsewhere (Geary, 1948) that when the disturbance variance is known or can be estimated and when certain other conditions are satisfied, methods of estimation of the coefficients, more efficient in the statistical sense than those contemplated here, can be devised.

Most of the present paper is devoted to the frequency distribution and the sampling errors of the coefficient in the simplest case, that in which a single relation is assumed to subsist between the systematic parts of two variables. In the final section, however, the errors in the coefficients of relations in more than two variables are dealt with.

II THE PROBLEM

Measures of a random sample of n of $(2p - 1)$ variables in two sets are observed:

$$\left. \begin{array}{l} \text{(i)} \quad x_{it} \ (i = 1, 2, \dots, p) \\ \text{(ii)} \quad x_{rt} \ (r = p + 1, p + 2, \dots, 2p - 1) \end{array} \right\} \quad (t = 1, 2, \dots, n),$$

where the x_{it} represent the measures of the investigational set and the x_{rt} the measures of the instrumental set. Each variable of the latter set is significantly correlated with at least one variable of the former set. It will be shown later that the higher the correlation (in a sense defined) between the two sets the more efficient the estimates of the coefficients of the linear relation.

Each observation of the investigational set is made subject to *error* or *disturbance*, i.e.,

$$x_{it} = x'_{it} + x''_{it}, \quad (1)$$

where x'_{it} is the *systematic* part and x''_{it} the error or disturbance. The following linear relation holds *exactly* (i.e., for all t) between the systematic parts:

$$\sum_{i=1}^p \alpha_i x'_{it} = C, \quad (2)$$

where C is a constant. It is assumed that the relation cannot be expressed in fewer variables than p . Estimates that are consistent in the statistical sense are derived for the coefficient ratios α_i/α_1 . The main object of the paper is to discuss the frequency distribution of these estimates which are found as follows. From (2)

$$\sum_i \alpha_i (x'_{it} - \bar{x}'_i) = 0, \quad (3)$$

where $n\bar{x}'_i = \sum_t x'_{it}$. On multiplying by x_{rt} and taking means of n , we have

$$\sum_{i=1}^p \alpha_i n_{ir} = 0 \quad (r = p+1, p+2, \dots, 2p-1), \quad (4)$$

where

$$n_{ir} = \frac{1}{n} \sum_t (x'_{it} - \bar{x}'_i) x_{rt}. \quad (5)$$

Assume that the disturbances x''_{it} are independent of one another, of the systematic parts of the investigational set, and of the instrumental set, that their universal means are zero, and that the variances of the variables of the instrumental set are finite, which assumptions rule out, for the moment, a lagged investigational variable as an instrumental variable. The systematic parts x'_{it} can be any numbers whatever; and it is not necessary to assume that the error variance is the same for each t . It follows that the Reiersøl method is applicable when the variables are stationary time series, under unrestrictive conditions. Let

$$m_{ir} = \frac{1}{n} \sum_t (x_{it} - \bar{x}_i) x_{rt} = \frac{1}{n} \sum_t (x_{it} - \bar{x}_i) (x_{rt} - \bar{x}_r) \quad (6)$$

which, different from n_{ir} , may be computed from the observations. Then

$$m_{ir} - n_{ir} = \frac{1}{n} \sum_t (x''_{it} - \bar{x}_i) x_{rt}, \quad (7)$$

so that

$$E(m_{ir} - n_{ir}) = 0 \quad (8)$$

and

$$\text{var}(m_{ir} - n_{ir}) = E(m_{ir} - n_{ir})^2 = 0(1/n). \quad (9)$$

It follows that $(m_{ir} - n_{ir})$ tends in probability towards zero with increasing n . We write

$$\sum_i m_{ir} a_i = 0 \quad (r = p+1, p+2, \dots, 2p-1). \quad (10)$$

The ratios a_i / a_1 ($i = 2, 3, \dots, p$), solutions of the simultaneous equations (10), are continuous functions of the m_{ir} and hence tend in probability towards the same functions of the n_{ir} — which, by hypothesis, are determinate — i.e., the α_i / α_1 .

The method can, of course, be used to estimate the coefficients of more than one linear relation between the systematic parts of the variables, provided that these relations are in the fewest variables, which implies, incidentally, that no two relations are such that all the variables in one appear in the other. The method is accordingly not well adapted to the discovery of *structural* relations unless the form of these relations (i.e., the variables that they contain) is given in advance and satisfies the condition just specified.

The difficulty of obtaining a number of instrumental variables highly correlated with the investigational variables is, of course, a circumstance that limits the usefulness of the method described in this section. When the variables are economic time series Reiersøl has made the ingenious suggestion that lagged or forwarded investigational variables can be used as instrumental variables. This method usefully exploits the property characteristic of economic time series, namely serial correlation. In applying this method we would take x_{it} ($t = 1, 2, \dots, n$) for one of the x_{rt} . This introduces a slight complication into the convergence in probability theorem, in the proof that $(n_{ir} - m_{ir})$ converges in probability to zero when $r = i$. We can easily show that in this case $E(m_{ir} - n_{ir}) = 0(1/n)$ and (as before) $E(m_{ir} - n_{ir})^2 = 0(1/n)$. It is necessary to assume in addition the existence of the fourth moments of the errors.

III THE EXACT-FREQUENCY DISTRIBUTION OF THE COEFFICIENT a

The problem dealt with in this section is the simplest one that can arise in the order of ideas of this paper, namely that in which the observed sample is (x_{1t}, x_{2t}, x_{3t}) ($t = 1, 2, \dots, n$), of which x_{1t} and x_{2t} are investigational and x_{3t} is instrumental. The following assumptions are made:

(i) (x_{1t}, x_{2t}, x_{3t}) is distributed on the normal surface with known variance-covariance matrix $\|\mu_{ij}\|$ ($i, j = 1, 2, 3$) the same for each t ; in practice this matrix will usually have to be estimated from the sample;

(ii) the sets of observations are statistically independent for different t ;

(iii) the investigational variables x_{it} ($i = 1, 2$) are the sum of a systematic part x'_{it} and a disturbance in error x''_{it} ; as the latter are independent of one another, of the systematic parts, and of the instrumental variables, both systematic parts and disturbances must, by the Cramér-Lévy theorem, be each normally distributed.

It is recognized that this theory is not formally applicable to any plausible model of economic time series since, on account of the phenomenon of serial correlation, we cannot reasonably assume that the frequency distributions are the same for different t : at least the means should be deemed to alter. The application of the theorem to time series is discussed in detail in the following section.

(iv) The relation $x'_{it} = \alpha x'_{2t} + C$ holds exactly between the systematic parts. The problem is to determine the frequency distribution of the estimate a of the coefficient α given by

$$a = \frac{X_1}{X_2}, \tag{11}$$

where

$$\begin{aligned} nX_1 &= \sum_{t=1}^n (x_{1t} - \bar{x}_1)(x_{3t} - \bar{x}_3), \\ nX_2 &= \sum_t (x_{2t} - \bar{x}_2)(x_{3t} - \bar{x}_3). \end{aligned} \tag{12}$$

Let the joint frequency of (x_{1t}, x_{2t}, x_{3t}) be

$$\frac{A^{\frac{1}{2}}}{(2\pi)^{3/2}} \left\{ \exp \left(-\frac{1}{2} Q_t \right) \right\} dx_{1t} dx_{2t} dx_{3t}, \tag{13}$$

where

$$Q_t = \sum_{i=1}^3 \sum_{j=1}^3 \alpha_{ij} x_{it} x_{jt}, \tag{14}$$

with

$$\alpha_{ij} = \alpha_{ji} = \mu^{ij}, \quad (15)$$

the latter being the element of the reciprocal matrix $\|A\|$ (of which the determinant is A) of the variance-covariance matrix $\|\mu_{ij}\|$. The universal means of x_{it} may be assumed to be zero, without loss of generality.

