Econometrica 26 (1958) 393-415 *North-Holland Publishing Co. Amsterdam*

THE ESTIMATION OF ECONOMIC RELATIONSHIPS USING INSTRUMENTAL VARIABLES

BY J. D. SARGAN

1. INTRODUCTION

THE USE OF INSTRUMENTAL variables was first suggested by Reiersøl [**13, 14**] for the case in which economic variables subject to exact relationships are affected by random disturbances or measurement errors. It has since been discussed for the same purpose by several authors, notably by Geary [**9**] and Durbin [**7**]. In this article the method is applied to a more general case in which the relationships are not exact, so that a set of ideal economic variables is assumed to be generated by a set of dynamic stochastic relationships, as in Koopmans [**12**], and the actual economic time series are assumed to differ from the ideal economic variables because of random disturbances or measurement errors. The asymptotic error variance matrix for the coefficients of one of the relationships is obtained in the case in which these relationships are estimated using instrumental variables. With this variance matrix we are able to discuss the problem of choice that arises when there are more instrumental variables available than the minimum number required to enable the method to be used. A method of estimation is derived which involves a characteristic equation already considered by Hotelling in defining the canonical correlation [**10**]. This method was previously suggested by Durbin [**7**].

The same estimates would be obtained by the maximum-likelihood limited-information method if all the predetermined variables which are assumed subject to disturbances or errors were treated as if they were jointly determined, and the instrumental variables treated as if they were predetermined variables. Such a procedure was suggested by Chernoff and Rubin [**5**]. It is possible to use the smallest roots of the characteristic equation for significance tests in exactly the same way as when using the maximum-likelihood method, and similar confidence regions can be defined.

All the results listed so far depend on the use of asymptotic approximations. A few calculations were made by the author on the order of magnitude of the errors involved in this approximation. They were found to be proportional to the number of instrumental variables, so that, if the asymptotic approximations are to be used, this number must be small.

2. THE STRUCTURE OF RANDOM SHOCKS AND DISTURBANCES

This article is concerned with a model in which there exist both disturbances with properties first outlined by Frisch [**8**] and random shocks as in the models used by Koopmans and others [**12**]. The only problem which will be considered is that of determining the coefficients of a single relationship.

The actual time series with which the economist is concerned are represented by x_{it}, $i = 1 \ldots n$, $t = 1 \ldots T$, and are assumed to be of the form

$$x_{it} = x'_{it} + x''_{it} \tag{2.1}$$

where x'_{it} is the systematic part of the variable and x''_{it} accounts for the measurement error and the random disturbances.

The x'_{it}, except for a few which have no measurement error (the constant term, the trend and the seasonal components), are either exogenous variables, or, if endogenous, are assumed to have been generated by a stochastic model of the kind considered by Koopmans [**12**]. The relationship under consideration will then be written

$$\sum_{i=1}^{n} a_i x'_{it} = \varepsilon_t \tag{2.2}$$

where ε_t, the random shock, is assumed to be independent of the systematic parts of all the predetermined variables.

We now assume that there are some predetermined variables whose measurement errors are independent of the measurement errors of all the variables in the relationship, and of the random shock. This requirement excludes certain categories of predetermined variables: those in the relationship (unless they have no measurement error), lagged values of variables in the relationship (unless one makes the unrealistic assumption that measurement errors are not autocorrelated), and any predetermined variable which is estimated from the same data as one of the variables in the relationship. Thus, it is necessary that the sources of data used for estimating the instrumental variables should be largely independent of those used to estimate the variables in the relationship. The instrumental variables will be denoted by u_{jt}, $j = 1, \ldots, N$.

If

$$\sum_{i=1}^{n} a_i x_{it} = E_t \tag{2.3}$$

it follows from (2.1) and (2.2) that

$$E_t = \varepsilon_t + \sum_{i=1}^{n} a_i x''_{it}. \tag{2.4}$$

E_t will be called the residual, and it follows from (2.4) and the discussion above that it is independent of all the instrumental variables. The notation used takes no account of the fact that some of the instrumental variables may be in the relationship. As noted above such variables must have zero measurement error, and this restricts them in practice mainly to the constant term or trend or seasonal factors.

In this notation, then, Reiersøl's method amounts to positing a zero sample covariance between the residual and each instrumental variable. One therefore obtains the following equations:

$$\frac{1}{T}\sum_{t=1}^{T} E_t u_{jt} = 0, \qquad (j = 1, \ldots, N), \tag{2.5}$$

or

$$\sum_{i=1}^{n} a_i \left(\frac{1}{T} \sum_{t=1}^{T} x_{it} u_{jt} \right) = 0, \qquad (j = 1, \ldots, N).$$

Equations (2.5) provide N equations for the n—1 ratios of the coefficients, so that if $N = n$—1 they give a unique set of estimates of the coefficients a_i.

The interpretation of the "ideal economic variable" and of the "disturbance" is a little difficult. Probably the simplest interpretation is to suppose that, if the data were available for sufficiently short periods, it would be possible to use a model in which each equation explains how some class of economic agents determine the value of one variable as a function of the previous values of other variables. Then it will be assumed that the "ideal economic variables" are the actual variables to which the economic agents react. At the same time a random component or shock is used in the model equation to represent all those factors which affect the variable being determined by the relationship, including factors which are not easy to measure or to represent numerically, factors which have individually such a small effect that they are not worth attempting to measure, and a catch-all factor providing for the economist's ignorance of human relationships and other social, institutional, and technological factors. The random component is treated like a random variable and is assumed to be independent of the other variables (the lagged variables) in the relationship.

It is not necessarily true that the determined variable is also an ideal economic variable in the sense that it is exactly equal to the variable to which some other economic agent later reacts, or that if an economic variable appears as a cause in two different equations the appropriate values of the ideal economic variable are the same. It will be assumed that if there are differences they can be absorbed into the random shock, or equation error.

Owing to lack of data and the need to simplify the estimation problem this ideal model must be replaced by a simpler model in which short lags must be ignored or approximated by distributed lags. The ideal economic variables of this simpler model correspond to the ideal economic variables of the ideal model. The disturbances or non-systematic parts of the variables, being the differences between the actual times series and the ideal economic variables, can be regarded as really measurement errors perhaps partly due to differences in definition. It is not easy to justify the basic assumption concerning these errors, namely, that they are independent of the instrumental variables. It seems likely that they will vary with a trend and with the trade cycle. In so far as this is true the method discussed here will lead to biased estimates of the coefficients. Nothing can be done about this since presumably, if anything were known about this type of error, better estimates of the variables could be produced. It must be hoped that the estimates of the variables are sufficiently accurate, so that systematic errors of this kind are small.

In any case it will be noted that the method of this article uses the minimum assumption about the measurement error. There is no need to assume, for example, that the errors on the variables in the equation are independent of each other, and the estimated coefficients are still consistent even if the errors are autocorrelated.

Throughout this paper it will be assumed that the E_t are not autocorrelated.

3. THE ASYMPTOTIC COVARIANCE MATRIX

The main results of this article will be concerned with the asymptotic properties of the estimates. The following general notation will be used. The sample covariance matrix of the variables in the relationship will be denoted by M_{xx}; the sample covariance matrix of the instrumental variables will be denoted by M_{uu}; and the sample covariance matrix between a variable in the relationship and an instrumental variable will be denoted by M_{xu}.

The asymptotic limit of a sample function of variables will be denoted by placing a bar over the corresponding symbol. Thus $\bar{M}_{xu}$, representing the probability limit of M_{xu}, will be equal to the stationary limit of $\mathscr{E}\,(x_t' u_t)$ providing that the model is not explosive, and similarly for the other covariance matrices.

Considering first the case in which $N = n - 1$ it is convenient to introduce the functions

$$w_j = \frac{1}{\sqrt{T}} \sum_{r=1}^{T} E_t u_{jt}, \qquad (j = 1, \ldots, N).$$

For finite T, the w_j are not usually normally distributed. But their asymptotic joint distribution is normal. From a direct expansion of the w_j and w_k and the fact that E_t is independent of $u_{jt'}$, for $t \geqslant t'$ and of $E_{t'}$, for $t \neq t'$, it follows that

$$\mathscr{E}\,(w_j w_k) = \mathscr{E}\left(\frac{1}{T}\sum_{t=1}^{T} u_{jt}u_{kt}\right)\sigma^2 = \mathscr{E}\,(M_{uu})_{jk}\,\sigma^2$$

with

$$\sigma^2 = \mathscr{E}\,(E_t{}^2)$$

so that

$$\mathscr{E}(\bar{w}_j\bar{w}_k) = (\bar{M}_{uu})_{jk}\sigma^2.$$

We adopt the standardization that the mth parameter (an arbitrary parameter) be equal to unity. We then denote by a_m the vector of the n—1 parameters which are not assumed equal to 1, by M_{xum} the $N \times N$ matrix M_{xu} with its mth row missing and by q_m the vector whose jth component is equal to

$$\frac{1}{T}\sum_{t=1}^{T} x_{mt}u_{jt}.$$

Then the estimate derived from (2.5) can be written

$$\hat{a}_m = -q_m\,M_{xum}^{-1}$$

where the $\hat{a}_m$ is obtained from the estimated vector by the omission of the unit in the mth position.

Now,

$$\sum_{i=1}^{n} a_i\left(\sum_{t=1}^{T} x_{it}u_{jt}\right) = \sum_{t=1}^{T} E_t u_{jt} = w_j\sqrt{T},$$

and this can be written in terms of the previously defined vectors

$$\sqrt{T}(a_m M_{xum} + q_m) = w$$

where w is the vector with components equal to w_j.

From this it follows that

$$\sqrt{T}(a_m - \hat{a}_m) = wM_{xum}^{-1}.$$

Asymptotically $\sqrt{T}(a_m - \hat{a}_m)$ is distributed like $\bar{w}\bar{M}_{xum}^{-1}$ and its asymptotic variance-covariance matrix is given by:

$$\bar{M}_{uxm}^{-1}\,\mathscr{E}(\bar{w}'\bar{w})\,M_{xum}^{-1} = \sigma^2\,(\bar{M}_{uxm}^{-1}\,\bar{M}_{uu}\,\bar{M}^{-1}{}_{xum}).$$

4. THE REDUCED SET OF INSTRUMENTAL VARIABLES

In general there will be more instrumental variables available than the number of parameters to be determined. Then there arises a problem of choice as to which set of instrumental variables should be used. More generally, one may consider the problem of choosing $n-1$ linear transformations of the available N instrumental variables, so as to provide a reduced set of instrumental variables.

This reduced set of instrumental variables will be denoted by u^*_{it}, $i = 1 \ldots n-1$, and it will be assumed that

$$u^*_{it} = \sum_{j=1}^{N} \theta_{ij} u_{jt}.$$

The problem is then to choose an optimum set of u^*_{it}. It is clear that if any particular set of u^*_{it} is considered any linear transformation of this set is an equivalent set in the sense that it will lead to the same estimates of the coefficients a_i. The choice of the optimum set will be made only with reference to the asymptotic covariance matrix of the estimates of the coefficients.

From the results of the previous section it follows that the asymptotic variance matrix is given by

$$(T\bar{V}_m) = \sigma^2 \left((\theta M_{uxm})^{-1} (\theta M_{uu}\, \theta') (M_{xum}\, \theta')^{-1}\right)$$

where $\theta = \theta(_{ij})$. Now, if the elements of this matrix are denoted by V_{ij} it is convenient to introduce an arbitrary positive definite weighting matrix c_{ij} and to determine the θ_{ij} so as to minimise

$$\begin{aligned}\sum_{i=1}^{n-1} \sum_{j=1}^{n-1} c_{ij} V_{ij} &= \operatorname{tr}\,(cV) \\ &= \operatorname{tr}[c(\theta\bar{M}_{uxm})^{-1} (\theta\bar{M}_{uu}\theta') (\bar{M}_{xum}\theta')^{-1}] \\ &= \operatorname{tr}\,[(\bar{M}_{xum}\theta')^{-1} c(\theta\bar{M}_{uxm})^{-1} (\theta\bar{M}_{uu}\theta')]\,.\end{aligned}$$

Now, the optimum θ will be indeterminate to the extent that a linear transformation can be applied on the left. Further restrictions must, therefore, be imposed on the θ_{ij} to make the solution determinate, and the following seem the most appropriate restrictions.

$$(\theta\bar{M}_{uxm})c^{-1}(\bar{M}_{xum}\theta') = I,$$
$$\theta\bar{M}_{uu}\theta' = \text{a diagonal matrix.}$$