The characteristic function $f(s, t)$ of (x_1, x_2) is

$$f(s, t) = \frac{A^{n/2}}{(2\pi)^{3n/2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left\{i(sX_1 + tX_2) - \frac{1}{2} \sum_t \sum_i \sum_j \alpha_{ij} x_{it} x_{jt}\right\} \prod_t dx_{1t} dx_{2t} dx_{3t} \quad (16)$$

which is known to equal¹

$$\left(\frac{A}{A^*}\right)^{(n-1)/2}, \quad (17)$$

where

$$A^* = \begin{vmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} - \frac{is}{n} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} - \frac{it}{n} \\ \alpha_{31} - \frac{is}{n} & \alpha_{32} - \frac{it}{n} & \alpha_{33} \end{vmatrix}. \quad (18)$$

Hence

$$f(s, t) = \left\{ 1 - \frac{2i}{n}(k_{10}s + k_{01}t) + \frac{1}{n^2}(k_{20}s^2 - 2k_{11}st + k_{02}t^2) \right\}^{-(n-1)/2}, \quad (19)$$

where, on using (15) and well-known determinantal properties, we have

$$\begin{aligned} k_{20} &= \mu_{11}\mu_{33} - \mu_{13}^2, \\ k_{10} &= \mu_{13}, \\ k_{11} &= \mu_{23}\mu_{13} - \mu_{12}\mu_{33}, \\ k_{01} &= \mu_{23}, \\ k_{02} &= \mu_{22}\mu_{33} - \mu_{23}^2. \end{aligned} \quad (20)$$

1. See, for example, H. Cramér (1946), *Mathematical Methods of Statistics*, p. 405.

Geary (1944), generalizing a result of H. Cramér (1937),² has shown that the frequency distribution of a given by (11) can, under very general conditions, be expressed in the form

$$\phi(a) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} ds \left[\frac{\partial f(s, t)}{\partial t} \right]_{t=-as}, \quad (21)$$

which, applied to (19), gives

$$\phi(a) = \frac{(n-1)}{2\pi i n} \int_{-\infty}^{\infty} ds \left\{ ik_{01} + \frac{s}{n} (k_{11} + k_{02}a) \right\} \left\{ 1 - \frac{2is}{n} (k_{10} - k_{01}a) + \frac{s^2}{n^2} (k_{20} + 2k_{11}a + k_{02}a^2) \right\}^{-(n+1)/2}. \quad (22)$$

Suppose n is an odd number. It is easy to show that $(k_{20}k_{02} - k_{11}^2)$ is always positive, hence that

$$k_{20} + 2k_{11}a + k_{02}a^2$$

is always positive. Accordingly set the last expression in the brackets $\{ \}$ in (22) equal to

$$\left(1 - \frac{Kis}{n} \right) \left(1 - \frac{K'is}{n} \right) \quad (23)$$

so that

$$\begin{aligned} K + K' &= 2(k_{10} - k_{01}a), \\ KK' &= -(k_{20} + 2k_{11}a + k_{02}a^2), \end{aligned} \quad (24)$$

from which we infer that

$$K - K' = 2 \left\{ (k_{10} - k_{01}a)^2 + (k_{20} + 2k_{11}a + k_{02}a^2) \right\}^{\frac{1}{2}}. \quad (25)$$

Hence K and K' are real quantities, positive and negative respectively.

In the integral on the right-hand side of (22) regard s as a complex variable. The function to be integrated has then two poles each of order $(n+1)/2$ at $-ni/K$ and $-ni/K'$, the former accordingly on the negative and the

2. The result is also given in Cramér, *op. cit.*, p. 317.

latter on the positive side of the imaginary axis. If the function be integrated around a closed contour consisting of the real axis and the great semicircle below the real axis, by Cauchy's theorem the integral must equal the residue at the pole $-ni/K$, which is found to be

$$\frac{(n-1)!(n-1)}{\left\{\left(\frac{n-1}{2}\right)!\right\}^2} |KK'|^{(n-3)/2} (K-K')^{-n} \{(k_{20}k_{01} + k_{11}k_{10}) + a(k_{11}k_{01} + k_{10}k_{02})\}. \quad (26)$$

This must be the function $\phi(a)$ required since the integral around the great circle is obviously zero. Making the substitution

$$\begin{aligned} y &= (k_{01}a - k_{01}) / (k_{20} + 2k_{11}a + k_{02}a^2)^{\frac{1}{2}} \\ &= (\mu_{23}a - \mu_{13}) \{(\mu_{22}\mu_{33} - \mu_{23}^2)a^2 - 2(\mu_{12}\mu_{33} - \mu_{23}\mu_{13})a + (\mu_{11}\mu_{33} - \mu_{13}^2)\}^{\frac{1}{2}} \\ &= \left\{ \frac{\mu_{33}(\mu_{22}a^2 - 2\mu_{12}a + \mu_{11})}{(\mu_{23}a - \mu_{13})^2} - 1 \right\}^{-\frac{1}{2}} \end{aligned} \quad (27)$$

and using (24), (25), and (27), we find

$$\phi(a) da = \frac{\left(\frac{n-2}{2}\right)!}{\left(\frac{n-3}{2}\right)! \sqrt{\pi}} (1+y^2)^{-n/2} dy, \quad (28)$$

so that $y\sqrt{n-1}$ is distributed as the Gosset-Fisher t with $(n-1)$ degrees of freedom.

This result³ is reminiscent of that of Geary (1930) that if X'_1 and X'_2 are normally distributed with X'_2 unlikely to assume negative values and

$$a' = X'_1 / X'_2 \quad (29)$$

then

$$y'\sqrt{n-1} = (\mu'_{01}a' - \mu'_{10}) / (\mu'_{02}a'^2 - 2\mu'_{11}a' + \mu'_{20})^{\frac{1}{2}}, \quad (30)$$

3. The simplicity of the result as compared with the manner of derivation of this theorem suggests that it might be possible to show directly that y is the quotient of a normal mean and standard deviation for a sample of n .

where μ'_{10} and μ'_{01} are the universal means of X'_1 and X'_2 and μ'_{20} , μ'_{11} , and μ'_{22} the variances and covariances, is normally distributed with mean zero and variance unity. What would this latter theorem show if X_1 and X_2 given by (12) were normally distributed (as, of course, they tend to be when n tends towards infinity)? We find

$$y' = \left\{ \frac{\mu_{33}(\mu_{22}a^2 - 2\mu_{12}a + \mu_{11})}{(\mu_{23}a - \mu_{13})^2} + 1 \right\}^{-\frac{1}{2}} \tag{31}$$

Comparing (31) with (27) (last expression) we see that they are identical except for a change of sign before the unit. At first sight this might appear to be a contradiction since both distributions must in the limit, as n tends towards infinity, be the same. The anomaly is explained by the fact that the identical first term in the brackets is of order n .

In applying formula (27) it will, of course, be necessary in almost every case to estimate the variance-covariance matrix $\|\mu_{ij}\|$ from the observations. The corresponding Studentized problem of finding the distribution of, e.g.,

$$m_{33}(m_{22}a^2 - 2m_{12}a + m_{11}) / m_{23}^2(a - \alpha)^2 \tag{32}$$

where

$$a = m_{13} / m_{23} \text{ and } \alpha = \mu_{13} / \mu_{23}, \tag{33}$$

the m 's being the sample values of the μ 's, would appear to be of considerable complexity which may render it the more attractive to statisticians with greater ingenuity than the writer.

In practical applications of the theorem one will set in the usual way

$$\left\{ \frac{\mu_{33}(\mu_{22}a^2 - 2\mu_{12}a + \mu_{11})}{(\mu_{23}a - \mu_{13})^2} - 1 \right\}^{-1} \leq \frac{\tau^2}{n - 1} \tag{34}$$

where τ is the probability point for the t -distribution corresponding to the predetermined probability 0.1, 0.05, 0.01, etc., so that

$$\frac{\mu_{33}(\mu_{22}a^2 - 2\mu_{12}a + \mu_{11})}{(\mu_{23}a - \mu_{13})^2} \geq 1 + \frac{n - 1}{\tau^2} = \kappa. \tag{35}$$

The limits of a are found from

$$0 \geq a^2(\kappa\mu_{23}^2 - \mu_{22}\mu_{33}) - 2a(\kappa\mu_{13}\mu_{23} - \mu_{12}\mu_{33}) + (\kappa\mu_{13}^2 - \mu_{11}\mu_{33}). \quad (36)$$

As a measure of the possible range of variation of the estimate a corresponding to a given probability level we may take the difference δ of the roots of the expression on the right of (36). We find

$$\delta = \frac{2\left\{\mu_{33}\kappa(\mu_{22}\mu_{13}^2 - 2\mu_{12}\mu_{13}\mu_{23} + \mu_{11}\mu_{23}^2) - \mu_{33}^2(\mu_{11}\mu_{22} - \mu_{12}^2)\right\}^{\frac{1}{2}}}{(\kappa\mu_{23}^2 - \mu_{22}\mu_{33})}. \quad (37)$$

Bearing in mind that κ is at the order of n , a large-sample approximation to δ is

$$\Delta = \frac{2}{\kappa^{\frac{1}{2}}\mu_{23}^2} \left\{ \mu_{33}(\mu_{22}\mu_{13}^2 - 2\mu_{12}\mu_{13}\mu_{23} + \mu_{11}\mu_{23}^2) \right\}^{\frac{1}{2}}. \quad (38)$$

Perhaps the most suggestive transformation is that found by substituting

$$\mu'_{ij} + \delta_{ij}\mu''_{ii}$$

for μ_{ij} , where the μ' and μ'' represent the variance-covariances of the systematic and error parts respectively of the observations. We find

$$\Delta = 2\mu_{33}^{\frac{1}{2}}(\mu''_{11} + \alpha^2\mu''_{22})^{\frac{1}{2}}/\kappa^{\frac{1}{2}}\mu_{23}. \quad (39)$$

The precision of the estimate in the large-sample case accordingly depends inversely on μ_{33} and directly on μ_{23} (when the variables x_{1t} and x_{2t} are given) which is tantamount to stating that we should select (if we have a choice) the instrumental variable with the highest correlation with x_{1t} and x_{2t} . This, of course, is just what would have been anticipated.

The question naturally arises here: if we had several instrumental variables available, would an appreciable improvement in the estimate be effected by using a combination of them and, if so, in what way? It is clear that if any instrumental variable is highly correlated with one of the investigational variables, x_{1t} or x_{2t} then the improvement to be expected from any combination of variables is slight. On the other hand if the correlation between members of the investigational and instrumental sets is of the order of, say, 0.7 it will be worth while to try to find an optimum combination. Let the instrumental sets be $x_{3t}^{(i)}$ ($i = 1, 2, \dots, k$; $t = 1, 2, \dots, n$) and let

$$x_{3t} = \sum_{i=1}^k c_i x_{3t}^{(i)} \tag{40}$$

Required to find the coefficients c_i so that the variance is unity and the covariance μ_{23} maximum, i.e.,

$$\sum_{i,j} v_{ij} c_i c_j = 1, \tag{41}$$

$$\mu_{23} = \sum_i c_i \lambda_i \text{ maximum} \tag{42}$$

where

$$v_{ij} = E x_{3t}^{(i)} x_{3t}^{(j)}, \quad \lambda_i = E x_{3t}^{(i)} x_{2t}. \tag{43}$$

In the usual way the solution is given by

$$\sum_j v_{ij} c_j = \beta \lambda_i, \tag{44}$$

where β is a Lagrange multiplier. Hence

$$c_j = \beta \sum_i v^{ij} \lambda_i, \tag{45}$$

which substituted in (41) gives β : so the c_i are known.

Consider the very simple case in which all the v_{ij} ($i \neq j$) are equal to v and all the λ_i to λ , both v and λ being positive. Suppose, further, that the variances of x_{2t} and $x_{3t}^{(i)}$ are all unity, so that λ and v are correlation coefficients. Then from (39) it will be seen to be advantageous in the large-sample case to take for instrumental variable a weighted mean of the individual variables $x_{3t}^{(i)}$ provided that

$$\begin{aligned} & \left(\sum_i \sum_j v_{ij} c_i c_j \right)^{\frac{1}{2}} / \sum_i c_i \lambda_i \\ & = \left(\sum_i c_i^2 + 2v \sum_{i < j} c_i c_j \right)^{\frac{1}{2}} / \lambda \sum_i c_i < \frac{1}{\lambda}. \end{aligned} \tag{46}$$

Obviously the c_i should be all taken as equal to give the minimum value, so that the inequality would become

$$\frac{k + vk(k-1)}{k^2} < 1, \tag{47}$$

which is always true provided $k > 1$ since $v < 1$. As an example suppose $k = 5$ and $v = 0.7$. Then the improvement effected by taking as instrumental variable an average of the five series, as compared with using any one of them, will be

$$(5 + 0.7 \times 20)^{\frac{1}{2}} / 5 = 0.87,$$

which is equivalent to an improvement in accuracy of 13 per cent as compared with using any one of them. Even if we had available an infinity of investigational sets the improvement would be only 1 to $v^{\frac{1}{2}}$, i.e., when $v = 0.7$ by 16 per cent. No matter what approach is made to this problem improved accuracy is difficult of attainment.

IV APPLICATION TO ECONOMIC TIME SERIES

The theory of relationship between statistics finds its most important application in economic time series and it is accordingly necessary to consider the suitability of the sampling model of Section III for dealing with these kinds of statistics. Formally the model is inappropriate. While we might take as the three-dimensional normal⁴ universe the three series indefinitely extended in time, i.e., as consisting of observations x_{it} ($i = 1, 2, 3$; $t = -N, -N + 1, \dots, N - 1, N$, where N is indefinitely large), for the theorem to apply the sample of n would have to be x_{it_j} ($j = 1, 2, \dots, n$), where the t_j are positive or negative integers selected at random. In practice this will hardly ever be the case since the sample will nearly always consist of a series of observations consecutive in time. What is wanted is a sampling theory appropriate to *sequences* of n , so that when we state that the probability is, say, $1/20$ that a differs from a given α (usually zero) by at least the amount found in the given sample, we mean that we should expect to find approximately a proportion of $1/20$ of such cases if the experiment were repeated a large number of times for sequences of n at different nonoverlapping parts of the indefinitely extended time series.

It will be shown, however, that for time series of moderate length the theorem in Section III can be applied with confidence and, as the practical application considered in Section V will make abundantly clear, the theorem, while applicable to samples of all sizes, will, in practice, yield useful results only when the samples are fairly large. It will be noted, in the first place, that a, given by (11), is symmetrical in the t , so that the order in which the sets of three (x_{1t}, x_{2t}, x_{3t}) are taken is immaterial — the sample sequence need not be envisaged as serially correlated. We are accordingly quite at liberty

4. The validity of the assumption of normality is discussed later.

to regard the particular sample as a random sample from *some* three-dimensional time universe, i.e., sequences indefinitely extended in time. The trouble is that the variance-covariance matrix $\|\mu_{ij}\|$ has to be estimated from the particular sample and we know that unless the sample series are very long (for example covering several periods if the series are periodic) the estimates of the matrix cannot be regarded as statistically consistent: the estimates of the variances in particular would usually be too low, if, for instance, the sample series covered only part of a period. In other words, different short sample sequences would yield estimates of the variance that for the given sample number would vary more widely than they should if computed from completely random samples all from the same universe.

Now it will have been seen, from (27), that the frequency distribution of a depends on that of

$$z = \frac{\mu_{33}(\mu_{22}a^2 - 2\mu_{12}a + \mu_{11})}{(\mu_{23}a - \mu_{13})^2}. \quad (48)$$

Set

$$\mu_{12} = \sqrt{\mu_{11}\mu_{22}} \rho_{12},$$

$$\mu_{13} = \sqrt{\mu_{11}\mu_{33}} \rho_{13},$$

$$\mu_{23} = \sqrt{\mu_{22}\mu_{33}} \rho_{23},$$

where the ρ_{ij} are coefficients of correlation. Then

$$z = (a^2 - 2\tau_{12}\rho_{12}a + \tau_{12}^2) / (\rho_{23}a - \rho_{13}\tau_{12})^2, \quad (49)$$

where $\tau_{12}^2 = \mu_{11}/\mu_{22}$. The population parameters are accordingly reduced to four, consisting of τ_{12} and the three correlation coefficients. While the estimates of μ_{11} and μ_{22} from the sample sequence will be biased it is plausible to assume that the estimate of their ratio τ_{12}^2 might be unbiased, particularly having regard to the fact that the system to be workable must be a highly correlated one. Nor does there appear to be any good reason why the three correlation-coefficient estimates should be biased. If it does no more, this aspect suggests that the theorem of Section III might be adapted to time series even if these are serially correlated.