The problem is then to minimise $\operatorname{tr}(\theta\bar{M}_{uu}\theta')$ subject to these conditions. A conventional minimisation using Lagrange multipliers then shows that the rows of θ are characteristic vectors satisfying the equations

$$\bar{M}_{uu}v' = \lambda\, (\bar{M}_{uxm}\, c^{-1}\bar{M}_{xum})\, v'.$$

Thus, $$\bar{M}_{uu}\theta' = (\bar{M}_{uxm}\, c^{-1}\bar{M}_{xum})\, \theta'\, \Delta,$$

where Δ is the diagonal matrix of characteristic roots. Thus,

$$\theta' = (\bar{M}_{uu}^{-1} \bar{M}_{uxm}) (c^{-1} \bar{M}_{xum}\theta' \Delta).$$

The last pair of parentheses enclose a product matrix which is a square $n-1 \times n-1$ matrix, so that θ' is a linear transformation of $\bar{M}_{uu}^{-1} \bar{M}_{uxm}$. It follows that the optimum transformation matrix θ can be more simply taken to be

$$\theta = \bar{M}_{xum} \bar{M}_{uu}^{-1}. \tag{4.1}$$

With this value of θ the asymptotic variance matrix is

$$(\overline{TV}_m) = \sigma^2 (\bar{M}_{xum} \bar{M}_{uu}^{-1} \bar{M}_{uxm})^{-1}. \tag{4.2}$$

Since the optimum value of θ is independent of the matrix c it follows that c can be allowed to tend to a semi-definite matrix. In particular it follows that this choice of θ minimised the asymptotic variance of any linear function of the estimates.

Now of course the matrices $\bar{M}_{uu}$ and $\bar{M}_{xum}$ are unknown, but they are the asymptotic limits of M_{uu} and M_{xum}. This suggests that the transformation $\theta = M_{xum} M_{uu}^{-1}$ should be considered. The corresponding estimates are given by

$$\hat{a}_m M_{xum} M_{uu}^{-1} M_{uxm} = -q_m M_{uu}^{-1} M_{uxm}.$$

Provided $\bar{M}_{xum} \bar{M}_{uu}^{-1} \bar{M}_{uxm}$ is nonsingular these estimates have indeed an asymptotic variance matrix equal to (4.2). This variance matrix is the probability limit of

$$s^2 (M_{xum} M_{uu}^{-1} M_{uxm})^{-1} \tag{4.3}$$

where

$$s^2 = \frac{\sum_{t=1}^{T} \left(\sum_{i=1}^{n} \hat{a}_i \times x_{it}\right)^2}{T}.$$

From these results the advantage of adding one new instrumental variable can be discussed. It is clear that eliminating one variable from a set of instrumental variables is equivalent to obtaining the optimum subject to the condition that the coefficients of one of the instrumental variables are zero, so that eliminating that variable cannot improve the asymptotic variance matrix. It follows that the addition of a new instrumental variable will improve the variance matrix unless the partial correlation between each variable in the relationship and the new instrumental variable is zero after the effects of the other instrumental variables have been allowed for. In practice the addition of a new instrumental variable will usually improve the estimated variance matrix (4.3) unless it leads to an increase in s. However, the improvements are usually small after the first three or four instrumental variables have been added. Thus there may be no great advantage in increasing the number of instrumental variables, and from the later discussion it emerges that the estimates have large biases if the number of instrumental variables becomes too large. We return to this problem of whether there are advantages or disadvantages in using a large number of instrumental variables in our conclusions.

Now it is clear that for each m a different set of estimates is obtained, and that these can be summarised by saying that they are obtained from the set of n equations

$$(M_{xu} M_{uu}^{-1} M_{ux}) \hat{a} = 0 \tag{4.4}$$

upon deleting the mth equation. In Section 7 it will be shown that in fact these estimates of $\hat{a}$ will differ asymptotically by quantities of order $1/T$ provided that $|\bar{M}_{xu} \bar{M}_{uu}^{-1} \bar{M}_{ux}|$ is of rank n—1.

5. THE CANONICAL CORRELATION APPROACH

The symmetrical equations (4.4) immediately suggest the canonical correlation equations introduced by Hotelling [10]. In the latter analysis the intercorrelation between two sets of variables is considered, and it is shown that there are linear transformations of the two sets such that the transformed sets have the following property: The variance matrix for two variables both in the same set is the unit matrix, and the covariance between two variables, one from each set, is zero unless both have the same suffix when it is a canonical correlation.

This analysis can be applied to the present problem by taking the variables in the relationship as one set of variables, and the instrumental variables as the other set of variables. If there is actually a relationship between the variables whose residual is independent of the instrumental variables, then one linear transformation of the variables in the relationship will have zero correlation with all the instrumental variables. It follows that the smallest population canonical correlation is zero. This suggests that the smallest sample correlation should be small, and that the corresponding transformation provides an estimate of the relationship's coefficients. In fact the sample correlation coefficient is the square root of the characteristic root λ of the equation

$$(M_{xu} M_{uu}^{-1} M_{ux} - \lambda M_{xx}) \hat{a}' = 0. \tag{5.1}$$

The characteristic vector is an estimate of the coefficients of the corresponding linear transformation, and, when λ is the smallest root, this provides an estimate of the coefficients of the economic relationship.

If there are two or more small canonical correlations it is clear that the estimates of the coefficients will be rather badly determined, since any linear combination of the two characteristic vectors will be nearly independent of all the instrumental variables. There will be no way of deciding between these different possible vectors unless one has some information about the relationship other than the variables it contains and the independence between the residual and the present set of instrumental variables.

6. A MINIMAX APPROACH

The general properties of the canonical correlation suggest the following alternative derivation. If a_i is any possible set of coefficients, the linear combination

$$\sum_{i=1}^{N} b_i u_{it}$$

is considered which has the maximum correlation with

$$\sum_{i=1}^{n} a_i x_{it}.$$

Then the a_i are chosen to minimise this maximum correlation.

It is clear that the correlation studied can be written vectorally as

$$\varrho^2 = \frac{(aM_{xu}b')^2}{(aM_{xx}a')\ (bM_{uu}b')}.$$

If a is given, ϱ^2 is maximised for the values of b satisfying

$$M_{ux}a = \mu M_{uu}b$$

and then

$$\varrho^2 = \frac{a(M_{xu}\ M_{uu}^{-1}\ M_{ux})\ a'}{a\ M_{xx}\ a'}.$$

Now, the minimum value for ϱ^2 with respect to a, is equal to the smallest root of the characteristic equation

$$(M_{xu}\ M_{uu}^{-1}\ M_{ux} - \lambda\ M_{xx})\ a' = 0.$$

This analysis suggests again that the value of λ would be a suitable criterion for the presence of a relationship of the suggested type.

This approach to the problem is similar to that of Durbin [7] who considers specifically the case in which only one of the variables in the relationship is jointly determined, all the other variables in the relationship being predetermined variables with measurement errors. He suggests that the appropriate procedure is to minimise the canonical correlation between the single variable E_t, which is the residual of the relationship, and the set of variables u_{jt}, $j = 1, \ldots, N$. This canonical correlation is given in the previous notation by

$$T \varrho^2 = \frac{w \, M_{uu}^{-1} \, w'}{s^2}$$

or

$$\varrho^2 = \frac{a \, M_{xu} \, M_{uu}^{-1} \, M_{ux} \, a'}{a \, M_{xx} \, a'}.$$

Thus Durbin's method is equivalent to the method developed here, although he has apparently only considered the restricted case in which n—1 of the variables x_{it} can be expressed as linear functions of the u_{jt} of the form

$$x_{it} = \sum_{j=1}^{N} \beta_{ij} \, u_{jt} + \eta_{it} \qquad (i = 2, \ldots, n)$$

and the η_{it} are all random variables independent of all previous variables in the model and non-autocorrelated. These are clearly more restrictive assumptions than those used in this article.

7. THE ASYMPTOTIC DISTRIBUTION OF THE SMALLEST CHARACTERISTIC ROOT

In this section the distribution of the smallest root of (5.1) is considered on the assumption that $\bar{M}_{xu}$ is of rank n—1, so that only one population characteristic root is zero.

Now there will be a unique matrix H which will satisfy the equation $HH' = M_{uu}^{-1}$ and a suitable arbitrary set of linear restrictions of number $\frac{1}{2}N(N-1)$. It follows that $\overline{HH'} = \bar{M}_{uu}^{-1}$, and, if $t = wH/\sigma$, then $\bar{t} = \overline{wH}/\sigma$. Thus the components of $\bar{t}$ are all asymptotically independently normally distributed with unit variance.

arbitrary Now, if

$$\frac{a \, M_{xu} \, M_{uu}^{-1} \, M_{ux} \, a'}{a \, M_{xx} \, a'}$$

is considered as a function of a, the smallest characteristic root λ_1 is its minimum. Hence:

$$T \lambda_1 \leqslant \frac{T \, a \, M_{xu} \, M_{uu}^{-1} \, M_{ux} \, a'}{a \, M_{xx} \, a'} = F\,(a)$$

and

$$T\,(a \, M_{xu} \, M_{uu}^{-1} \, M_{ux} \, a') = w \, M_{uu}^{-1} \, w'$$

and, of course,

$$\text{plim}\,(a \, M_{xx} \, a') = \sigma^2.$$

Thus the asymptotic distribution of $F(a)$ is the same as that of

$$\sum_{i=1}^{N} t_i^2$$

and, since the latter is a continuous function of the t_i, it follows that

$$\bar{F}(a) = \sum_{i=1}^{N} \bar{t}_i^2.$$

This has a χ^2 distribution of N degrees of freedom, so that λ is asymptotically of order $1/T$.

Now, if $\bar{M}_{xu}\,\bar{M}_{xu}^{-1}\,\bar{M}_{ux}$ is of rank n—1, there is a unique a satisfying

$$(\bar{M}_{xu}\,\bar{M}_{uu}^{-1}\,\bar{M}_{ux})\,a' = 0$$

and $\operatorname{plim}_{T\to\infty} \hat{a} = a.$

From the equations determining them and the previous result it follows that if $\hat{a}$ is suitably standardised it differs asymptotically from $\hat{a}_m$ by quantities of order $1/T$. The asymptotic error variance matrix of $\hat{a}_m$ shows $\Delta a = a - \hat{a}$ is of order $1/\sqrt{T}$.

Now

$$\begin{aligned} &a\,(M_{xu}\,M_{uu}^{-1}\,M_{ux} - \lambda_1\,M_{xx})\,a' \\ &= (\hat{a} + \Delta a)\,(M_{xu}\,M_{uu}^{-1}\,M_{ux} - \lambda_1\,M_{xx})\,(\hat{a}' + \Delta a') \\ &= \Delta a\,(M_{xu}\,M_{uu}^{-1}\,M_{ux} - \lambda_1\,M_{xx})\,\Delta a', \end{aligned}$$

using

$$(M_{xu}\,M_{uu}^{-1}\,M_{ux} - \lambda_1\,M_{xx})\,\hat{a}' = 0$$

so that

$$\lambda_1 = \frac{a\,M_{xu}\,M_{uu}^{-1}\,M_{ux}\,a' - \Delta a(M_{xu}\,M_{uu}^{-1}\,M_{ux})\,\Delta a}{a\,M_{xx}\,a' - \Delta a(M_{xx})\,\Delta a'}.$$

Now, since Δa is asymptotically zero, it follows that

$$\operatorname*{plim}_{T\to\infty}\,(a\,M_{xx}\,a' - \Delta a(M_{xx})\,\Delta a') = \sigma^2$$

so that the asymptotic distribution of $T\lambda_1$ is the same as that of

$$\frac{T(a\,M_{xu}\,M_{uu}^{-1}\,M_{ux}\,a')}{\sigma^2} - \frac{T(\Delta a\,(M_{xu}\,M_{uu}^{-1}\qquad x)\,\Delta a')}{\sigma^2}. \tag{7.1}$$

Now, if Δa is replaced by $a_m - \hat{a}_m$, the difference in (7.1) is easily seen to be of order $1/\sqrt{T}$. Hence, no difference will be produced in the asymptotic distribution of (7.1). Both a_m and $\hat{a}_m$ are standardised so that their mth components are unity. In the second part of (7.1) will only appear those terms which do not involve these components. Further, the components of $\sqrt{T}\,(a_m - \hat{a}_m)$ can be represented as linear functions of the w_t. They are asymptotically jointly normally distributed with variance matrix equal to

$$\frac{\sigma^2}{T}\,(\bar{M}_{xum}\,\bar{M}_{uu}^{-1}\,\bar{M}_{uxm})^{-1}.$$

Thus, as before, a linear transformation of the w_t can be defined so that

$$\frac{T}{\sigma^2}\,(a\,M_{xu}\,M_{uu}^{-1}\,M_{ux}\,a') = \sum_{i=I}^{N} t_i^2$$

$$\frac{T}{\sigma^2}\,(a_m - \hat{a}_m)\,(M_{xu}\,M_{uu}^{-1}\,M_{ux})\,(a_m' - \hat{a}_m') = \sum_{i=1}^{n-1} t_i^2$$

then

$$(\overline{T\lambda_1}) = \sum_{i=n}^{N} (\bar{t}_i)^2$$

where all the t_i are normally and independently distributed with unit variance, so that $(\overline{T\lambda}_1)$ is distributed as χ^2 with $N - n + 1$ degrees of freedom.