Comparison of Simplest Time-Series Frequency with Frequency of Section III

The frequency distribution of the estimate a of the coefficient α on the assumptions of the previous section will now be compared with the frequency

distribution of the estimate from a time-series model, of the simplest stationary type. This will be termed *Model B* (in contra-distinction to *Model A* of Section III) and is as follows:

Three series of observations x_{it} ($i = 1, 2, 3$; $t = 1, 2, \dots, n$) are made at equal time intervals of which the investigational sets x_{1t} and x_{2t} are given by

$$\begin{aligned}x_{1t} &= x'_{1t} + x''_{1t}, \\x_{2t} &= x'_{2t} + x''_{2t},\end{aligned}\tag{50}$$

where the systematic parts x'_{1t} and x'_{2t} are connected by the exact relation

$$x'_{1t} = \alpha x'_{2t},$$

the same for all t . We also assume that

$$\sum_t x'_{1t} = 0 = \sum_t x'_{2t}.$$

The actual sample investigational series are then assumed to consist in systematic parts x'_{1t} and x'_{2t} fixed once for all not only in magnitude but in order, disturbed by x''_{1t} and x''_{2t} assumed to be normally distributed with means zero and variances μ''_{11} and μ''_{22} , independent of one another of the systematic parts, and with instrumental series x_{3t} (with $\sum_t x_{3t} = 0$) also regarded as fixed from sample to sample: this implies, of course, that the series x_{3t} is not a lagged investigational variable, a case considered later. The estimate a of α on Model B is then

$$a = \frac{X_1}{X_2},\tag{51}$$

where

$$X_1 = \sum_t x_{1t} x_{3t},$$

$$X_2 = \sum_t x_{2t} x_{3t}.$$

Clearly

$$\begin{aligned}EX_1 &= \sum_t x'_{1t} x_{3t} = \alpha \sum_t x'_{2t} x_{3t}, \\EX_2 &= \sum_t x'_{2t} x_{3t},\end{aligned}\tag{52}$$

so that a is a consistent estimate of α . The variances of X_1 and X_2 are

$$\begin{cases} \sigma_{x_1}^2 = \mu''_{11} \sum_t x_{3t}^2, \\ \sigma_{x_2}^2 = \mu''_{22} \sum_t x_{3t}^2, \end{cases} \quad (53)$$

and X_1 and X_2 are stochastically independent. Now a is the quotient of two normal variates X_1 and X_2 of which it may be assumed that the denominator X_2 is unlikely to assume negative values. Hence by Geary (1930)

$$u = \frac{aEX_2 - EX_1}{\sqrt{(\sigma_{x_2}^2 a^2 + \sigma_{x_1}^2)}} \quad (54)$$

is normally distributed with mean zero and variance unity for all sample sizes. This sampling model is so much more simple than Model A that it is natural to inquire the reason why it should not be used in preference to Model A in connection with the theory of linear relationship in time or otherwise. The answer is, of course, that the essential feature of the theory developed in this communication is that the error variances μ''_{11} and μ''_{22} are not known or cannot be efficiently estimated from the observations. We can, however, assume the error variances known for the purpose of assessing the reliability of Model A as applied to time series. In order to apply Model A formally the values of the variances and covariances required are given by:

$$\begin{aligned} n\mu_{11} &= n\mu''_{11} + \sum_t x'_{1t}{}^2, & n\mu_{13} &= \sum_t x'_{1t}x'_{3t} \\ n\mu_{12} &= \sum_t x'_{1t}x'_{2t}, & n\mu_{23} &= \sum_t x'_{2t}x'_{3t} \\ n\mu_{22} &= n\mu''_{22} + \sum_t x'_{2t}{}^2, & n\mu_{33} &= \sum_t x'_{3t}{}^2. \end{aligned} \quad (55)$$

We proceed as follows: Given sample size n , the variances and covariances, and a given probability (say 0.05), we find the confidence limits a_1 and a_2 of the estimate a using the Model A theorem, i.e., derived from (36). The values a_1 (or a_2) are substituted for a in (54) and the (normal) probability of this unit-variance deviation read off for comparison with the given probability (say 0.05). The results for five examples, with three sample sizes ($n = 10, 25, 120$) for each, are given in Table 1. The variances, covariances, and coefficient α for each example are shown at the head of the table. The examples are designed to illustrate the different kind of cases that can occur, in particular (i) different magnitudes of error variances, (ii) different correlations between instrumental variable and investigational variables, (iii) different values of α (which without loss of generality may be assumed not to exceed unity). Actually the units in which the investigational variables are measured are deemed to be such that the error variances, i.e., μ''_{11} and μ''_{22} , are each unity.

Table 1

<i>Example</i>	μ_{11}	μ_{12}	μ_{22}	μ_{13}	μ_{23}	μ_{33}	
I	6	10	21	8	16	15	1/2
II	3	2	3	5	5	10	1
III	3	2	3	5	5	20	1
IV	2	4	17	4	16	20	1/4
V	4	5	28/3	3	5	3	3/5

<i>Example</i>	<i>Sample size</i> (<i>n</i>)	<i>Lower 0.05</i> <i>prob. point</i> (<i>a</i> ₁) <i>on</i> <i>Model A</i>	<i>Model B</i>	
			<i>Normal deviate</i> [<i>u</i> (<i>a</i> ₁)]	<i>Probability</i>
I	10	0.3488	1.865	0.031
	25	.4097	1.726	.042
	120	.4596	1.661	.048
II	10	.5920	1.755	.040
	25	.7350	1.688	.046
	120	.8732	1.654	.049
III	10	.3868	2.022	.022
	25	.6262	1.771	.038
	120	.8234	1.670	.047
IV	10	.0823	1.891	.029
	25	.1520	1.733	.042
	120	.2067	1.663	.048
V	10	.3872	1.812	.035
	25	.4693	1.708	.044
	120	0.5405	1.658	0.049

Since, as shown later on, the variances on Model A and Model B tend to the same value when n tends towards infinity, it is to be expected that the two models would yield fairly similar results (i.e., would give much the same limits for the range of values of the estimate a of α corresponding to a given probability) for samples of moderate size. Table 1 shows that this is actually the case. Even for samples as small as 25 the probability of the lower limit a_1 (shown in the final column) is not very different from the probability 0.05. Since in all cases the Model B probability is below that of Model A (0.05) it is clear that the limits derived from the latter are on the "safe side". This also is to be expected since in applying Model B we assume more information than in applying the other model, namely that the error variances are known. In the table attention has been confined to the lower limit a_1 ; the upper limit a_2 would not give significantly different probabilities in the last column.

The table shows that not only does the use of an instrumental variable

highly correlated with the investigational variables yield more accurate estimates of the coefficient (as shown in the paper) but it results in more similar inferences from Models A and B. This is clearly seen by comparing examples II and III which are identical except for μ_{33} which is twice as large in III as in II. Relatively large error variances (as indicated by relatively small values of μ_{11} and μ_{22}) do not appear to render the results more discordant, i.e., in giving Model A probabilities very different from 0.05; this will be seen from example II in which the error variances are relatively large and yet the probabilities are nearest to 0.05 for all sample sizes.

It is emphasized that the two frequency distributions utilized in Models A and B are *exact*, assuming of course, that the conditions of the theorems are satisfied. The examples strongly suggest that the Model A approach yields sampling limits for estimates of α (corresponding to a given probability) which do not differ widely from those of the "time series" Model B, for samples of moderate size.

It is interesting to compare the quadratic inequalities yielding the limits from the two models. The Model A equation, which is derived from (36) by substituting $(\mu'_{11} + \mu''_{11})$ and $(\mu'_{22} + \mu''_{22})$ for μ_{11} and μ_{22} respectively, may be rewritten as follows:-

$$(\kappa\mu_{23}^2 - \mu'_{22}\mu_{33})(a - \alpha)^2 - \mu_{33}(a^2\mu''_{22} + \mu''_{11}) \leq 0, \quad (56)$$

whereas the Model B quadratic inequality, derived from (54) is

$$\kappa'\mu_{23}^2(a - \alpha)^2 - \mu_{33}(a^2\mu''_{22} + \mu''_{11}) \leq 0. \quad (54')$$

In (56)⁵ $\kappa = 1 + n/\tau^2$ where τ is the probability point from the t-distribution corresponding to a given probability and $\kappa' = n/\tau'^2$ where τ' is the normal probability point corresponding to the same probability. Since κ and κ' tend towards the same limit of order n when n tends towards infinity it is clear that the limits derived from the two equations must tend to be the same.