This provides a significance test for the hypothesis that there is a relationship between the suggested variables with a residual independent of all the instrumental variables.

This is a suitable test even when $\bar{M}_{xu}$ is of rank less than n—1 since it can be shown that in this case the probability of rejecting the hypothesis will be less than in the other case.

8. THE A PRIORI UNIDENTIFIED CASE

The results of the previous section depend on the assumption that $\bar{M}_{xu}$ is of rank $n - 1$. If its rank is less than this, the relationship will be said to be a priori unidentified. At least two of the population characteristic roots are then zero and the equation

$$(\bar{M}_{xu}\,\bar{M}_{uu}^{-1}\,\bar{M}_{ux})\,a' = 0$$

no longer has a unique solution.

Let us now assume that $\bar{M}_{xu}$ is of rank $n - 2$ and that there is no serial correlation for any linear combination ax'_t which corresponds to the double-zero characteristis root. If λ_1 and λ_2 are the two smallest sample characteristic roots, one may prove, by the methods of the last section, that $T(\lambda_1 + \lambda_2)$ is asymptotically distributed as χ^2 with $2\ (N - n + 2)$ degrees of freedom.

This result can now be used as an approximate significance test of the hypothesis that the relationship is a priori unidentified and that any possible relationship has a non-autocorrelated residual. This hypothesis is not very likely to be true a priori since even if there is a relationship between the suggested variables with a non-autocorrelated residual it is unlikely that there would be a second combination of these variables not only independent of all the instrumental variables but non-autocorrelated as well. In practice the significance test fairly often indicates that the hypothesis may be true, but this is probably because the smallest non-zero root of the population characteristic equation is small, the corresponding residual is not very highly autocorrelated and T is not large enough to make $T(\lambda_1 + \lambda_2)$ significant. The use of the test, however, provides a useful qualitative answer as to whether the estimates are reasonably well identified, although as noted in Section 12 the fact that the estimates are well identified does not necessarily mean that they have reasonably small standard errors.

9. ON THE USE OF VARIABLES IN THE RELATIONSHIP AS INSTRUMENTAL VARIABLES

It is now worthwhile to introduce a notation which explicitly recognises that some of the variables in the relationship may be used as instrumental variables. As noted in Section 2 this is only possible if they are predetermined variables with zero measurement errors. The most important variables which are of this type are those representing the constant term, the trend, and seasonal factors. It is convenient to use a notation very similar to that used by Anderson [3].

Let H and K^* be, respectively, the number of variables in the relationship which cannot and which can be used as instrumental variables ($K^* + H = n$). Let K^{**} be the number of instrumental variables not in the relationship ($K^{**} + K^* = N$).

Let the variables in the relationship which cannot be used as instrumental variables be denoted by $y_{it'}$, $i = 1, \ldots, H$; let the instrumental variables in the relationship be denoted by u^*_{it}, $i = 1, \ldots, K^*$; and let the other instrumental variables be denoted by u_{it}, $i = 1, \ldots, K^{**}$. Then M_{uu} can be partitioned as below

$$M_{uu} = \left(\begin{array}{c|c} M_{u^*\ u^*} & M_{u^*\ u^{**}} \\ \hline M_{u^{**}\ u^*} & M_{u^{**}\ u^{**}} \end{array}\right).$$

Similarly,

$$M_{xu} = \left(\frac{M_{yu}}{M_{u^*\ u}}\right) = \left(\begin{array}{c|c} M_{yu^*} & M_{yu^{**}} \\ \hline M_{u^*\ u^*} & M_{u^*u^{**}} \end{array}\right),$$

and

$$M_{xx} = \left(\frac{M_{yy}}{M_{u^*y}}\middle|\frac{M_{yu^*}}{M_{u^*u^*}}\right).$$

Now clearly

$$M_{u^*u}\, M_{uu}^{-1} = (I \mid 0)$$

so that

$$M_{xu}\, M_{uu}^{-1} = \left(\frac{M_{yu}}{M_{u^*u}}\right) M_{uu}^{-1} = \left(\frac{M_{yu}\, M_{uu}^{-1}}{I \mid 0}\right).$$

Thus

$$M_{xu} M_{uu}^{-1} M_{ux} = \left(\frac{M_{yu} M_{uu}^{-1} M_{uy}}{M_{u^*y}}\middle|\frac{M_{yu} M_{uu}^{-1} M_{uu^*}}{M_{u^*u^*}}\right)$$

$$= \left(\frac{M_{yu}\, M_{uu}^{-1}\, M_{uy}}{M_{u^*y}}\middle|\frac{M_{yu^*}}{M_{u^*u^*}}\right).$$

Writing $\hat{a} = (\hat{b} \mid \hat{c})$ equations (5.1) take the form

$$\left(\frac{M_{yu}\, M_{uu}^{-1}\, M_{uy}}{M_{u^*y}}\middle|\frac{M_{yu^*}}{M_{u^*u^*}}\right)\left(\frac{\hat{b}'}{\hat{c}'}\right) = \lambda\left(\frac{M_{yy}}{M_{u^*y}}\middle|\frac{M_{yu^*}}{M_{u^*u^*}}\right)\left(\frac{\hat{b}'}{\hat{c}'}\right)$$

or

$$(M_{yu}\, M_{uu}^{-1}\, M_{uy} - \lambda M_{yy})\, \hat{b}' + (1 - \lambda)\, M_{yu^*}\, \hat{c}' = 0 \tag{9.1}$$

and

$$(1 - \lambda)\, (M_{u^*y}\, \hat{b}' + M_{u^*u^*}\, \hat{c}') = 0. \tag{9.2}$$

It follows that in this case K^* of the canonical correlations are unity, the corresponding canonical transformations being those for which $\hat{b} = 0$. For the other canonical correlations $\lambda < 1$, and it is then possible to solve equations (9.2) for $\hat{c}$ in terms of $\hat{b}$ in the form

$$\hat{c}' = -\, M_{u^*u^*}^{-1}\, M_{u^*y}\, \hat{b}'.$$

Substituting in equations (9.1) and rearranging, we obtain

$$(M_{yu}M_{uu}^{-1}M_{uy} - M_{yu^*}M_{u^*u^*}^{-1}M_{u^*y})\hat{b}' = \lambda(M_{yy} - M_{yu^*}M_{u^*u^*}^{-1}M_{u^*y})\hat{b}'). \tag{9.3}$$

This has eliminated the roots $\lambda = 1$ from the characteristic equation, and so has reduced the degree of the characteristic equation to H.

10. THE CONFIDENCE REGIONS

It is now possible to derive two confidence regions for the unknown coefficients of the relationship. We use the notation

$$F_1(a) = \frac{T}{\sigma^2}(bM_{yy}b' + 2cM_{u^*y}b' + cM_{u^*u^*}c')$$

$$= \frac{T}{\sigma^2}(aM_{xx}a') = \sum_{t=1}^{T} \frac{E_t^2}{\sigma^2}$$

and has a χ^2 distribution with T degrees of freedom.
Similarly,

$$F_2(a) = \frac{T}{\sigma^2}(b\,M_{yu}\,M_{uu}^{-1}\,M_{uy}\,b' + 2cM_{u^*y}\,b' + cM_{u^*u^*}\,c')$$

$$= \frac{T}{\sigma^2}(aM_{xu}\,M_{uu}^{-1}\,M_{ux}\,a')$$

and from the argument of Section 7 it follows that asymptotically $F_2(a)$ is distributed as χ^2 with $K^* + K^{**}$ degrees of freedom. But the argument of Section 7 can be repeated for only the instrumental variables in the relationship, and it will then be found that

$$F_3(a) = \frac{T}{\sigma^2}(bM_{yu^*}\,M_{u^*u^*}^{-1}\,M_{u^*y}\,b' + 2c\,M_{u^*y}\,b' + cM_{u^*u^*}\,c')$$

$$= \frac{T}{\sigma^2}(aM_{xu^*}\,M_{u^*u^*}^{-1}\,M_{u^*x}\,a')$$

is asymptotically distributed as χ^2 with K^* degrees of freedom. Further, it can be shown by the methods of Section 7 that $F_1 - F_2$, $F_2 - F_3$, and F_3 are all asymptotically independent of each other and are therefore all asymptotically distributed as independent χ^2's with $T - K^* - K^{**}$, K^{**}, and K^* degrees of freedom, respectively.

Thus $F^2/(F_1 - F_2)$ is asymptotically distributed as the ratio of two independent χ^2's and so

$$\varphi_1 = \frac{F_2}{F_1 - F_2}\left(\frac{T - K^* - K^{**}}{K^* + K^{**}}\right)$$

is distributed as the ratio of two independent estimates of the same variance based upon samples with degrees of freedom N and $T - N$ respectively. A confidence region can now be used representing all values of b and c which satisfy the inequality

$$\frac{b\,M_{yu}\,M_{uu}^{-1}\,M_{uy}\,b' + 2\,b\,M_{yu^*}\,c' + c\,M_{u^*u^*}\,c'}{b\,(M_{yy} - M_{yu}\,M_{uu}^{-1}\,M_{uy})\,b'} = \frac{F_2}{F_1 - F_2} \leqslant \varphi_{1L}\,\frac{N}{T - N}$$

where φ_{1L} is a limit which is likely to be exceeded with only a small probability, say 5%. Then, if we consider the possibility that the actual values of b and c correspond to a point outside the confidence region, it is clear that the probability of the sample φ_1 being greater than φ_{1L} would be less than 5%. Thus, using this criterion, it is unlikely that the data could have been produced if the relationship had coefficients corresponding to a point outside the confidence region. This type of confidence region will be referred to as a confidence region of Type I.

In the same way $(F_2 - F_3)/(F_1 - F_2)$ is distributed asymptotically as the ratio of two independent χ^2's and so

$$\varphi_2 = \frac{F_2 - F_3}{F_1 - F_2}\left(\frac{T - K^* - K^{**}}{K^{**}}\right)$$

is asymptotically distributed as the ratio of two independent variance estimates of degrees of freedom K^{**} and $T - N$ and, if φ_{2L} is an appropriate limit,

$$\left(\frac{F_2 - F_3}{F_1 - F_2} = \frac{b\,(M_{yu} M_{uu}^{-1} M_{yu} - M_{uy^*} M_{u^*u^*}^{-1} M_{u^*y})\,b'}{b\,(M_{yy} - M_{yu} M_{uu}^{-1} M_{uy})\,b'} \leqslant \frac{K^{**}\,\varphi_{2L}}{T - N}\right)$$

provides an appropriate confidence region for the components of b, which is usually more useful than the region of Type I since the coefficients c are usually not very interesting. For computational purposes it is often useful to rewrite the last equation in the form

$$\left(\frac{b\,(M_{yu} M_{uu}^{-1} M_{uy} - M_{yu^*} M_{u^*u^*}^{-1} M_{u^*y})\,b'}{b\,(M_{yy} - M_{yu^*} M_{u^*u^*}^{-1} M_{u^*y})\,b'} \leqslant \frac{K^{**}\,\varphi_{2L}}{T - N + K^{**}\,\varphi_{2L}}\right).$$

This type of region will be referred to as the Type II confidence region.

These confidence regions are exact for finite T provided all the instrumental variables are completely exogenous, that is, provided all the u_{it} are independent of all the E_t' for all t and t'. Otherwise, they are only accurate as $T \to \infty$. It is to be noted that they are valid even in the a priori unidentified case, although in this case they will usually be hyperbolic conics or hyperconics. Thus they are certainly usable in the more usual almost-unidentified case. Further, in view of the asymptotic approximation, they are equally well defined by the conditions that $T\,F_2/F_1$ and $T\,(F_2 - F_3)/(F_1 - F_3)$ are distributed as χ^2's with N and K^{**} degrees of freedom.

It is also noteworthy that these regions depend upon the assumption that the residuals are non-autocorrelated. If the residuals are positively autocorrelated the confidence regions (and of course the computed standard errors) will understate the indeterminacy of the coefficients.

11. THE ACCURACY OF THE ASYMPTOTIC APPROXIMATION

The previous sections have been concerned only with the purely theoretical problems of the asymptotic behaviour of the estimates. In practice T is unlikely to be greater than 100, and in some cases it may be necessary to attempt estimation where T is as small as 15, so that the usefulness of the asymptotic approximations depends upon their accuracy for finite T. It is very difficult to work out the actual distributions of the sample functions that have been used for finite T, and an obviously simpler approach is to attempt to calculate some of the important properties of these distributions (e.g., their moments).