Comparison of Variances of Consistent Estimates of the Coefficient α

In the four following subsections large-sample approximations are placed on record of means and variances of estimates of the coefficient α using four different models or methods, including Models A and B already discussed. The general objective is to show that where the samples are moderately large, and the error or disturbance variances relatively small, the approximations to

5. Note n instead of $(n-1)$ in (35), since in this application the data are measured from the supposed known universal means, whereas in Section III the data are measured from sample means.

means and variances of a on the different assumptions do not differ much. Somewhat heuristically the inference is made that the frequency distribution appropriate to one is approximately applicable to all; in simple terms that the theory of Section III may be used with confidence in time series, unless the sequence is short. The approximations to the means and variances when the instrumental variable is a lagged observational variable will probably be found useful.

To translate the variance-covariance matrix $\|\mu_{ij}\|$ into time-series terms we take

$$\begin{aligned}\mu_{ij} &= E \frac{1}{n} \sum_t (x'_{it} + x''_{it})(x'_{jt} + x''_{jt}) \\ &= \delta_{ij} \mu''_{ii} + \frac{1}{n} \sum_t x'_{it} x'_{jt} \quad (i, j = 1, 2, 3),\end{aligned}$$

where observational variables are x_{1t} and x_{2t} and the instrumental variable is x_{3t} — the latter may be a lagged observational variable — the μ''_{ii} are the error or disturbance variances, and the systematic parts of the variables x'_{jt} ($j = 1, 2, 3$) are regarded as fixed once for all.

Case when Instrumental Variable is a Lagged (or Advanced) Investigational Variable: Model C

It will presently be shown that, while the sampling theory in this case is much more complicated than when the instrumental variable is not an investigational variable, the error will be slight if the simple theory appropriate to the latter case be assumed to apply formally to the former case, for samples of moderate size. We will, in fact, proceed to compute the approximate mean and variances of the estimate a of α for the two cases. Let

$$b = \frac{\frac{1}{n} \sum_{t=1}^n (X'_{1t} + X''_{1t})(X'_{2,t-1} + X''_{2,t-1})}{\frac{1}{n} \sum_{t=1}^n (X'_{2t} + X''_{2t})(X'_{2,t-1} + X''_{2,t-1})} \quad (57)$$

and

$$X'_{1t} = \beta X'_{2t} \quad (58)$$

exactly, where the systematic parts X'_{it} ($i = 1, 2$) are fixed once for all and stochastic variation is introduced only through the errors or disturbances X''_{it} , assumed normally distributed with means zero and variance finite, the same for all t .

Here and throughout the remainder of the section we assume that

$$\sum_t X'_{1t} = 0 = \sum_t X'_{2t},$$

i.e., that while the systematic parts, deemed fixed from sample to sample, are unknown, their sums are zero, effected in practice by measuring the observations from their means. By the transformations

$$\begin{aligned} x'_{it} &= X'_{it} / \sigma_{x'_{it}}, \\ x''_{it} &= X''_{it} / \sigma_{x''_{it}}, \end{aligned} \quad (i = 1, 2) \quad (59)$$

$$b = a\sigma_{x''_{1t}} / \sigma_{x''_{2t}}, \quad \beta = \alpha\sigma_{x'_{1t}} / \sigma'_{2t},$$

we find

$$a = \frac{\frac{1}{n} \sum_t (x'_{1t} + x''_{1t})(x'_{2,t-1} + x''_{2,t-1})}{\frac{1}{n} \sum_t (x'_{2t} + x''_{2t})(x'_{2,t-1} + x''_{2,t-1})} = \frac{\sigma_{x_{2t}}}{\sigma_{x'_{1t}}} b, \quad (60)$$

with

$$x'_{1t} = \alpha x'_{2t}, \quad (61)$$

and now with the error variances (of x'_{1t} and x''_{2t}) equal to unity. The numbers x'_{1t} (and x'_{2t} from (61)) are fixed from sample to sample and can assume any finite values whatever. From (60) it will be seen that the coefficient of variation (the ratio of the mean to the standard deviation) of a will be equal to that of b for different samples of n . To estimate large-sample approximations of mean and variance of a , set

$$\begin{aligned} \sum_t x'_{2t} x''_{2,t-1} &= \sqrt{n} u, & \sum_t x'_{1t} x''_{2,t-1} &= \sqrt{n} x, \\ \sum_t x'_{2,t-1} x''_{1t} &= \sqrt{n} v, & \sum_t x'_{2t} x''_{2,t-1} &= \sqrt{n} y, \\ \sum_t x'_{2,t-1} x''_{2t} &= \sqrt{n} w, \end{aligned} \quad (62)$$

whence

$$\begin{aligned} nEu^2 &= \sum_t x'^2_{2t} = n\mu'_{22} = n(\mu_{22} - 1), \\ nEv^2 &= nEw^2 = \sum_t x'^2_{2,t-1} = n\mu'_{33} = n(\mu_{33} - 1), \\ Ex^2 &= Ey^2 = 1, \\ nEuw &= \sum_t x'_{2t} x'_{2,t+2} = nv, \text{ say.} \end{aligned} \quad (63)$$

From (60) and (61)

$$\mu_{23} \sqrt{n} (a - \alpha) = \{v + x - \alpha(w + y)\} \cdot \left\{ 1 + \frac{1}{\mu_{23} \sqrt{n}} (u + w + y) \right\}^{-1}, \quad (64)$$

where

$$n\mu_{23} = \sum_t x'_{2t} x'_{2,t-1}. \quad (65)$$

Whence [noting that $\mu_{33} = \mu_{22} + 0(1/n)$]

$$\begin{aligned} n\mu_{23}^2 E(a - \alpha) &\doteq \mu_{22} + v)\alpha \\ n\mu_{23}^2 E(a - \alpha)^2 &\doteq \mu_{33} (1 + \alpha^2) + \frac{1}{n\mu_{23}^2} \left\{ 2(3\mu_{22}^2 - 4\mu_{23}^2)(1 + 2\alpha^2) - 3\mu_{22}(1 + \alpha^2) \right\} \\ &\quad + \frac{6v}{n\mu_{23}^2} \left\{ \mu_{22}(1 + 3\alpha^2) + v\alpha^2 \right\}. \end{aligned} \quad (66)$$

These formulae give $E(a - \alpha)$ correct to $O(n^{-1})$ and $E(a - \alpha)^2$ to $O(n^{-2})$. Hence $\text{var}(a)$ is derivable to $O(n^{-2})$. Using (59) there will be no difficulty about finding means (b) and $\text{var}(b)$ in terms of the covariances and variances of the original variables $X_{it} = X'_{it} + X''_{it}$ ($i = 1, 2$) and of the error variances $\text{var}(X''_{it})$.

Instrumental Variable not a Displaced Investigational Variable: Model D

In this case in (57) we have X'_{3t} and X''_{3t} instead of $X'_{2,t-1}$ and $X''_{2,t-1}$ respectively, where the observed instrumental variable $X_{3t} = X'_{3t} + X''_{3t}$. The disturbances X''_{1t} , X''_{2t} , X''_{3t} are completely independent of one another. By analogous transformations we find instead of (60)

$$a = \frac{\frac{1}{n} \sum_t (x'_{1t} + x''_{1t})(x'_{3t} + x''_{3t})}{\frac{1}{n} \sum_t (x'_{2t} + x''_{2t})(x'_{3t} + x''_{3t})} = \frac{\sigma_{x''_{2t}}}{\sigma_{x''_{1t}}} b, \quad (67)$$

with $x'_{1t} \doteq \alpha x'_{2t}$ and where, as before, the error variances, i.e., of x''_{1t} , x''_{2t} , and x''_{3t} , are now unity and their means zero. We readily find

$$n\mu_{23}^2 E(a - \alpha) \doteq \mu_{33} \alpha, \quad (68)$$

$$n\mu_{23}^2 E(a - \alpha)^2 \doteq (1 + \alpha^2) \left\{ \mu_{33} - \frac{4}{n} + \frac{3\mu_{33}(\mu_{22} - 1)}{n\mu_{23}^2} \right\} + \frac{3\mu_{33}^2(1 + 3\alpha^2)}{n\mu_{23}^2}. \quad (69)$$

Displacement Effects

Suppose that we substitute formally in (68) and (69) the variances and covariances that would be found if $x_{2,t-1}$ were used as instrumental variable instead of x_{3t} . Denote by X_1 and X_2 the resulting pseudo-values of $E(a - \alpha)$ and $E(a - \alpha)^2$. Then

$$n\mu_{23}^2 X_1 = \mu_{22}\alpha, \quad (70)$$

$$n\mu_{23}^2 X_2 = (1 + \alpha^2) \left(\mu_{33} - \frac{4}{n} \right) + \frac{3}{n\mu_{23}^2} \left\{ 2\mu_{22}^2 (1 + 2\alpha^2) - \mu_{22} (1 + \alpha^2) \right\}. \quad (71)$$

Then from (65), (66), (70), and (71),

$$n\mu_{23}^2 \{E(a - \alpha) - X_1\} = v\alpha, \quad (72)$$

$$n^2 \mu_{23}^4 \{E(a - \alpha)^2 - X_2\} = 2(1 + 3\alpha^2)(3v\mu_{22} - 2\mu_{23}^2) + 6v^2\alpha^2. \quad (73)$$

Formulae (72) and (73) indicate the approximate effect of using a lagged investigational variable as an instrumental variable: the expressions on the right may be regarded as the "displacement effects". To form a more precise idea of their magnitude, set

$$\mu_{23} = \rho_1 \mu_{22},$$

$$v = \rho_2 \mu_{22},$$

where ρ_1 and ρ_2 are serial correlations lagged 1 and 2 respectively. Then

$$n\rho_1^2 \mu_{22} \{E(a - \alpha) - X_1\} = \rho_2 \alpha, \quad (72')$$

$$n^2 \rho_1^4 \mu_{22}^2 \{E(a - \alpha)^2 - X_2\} = 2(1 + 3\alpha^2)(3\rho_2 - 2\rho_1^2) + 6\rho_2^2 \alpha^2. \quad (73')$$

Usually ρ_1 is about 0.9 and ρ_2 about 0.7. Substituting these values tentatively we find

$$n\mu_{22}^2 \{E(a - \alpha) - X_1\} \sim 0.9\alpha, \quad (72'')$$

$$n^2 \mu_{22}^2 \{E(a - \alpha)^2 - X_2\} \sim 1.5 + 8.9 \alpha^2. \quad (73'')$$

The value of α may be assumed to be at most unity and the value of μ_{22} will, of course, depend on the error variance since it is expressed in units of the

error standard deviation: in fact the smaller the error the larger will be μ_{22} . It is clear that for samples of moderate size no great distortion is introduced into the probabilistic inferences by using a displaced investigational variable as the instrumental variable and treating it exactly as if its error constituent were independent of those of the investigational variables.

Model B Approximation

It will be necessary also to consider the large-sample approximations to the mean and variables for Model B, the exact sampling distribution of which was discussed above, and which, in particular, was shown to be very close to that for Model A except for very small samples (or short sequences). These will clearly be a special case of Model D: that in which the variance of x_{3t}'' is zero. The formulae are as follows, when $\mu_{11}'' = \mu_{22}'' = 1$:

$$n\mu_{23}^2 E(a - \alpha) = \mu_{33}\alpha, \quad (74)$$

$$n\mu_{23}^2 E(a - \alpha)^2 = (1 + \alpha^2)\mu_{33} + \frac{3\mu_{33}^2}{n\mu_{23}^2} (1 + 3\alpha^2). \quad (75)$$

Model A Approximation

Finally we require Model A approximations of mean and variance. These are derivable from

$$n\mu_{23}^2 E(a - \alpha) = \mu_{33}\alpha, \quad (76)$$

$$n\mu_{23}^2 E(a - \alpha)^2 = (1 + \alpha^2) \left\{ \mu_{33} - \frac{4}{n} \mu_{33} + \frac{3\mu_{33}}{n\mu_{23}^2} (\mu_{22}\mu_{33} + \mu_{23}^2) \right\} + \frac{6\alpha^2 \mu_{33}^2}{n\mu_{23}^2}, \quad (77)$$

where the error variances μ_{11}'' and μ_{22}'' are taken as unity. The expressions on the right-hand side simplify somewhat on taking

$$\mu_{23}^2 = \rho^2 \mu_{22} \mu_{33},$$

where ρ is the population coefficient of correlation between the observations x_{1t} and x_{2t} . We then find

$$n\rho^2 \mu_{22} E(a - \alpha) = \alpha, \quad (78)$$

$$n\rho^2 \mu_{22} E(a - \alpha)^2 = (1 + \alpha^2) \left\{ 1 - \frac{4}{n} + \frac{3(1 + \rho^2)}{n\rho^2} \right\} + \frac{6\alpha^2}{n\rho^2 \mu_{22}}. \quad (79)$$

The Assumption of Normality in Time Series

An additional formidable objection to the application to time series of the theory of Section III would appear to lie in the assumption of population normality in the case of such series. In fact the linear trend must have a rectangular distribution and a sinusoidal approximation of the systematic part

$$x'_t = \sum_{i=1}^k A_i \sin(\alpha_i t + \beta_i) \quad (t = 1, 2, \dots, n), \quad (80)$$

where the A_i and the α_i may be assumed all different, has the following moments when n is large:

$$\begin{cases} \mu_0 = 1 & \mu_1 = 0 = \mu_3, \\ \mu_2 = \frac{1}{2} \sum_i A_i^2, & \mu_4 = \frac{3}{8} \sum_i A_i^4 + \frac{3}{2} \sum_{i>j} A_i^2 A_j^2. \end{cases} \quad (81)$$

To the extent to which this system represents time series there is no reason why $\beta_2 = \mu_4/\mu_2^2$ should tend to its normal value 3: in fact, if one of the A_i is much greater than all the others, β_2 will not be very different from 3/2. The inclusion of disturbances would, of course, tend generally to give the sample more of a "normal look". In the actual case of USA economic data during the 17 years 1922-1938 (used in an application later), Kuznets' and Barger's⁶ data for employees' compensation yield a value of 0.8156 for the test of normality α (Geary, 1936), which is not to be confused with the coefficient a . This is practically identical with the normal value. Since most of the series of USA economic data are highly correlated, no doubt much the same result would be found from other data during this period of years.

Conclusion as to Application of the Theory in Section II to Time Series

We have shown that for time sequences of moderate length (e.g., 50):

- (i) The Model A distribution yields a frequency distribution similar to the simplest time-series model, termed Model B;
- (ii) all models yield consistent estimates a of α ;
- (iii) to $O(n^{-1})$ all "time models" give the same expression for the variance of a as does Model A; and study of terms in n^{-2} in $E(A - \alpha)^2$ will show that the contributions of these terms is small relatively to the term in n^{-1} unless the error variances are substantial, in which case no theory will yield efficient estimates of a ; in algebraic form, however, the terms in n^{-2} are very dissimilar.

6. From H. Barger, 1942, *Outlay and Income in the United States 1921-38*.

We accordingly feel justified, if on somewhat empirical grounds, in submitting that the theory in Section III can be used with confidence in connection with time series. The great advantages in using Model A are:

(i) given the variance-covariance matrix the frequency distribution is exact, and the variance-covariance matrix can be estimated consistently from the observations;

(ii) knowledge of the error variances is not required, as is the case with all the other models discussed in this section.

It is assumed throughout this communication that the error or disturbance variances cannot be efficiently estimated from the observations. The emphasis is on the word "efficiently". It is easy to show that, given the coefficient α and μ_{12} , both of which parameters can be consistently estimated from the observations, the error variances μ''_{11} and μ''_{22} can *formally* be estimated. In fact

$$\mu''_{11} = \mu_{11} - \mu'_{11} = \mu_{11} - \alpha\mu'_{12} = \mu_{11} - \alpha\mu_{12},$$

and similarly for μ''_{22} . The trouble is that, as the application in Section V will make abundantly clear, the sampling range of estimates of α is very wide even in a highly correlated system, when the sample is not large; furthermore, the error variances in such a highly correlated system must be small and the estimates of μ_{11} and μ_{12} are themselves substantially subject to error. The estimates of the error variances must, in consequence, be deemed worthless — it can obviously happen, for instance, that the "estimates" yield negative values for the variances! — unless the samples are very large (or the time sequences are very long). When confidence can be reposed in the estimates of the error variances, the appropriate formulae for mean and variance of the estimate a of α , given in this section, can be used.