The author has derived the expressions for the first and second moments for the case in which the variables are generated by linear stochastic models of the type considered by Koopmans et al. [**12**], disturbed by measurement errors, and with all random elements normally distributed. The calculations are, however, too lengthy to be reproduced here.

The general conclusion is that the biases in the estimates $\hat{a}$ and $T\lambda$ are both of order $N/T\overline{\lambda}_2$ where $\overline{\lambda}_2$ is the square of the smallest non-zero population canonical correlation coefficient. The biases are of course large when $\overline{\lambda}_2$ is small, that is, when the relationship is almost unidentified.

Likewise, it is found that the size of the biases in the confidence regions is not, to a first approximation, dependent on $\overline{\lambda}_2$ but is still proportional to N. This means that even when T is large, of order 100, N must be limited if the asymptotic approximation is to be satisfactory. One may consider as a necessary requirement that $N \leqslant T/20$. This, indeed, limits severely the number of instrumental variables which may be used.

The approximation obtained by regarding $w\, M_{uu}^{-1}\, w'/\sigma^2$ as distributed as χ^2 with N degrees of freedom gives a distribution for the variable which is biased in a positive direction. In particular, confidence regions based on it will be too large, and significance tests based indirectly on it will have too large a chance of accepting a suggested form of relationship. A simple change, which would probably reduce the bias considerably, is to assume that

$$\frac{T + 4\, N}{\sigma^2} (a\, M_{xu}\, M_{uu}^{-1}\, M_{ux}\, a')$$

is distributed as χ^2 with N degrees of freedom.

12. A COMPARISON WITH ALTERNATIVE METHODS: (1) LEAST SQUARES

In discussing the use of least squares to estimate the coefficients of a relationship it is convenient to denote the dependent variable by y_t, and the remaining variables by z_{it}, $i = 1, \ldots, p$. The relationship may then be written

(12.1) $$y_t = \sum_{i=1}^{P} a_i z_{it} + \varepsilon_t \qquad (t = 1, \ldots, T).$$

and the least squares equations can be written

(12.2) $$m_{yz} = \hat{a} M_{zz}$$

where

$$m_{yz} = \frac{1}{T} \sum_{t=1}^{T} y_t z_{it}$$

and

$$M_{zz} = \frac{1}{T} \sum_{t=1}^{T} z_{it} z_{jt}.$$

Now, multiplying each of equations (12.1) by z_{jt} and summing, we obtain

$$\sum_{t=1}^{T} y_t z_{jt} = \sum_{i=1}^{P} a_i \sum_{t=1}^{T} z_{it} z_{jt} + \sum_{t=1}^{T} \varepsilon_t z_{jt}$$

or

$$m_{yz} = a M_{zz} + v$$

where

$$v = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t z_{jt}$$

so that

$$(\hat{a} - a) M_{zz} = v.$$

Now let

$$\mathscr{E}(M_{zz}) = A_{zz}.$$

It will be assumed that the variables are stationary so that A_{zz} is independent of T. Let $M_{zz} - A_{zz} = Z_{zz}$.

Then

(12.3) $$A_{zz} (\hat{a}' - a') = v' - Z_{zz} (\hat{a}' - a').$$

Now, in the asymptotic approximation the second term has a probability limit of zero. If $\mathscr{E}(\bar{v}) = u$ is not zero, then asymptotically the biases in the estimates $\hat{a}$ can be taken as

(12.4) $$\mathscr{E}(\hat{a}' - a') = A_{zz}^{-1} u'.$$

It is now convenient to write $|A_{zz}| = \Delta$ and to denote the elements of the matrix adjugate to A_{zz} by Δ_{ij}. One may also write $u_i = \sigma_i \sigma \varrho_i$ where σ_i is the standard deviation of Z_{it}, σ the standard deviation of ε_t, and ϱ_i the correlation coefficient between Z_{it} and ε_t. Assuming that the variables are standardised so that $\sigma_i = 1$, the above equation (12.4) can be written

$$\Delta\,(\mathscr{E}(\hat{a}_i - a_i)) = (\sum_{j=1}^{P} \Delta_{ij}\,\varrho_j)\,\sigma.$$

An upper bound on the bias can be obtained by noting that $|\varrho_j| \leqslant 1$ and $\sigma^2 = 1 - R^2$, R being the multiple correlation coefficient. Also

$$\max_{i,j} \left(\frac{\Delta^2{}_{ij}}{\Delta^2}\right) = \max_{i} \left(\frac{\Delta_{ii}}{\Delta}\right)^2 = \left(\frac{1}{1 - R_m^2}\right)^2$$

where R_m is the maximum multiple correlation between any z_{it} and the other independent variables. From the properties of the multiple correlations one has $R_m \leqslant R$. Thus the bias is certainly less than

$$\frac{p\sqrt{1-R^2}}{1 - R_m^2} \leqslant \frac{p}{\sqrt{1 - R_m^2}} \leqslant \frac{p}{\sqrt{1-R^2}}.$$

The bias may become large if R_m is near unity or, in other words, if there is some Δ_{ii} which is large compared with Δ. Δ is then almost singular.

Returning to the notation of the previous sections, dropping the distinction between dependent and independent variables and denoting all the variables in the relationship by x_{it}, $i = 1, \ldots, n$, the biases will be large when the matrix $\bar{M}_{xx}$ is approximately of rank $n - 2$ or less. When this is true it is usual to say that the variables are confluent. In this case, even if the conditions are fulfilled which make the least squares estimates consistent, that is, even if $\varrho_i = 0$ for all i, the standard error of the estimates will still be large. When these conditions are not fulfilled it is clear by using equations (12.3) that the variance of the estimates is also large.

From the results of the previous sections it follows that if the instrumental variables have been correctly chosen their estimates are always consistent, in the sense that their biases tend to zero as T becomes infinite provided that the relationship is a priori identified. However, if the variables in the relationship are confluent so that $\bar{M}_{xx}$ is almost of rank $n - 2$ or less, it can be shown that the standard errors of the estimates are still large.