V AN APPLICATION CONSIDERED

In Richard Stone's paper (1947) on "The Interdependence of Blocks of Transactions" is presented a series of calculations of variances and covariances of 17 sets of USA economic data for the years 1922-1938, prepared by H. Barger and S. Kuznets (1942). Suppose that we are trying to determine whether a linear relation subsists between the systematic parts of variable 1 "employees' compensation" and variable 2 "consumers' perishable goods plus producers' durable goods". Admittedly variable 2 is somewhat artificial in content but (unfortunately) the purpose of the calculation presented here must be regarded as primarily arithmetical and of little economic significance, for the present theory requires observations more numerous and of a different character to be effective.

Table 2: *Estimated Value of the Coefficient in the Relation $x'_{1t} = ax'_{2t}$ Using as Estimates $a = \sum x_{1t}x_{it} / \sum x_{2t}x_{it}$, ($i = 3, 4, \dots, 17$), All Variables Measured from Means*

Instrumental Variable	Estimate of α	Correlation Coefficient	
		With 1	With 2
3	1.2539	0.59	0.66
4	1.4946	.92	.87
5	1.5910	.93	.82
6	1.6466	.78	.66
7	1.7296	-.63	-.51
8	1.2731	.62	.68
9	2.1415	.40	.26
10	1.7372	.64	.52
11	1.5896	.91	.80
12	1.4492	.91	.88
13	1.5510	-.94	-.85
14	16.1919	.13	.01
15	1.2669	-.48	-.53
16	1.4289	-.78	-.77
17	5.0503	0.17	0.05

Variables 1 and 2 constitute the investigational set. For the instrumental set we have no fewer than 15 series which Stone (*op. cit.*, p. 11) numbers 3 to 17: they need not be particularized here. In his Table II Stone gives the complete variance-covariance matrix for the 17 variables, as well as the correlation coefficients (Table III). From these tables Table 2 has been compiled without difficulty.

The correlation between variables 1 and 2 is very high, namely 0.97. Since instrumental variable 4 is (practically) the variable most highly correlated with variables 1 and 2 we use it to determine the estimate of a . We find $a = 1.4946$. To a probability of 1/10 (not to adopt too high a standard) the sampling limits (using 2.26) are found to be

$$1.3337 < a < 1.7378,$$

which are far wider than the regression limits given by

(i) $x_1 = 1.3543 x_2$ (x_1 on x_2),

(ii) $x_1 = 1.4491 x_2$ (x_2 on x_1).

Logically we must in this case adopt the regression lines as absolute limits, since for the derivation of the sampling limits of a , we had to assume known

the variance-covariance matrix of x_{1t} , x_{2t} , x_{3t} as determined by the sample: if, in particular, μ_{11} , μ_{12} , and μ_{22} are known, then from Frisch's theorem (1934) we must take the regression limits as the absolute limits of a . At the same time we must recognize the arbitrary nature of the assumption that for so small a sample as 17 the variance-covariance matrix should be regarded as given by the data. Using M.S. Bartlett's theorem (1933) as to the distribution of the Studentized regression coefficient, we find that the regression coefficient (ii) given as 1.4491 might (on probability 1/10) have come from a population with this coefficient ranging from 1.294 to 1.646, so that a high correlation is no guarantee of regression-coefficient stability when the sample is small.

Figure 1 based on Table 2 shows clearly that the higher the instrumental correlation, the more the estimates tend to cluster around the "true" figure which is probably in the neighbourhood of 1.4-1.5. When the correlations are insignificant the estimates are fantastic. When two instrumental variables are close together on the diagram, or even when they tend to give much the same estimate of a , usually we find them highly correlated (from Stone's table). Thus between 7 and 10 the correlation is -0.90 , and 0.96 between 6 and 10, 0.98 between 5 and 11, -0.97 between 5 and 13. This phenomenon is a reminder that, while close similarity in two or more estimates of a based on different instrumental variables may normally be regarded as good evidence that the relation is complete (in a sense to be defined in the next section) we should be chary about accepting it if the instrumental variables are highly correlated.

As far as the test goes, it does not contradict the hypothesis that the relation $x'_{1t} = \alpha x'_{2t}$ between the systematic parts of the variables is complete. The test is, however, insensitive for so small a sample.

VI DETERMINATION AND ASSESSMENT OF ACCURACY OF COEFFICIENTS IN EQUATIONS IN THE SYSTEMATIC PARTS OF MORE THAN TWO VARIABLES

In applying the method described in Section II, as well diversified a system of instrumental variables x_{it} as possible should be used, because if one uses two very highly correlated variables in the instrumental set, two equations in the a_i will be produced with covariant coefficients nearly proportional, so that, in effect, the rank of the $p \times (p - 1)$ matrix of covariances is less than $(p - 1)$. This circumstance will often preclude the use of both a time series and the same series lagged one interval, especially when the series is fairly smooth.

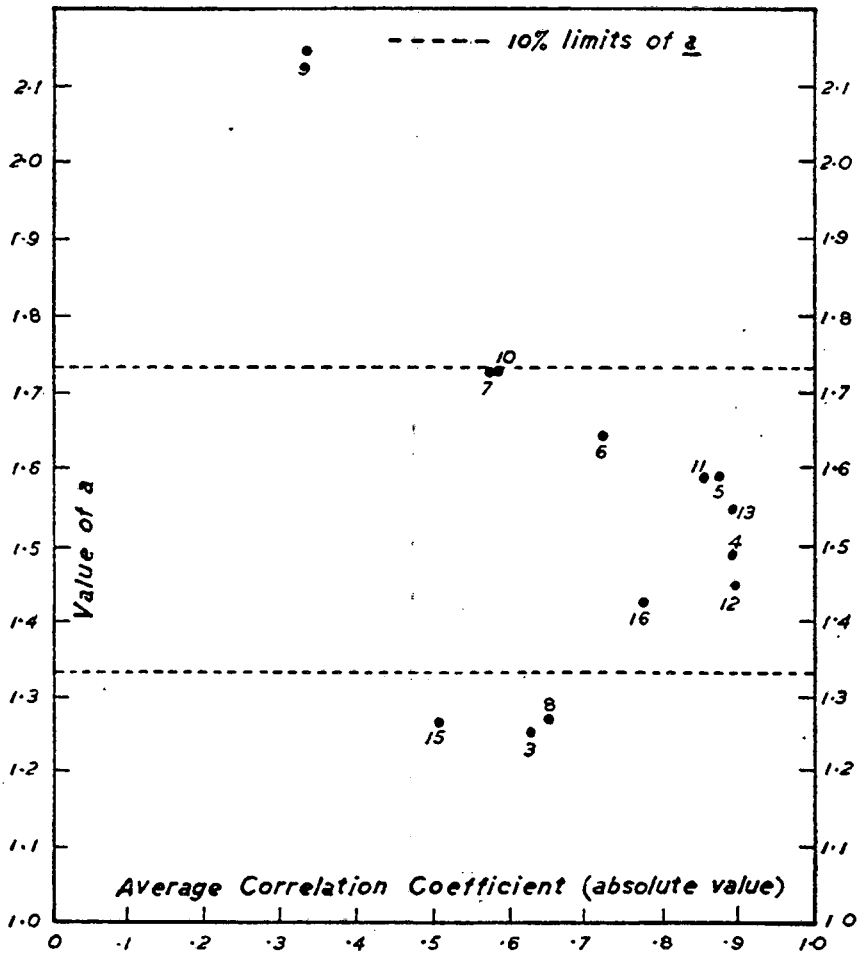


Figure 1: Estimates of Coefficient a of α Graphed against Average Correlation of Instrumental Variable with the Two Investigational Variables.

Note: Numbers indicate instrumental variables as shown in Table 2.

In the equation the number of variables may be

- (i) just enough, when the equation is said to be *complete*;
- (ii) more than sufficient, when the equation is *overdetermined*;
- (iii) too few, when the equation is said to be *incomplete*.

For case (i) the author in 1943 proposed as a test the determination of an equation additional to the $(p - 1)$ in (49) by means of another instrumental variable, computing the p th-order determinant of the covariances and attempting to assess whether it was significantly different from zero by

reference to its sampling standard deviation, the formula for which for large samples was given. It must be confessed in practice that the method has proved awkward for calculation and insensitive for inference, though in greater or lesser degree insensitivity bedevils most tests of significance of economic time series.

An alternative method that seems more likely to yield satisfactory results is the following. Having estimated the coefficients a_i from (10), set

$$y_t = \sum_{i=1}^p \alpha_i x_{it}, \quad (82)$$

taking the calculated a_i as estimates (proportionately) of the α_i . The systematic parts and the disturbances of y_t are given by

$$\begin{aligned} y'_t &= \sum_i \alpha_i x'_{it}, \\ y''_t &= \sum_i \alpha_i x''_{it}, \end{aligned} \quad (83)$$

though these are, of course, unknown separately. Let z_t be an investigational variable not included in the set x_{it} ($i = 1, 2, \dots, p$). Write

$$y'_t = \beta z'_t \quad (84)$$

and let w_t be an instrumental variable not included in the original sets. We are now in exactly the situation of previous sections for determining whether the coefficient β is significantly different from zero. The disturbances of y_t , z_t , w_t may be deemed independent in the manner required and the necessary variance-covariance matrix of (y_t, z_t, w_t) can be estimated just as was that of (x_{1t}, x_{2t}, x_{3t}) in the previous section. If the calculation (repeated if possible a few times using *new* functions z_t and w_t , each time) persists in showing that the estimate b of β is not significant, then the original relation (2) may be deemed complete. This method could be used for testing the validity of *structural* equations of given form.

Case (ii), that of *overdetermination*, will be indicated by small values of the coefficients of one or more of the variables. Having purged the equation of these doubtful variables, one proceeds exactly as in case (i) remembering, however, always to use new variables, of which one must assume an adequate supply of the right kind.

In case (iii), that of *incompleteness*, one will find the coefficient b of case (i) significantly different from zero for one variable, the systematic part of which

variable is then added to the relation with the coefficient as determined and the process of testing the completeness of the new relation is repeated. Care must be taken to use, when required, new variables for the instrumental set.

When the complete equation has been determined the sampling limits of each coefficient estimate may be found as follows, e.g., for a_p/a_1 . Write the relation in the form

$$\frac{\alpha_p x'_{pt}}{\alpha_1} + u'_t = 0 \tag{85}$$

where

$$u'_t = x'_{1t} + \frac{\alpha_2}{\alpha_1} x'_{2t} + \dots + \frac{\alpha_{p-1}}{\alpha_1} x'_{p-1,t} \tag{86}$$

and find the limits of a_p/a_1 using (36), again taking the computed a_i/a_1 as the values of α_i/α_1 , $i = 2, 3, \dots, p - 1$

It should be remarked that the test of significance proposed in this section for the several-variable case is exact only when the coefficient estimates are (proportionately) exactly equal to the α_i . Actually, as we have seen, the estimates are subject to wide sampling deviations unless the samples are very large; nevertheless they are consistent estimates which tend in probability towards the population values, and, while inferences as to significance may be wrong on account of sampling errors of estimate of the coefficients, they will be right in the long run.

We have seen that in the case of two variables the accuracy of the estimate of the coefficient depends largely on the correlation of the instrumental variable x_{3t} with the investigational variable x_{2t} . What is the corresponding property when the number of investigational variables exceeds two? From (10) it will be seen that the estimate a of the ratio of any two coefficients a_i may be expressed in the form

$$a = r''/r', \tag{87}$$

where the elements of the two determinants r' and r'' are the covariances m_{ir} . In the large-sample case and with some other wide assumptions it has been shown (Geary, 1943) that the confidence limits of a corresponding to a probability are given approximately by the roots of the quadratic equation (in a) (cf. (56)):

$$\lambda^2 (g_2 a^2 - 2g_1 a + g_0) = (\lambda^2 \overline{p-1^2} + n)(\rho' a - \rho'')^2, \tag{88}$$

where ρ' and ρ'' are the determinants r' and r'' when n is indefinitely large, λ

the normal probability point corresponding to probability π (e.g., when $\pi = 0.05$, $\lambda = 1.96$), g_0, g_1, g_2 homogeneous quadratic expressions in the first minors of ρ' and ρ'' of form specified in the original paper.

It is clear from (88) that generally the larger the value of ρ' the closer the values of the roots of the quadratic equation, and the smaller the value of ρ' the more dispersed these values are. Accordingly the instrumental set should be so well diversified as to give ρ' the greatest possible value. This is why, as remarked above, the estimates found using as investigational sets a particular series and the same series lagged or advanced say one term, though these may generally be highly correlated with the investigational set, may yield inefficient estimates of the coefficient a_i , because they may give two closely similar lines in the $p \times (p-1)$ matrix m_{ir} .

VII CONCLUSION

The method here outlined will certainly furnish consistent estimates of the coefficients of relations between systematic parts and determine their sampling limits, and provide the number of sets of observations, investigational and instrumental, if the number of observations is large enough. Classical regression theory on which confluence analysis so largely depends, will only afford consistent estimates in the almost trivial case of no disturbance or in which so many variables have been introduced into the equation (in relation to the number of observations) that the fit of the plane to the observations is very close (as indicated by the multiple-correlation coefficient). This is emphatically *not* to say that confluence analysis has not a value for determining just the set of variables constituting a complete set. Some progress has been made recently towards constructing large charts for omnibus use in confluence analysis designed to reduce, if not largely to eliminate, the tedium of making a "tilling" anew for each investigation.

Actually all the familiar difficulties of collinearity, etc., encountered in confluence analysis arise in the technique here discussed, but they have an entirely different character. They can all be resolved if the variables and observations are numerous enough, whereas in classical regression analysis, no matter how many observations there are, the estimates of the coefficient are biased.

It only remains now to build up a set of applications of the theory to test its practical efficacy!

The author wishes particularly to thank Mr Olav Reiersøl for constructive criticism which led to an extension in the paper as originally drafted.

REFERENCES

- BARTLETT, M.S., 1933. "On the Theory of Statistical Regression", *Proceedings of the Royal Society of Edinburgh*, Vol. 53, p. 260.
- CRAMÉR, H., 1937. *Random Variables and Probability Distributions*, Cambridge Tracts in Mathematics, No. 36, Cambridge.
- FRISCH, R., 1934. *Statistical Confluence Analysis by Means of Complete Regression Systems*, Universitetets Økonomiske Institutt, Oslo.
- GEARY, R.C., 1930. "The Frequency Distribution of the Coefficients of Two Normal Variates", *Journal of the Royal Statistical Society*, Vol. 93, Part III, pp. 442-446.
- GEARY, R.C., 1936. "Moments of the Ratio of the Mean Deviation to the Standard Deviation for Normal Samples", *Biometrika*, Vol. 28, Parts III and IV, pp. 295-305.
- GEARY, R.C., 1942. "Inherent Relations between Random Variables", *Proceedings of the Royal Irish Academy*, Vol. 47, A, 6, pp. 63-76.
- GEARY, R.C., 1943. "Relations between Statistics: the General and the Sampling Problems when the Samples Are Large", *Proceedings of the Royal Irish Academy*, Vol. 49, A, 10, pp. 177-196.
- GEARY, R.C., 1948. "Studies in Relations between Economic Time Series", *Journal of the Royal Statistical Society*, Series B (Methodological), Vol. 10, No. 1, pp. 140-158.
- REIERSØL, O., 1941. "Confluence Analysis by Means of Lag Moments and Other Methods of Confluence Analysis", *Econometrica*, Vol. 9, No. 1, pp. 1-24.
- REIERSØL, O., 1945. *Confluence Analysis by Means of Instrumental Sets of Variables*, Uppsala.
- STONE, R., 1947. "On the Interdependence of Blocks of Transactions", *Supplement to the Journal of the Royal Statistical Society*, Vol. 9, No. 1, pp. 1-45.
- TINTNER, G., 1945. "A Note on Rank, Multicollinearity and Multiple Regression", *Annals of Mathematical Statistics*, Vol. 16, pp. 304-308.