To show this it is convenient to return to the notation and results of Section 4. Adopting again the normalisation $a_m = 1$ for arbitrary m, the asymptotic error variance matrix of the estimates of the instrumental variables is

$$V_m = \frac{\sigma^2}{T}\,(\bar{M}_{xum}\,\bar{M}_{uu}^{-1}\,\bar{M}_{uxm})^{-1}$$

and the corresponding error variance matrix, which would be obtained for the least squares estimates on the assumption that the method is usable, would be

$$V_m^* = \frac{\sigma^2}{T}\,(\bar{M}_{xxm}^{-1})$$

where $\bar{M}_{xxm}$ means $\bar{M}_{xx}$ with the mth row and column omitted. If γ is any vector with $n - 1$ components, the function $\gamma\, \hat{a}'_m$ will have an asymptotic variance

$$\frac{\sigma^2}{T}\,(\gamma\,(\bar{M}_{xum}\,\bar{M}_{uu}^{-1}\,\bar{M}_{uxm})^{-1}\,\gamma').$$

But, if instead the a_m are assumed to have a variance matrix equal to that appropriate to least squares estimates, $\gamma\, \hat{a}'_m$ will have a variance

$$\frac{\sigma^2}{T}\,(\gamma\,\bar{M}_{xxm}^{-1}\,\gamma').$$

If a canonical correlation transformation is now introduced, and if H is the transformation matrix of $\bar{M}_{xxm}$, it follows that

$$H\,\bar{M}_{xxm}\,H' = I.$$

If then $H\gamma' = \delta$, it follows that

$$\gamma\,\bar{M}_{xxm}^{-1}\,\gamma' = \delta\delta' = \sum_{i=1}^{n-1} \delta_i{}^2,$$

and

$$\gamma\,(\bar{M}_{xum}\,\bar{M}_{uu}^{-1}\,\bar{M}_{uxm})\,\gamma' = \sum_{i=1}^{n-1} \frac{\delta_i{}^2}{\varrho_i{}^2}$$

where the ϱ_i are the canonical correlations. Thus, unless $\delta_i = 0$ except when $\varrho_i = 1$ $\gamma\, V_m \gamma' > \gamma\, V_m^* \gamma'$ and in any case $\gamma\, V_m \gamma' \geqslant \gamma\, V_m^* \gamma'$. Thus, the standard error of $\gamma\, \hat{a}'_m$ is usually greater than would have been obtained if the least squares method could be used. And, in particular, if the variables are confluent so that the least squares standard errors are large, the standard errors using instrumental variables will be even larger.

Theoretically then, if the asymptotic properties of the two kinds of estimates are compared, the instrumental variables method (provided the relationship is a priori identified) is the better, since the estimates are consistent, whereas the least squares estimates are not. However, for finite T, the advantage of using the instrumental variables method is less certain, since the instrumental variables estimates may have large biases especially in the almost unidentified case and in the event the number of instrumental variables is large.

The best practical test is obtained by comparing the results yielded by the two methods in practical cases. Let us suppose that, using a minimum of instrumental variables, confidence regions are obtained (these are still reasonably good approximations even in the almost unidentified case) and that the least squares estimates lie outside the confidence regions. This is sufficient evidence to show that the least squares estimates are probably significantly biased. In practice this happens when it would be expected theoretically, i.e., when there are at least two important jointly determined variables, or when some of the variables treated as independent in applying the method of least squares have large and obvious random variations.

13. A COMPARISON WITH ALTERNATIVE METHODS: (2) THE LIMITED-INFORMATION MAXIMUM-LIKELIHOOD METHOD

It is clear from a comparison of the equations (9.3) and the similar equations of the limited-information maximum-likelihood (L.I.M.L.) method as formulated by Anderson and Rubin [**1**, **2**] that the two methods are in practice very similar. Indeed the equations (9.3) can be easily transformed to the form

$$(M_{yy} - M_{yu} M_{uu}^{-1} M_{uy})\, \hat{b}' = \mu(M_{yy} - M_{yux} M_{u^*u^*}^{-1} M_{u^*y})\, \hat{b}'$$

where $\mu = 1/(1 - \lambda)$.

This differs from the L.I.M.L. equations only by the replacement of z, representing all the predetermined variables, with u, representing the instrumental variables. Indeed the L.I.M.L. method is equivalent to using the instrumental variables method with all the predetermined variables in the model used as instrumental variables. This procedure is reasonable since an essential assumption of the L.I.M.L. method is that there are no measurement errors. As argued in Section 2, it is then possible to use the predetermined variables in the relationship as instrumental variables. The difference between the two methods can be summarised as below.

(*i*) The L.I.M.L. method strictly interpreted requires that all predetermined and exogenous variables which occur in any relationship in the model should be used. The instrumental variables approach is wider in allowing the use of any suitable predetermined variable whether it occurs in any relationship or not.

(*ii*) The instrumental variable method, however, is narrower in not allowing the use of any predetermined variables occurring in the actual relationship being studied, unless it has no measurement error or disturbance, for the reasons discussed in Section 2. It is also suggested in Section 2 that it is probably not wise to use lagged values of a variable appearing in the relationship as instrumental variables.

With both methods practical difficulties of computation and theoretical considerations concerning the biases make it worthwhile in practice to use only a small selection of the vast number of possible predetermined variables or instrumental variables.

Alternatively, it may be said that the instrumental variables method consists in modifying the L.I.M.L. method by treating those predetermined variables in the relationship which have measurement errors as if they were jointly determined variables. This method of treating variables with measurement errors has already been suggested by Chernoff and Rubin [**5**].

14. SOME GENERAL CONCLUSIONS

From the previous results and from quite a large amount of work carried out by the author to test the method in practice, part of which it is hoped will be published later, several tentative conclusions may be drawn:

(*i*) The use of the instrumental variables method will produce consistent estimates of coefficients even when large measurement errors are apparent.

(*ii*) Better results may appear to be obtained if an attempt is made to reduce measurement errors somewhat. One may, for example, smooth the series containing obvious random errors by use of moving averages. Such a process, however, often increases the autocorrelation of the residual.

(*iii*) The use of the limited-information maximum-likelihood method is likely to produce, and in practice has produces, biased estimates when there are large measurement errors in any of the predetermined variables in the relationship. The least squares method is likely to produce large biases when there are large measurement errors, or when some of the independent variables are not predetermined variables and the relationship is confluent.

(*iv*) The use of large numbers of instrumental variables may not improve the accuracy of the estimates. In practice, the effect of increasing the number of instrumental variables has been tried by the author. It has been found that if the first few instrumental variables are well chosen, there is usually no improvement, and even a deterioration, in the confidence regions as the number of instrumental variables is increased beyond three or four. This might have been expected a priori from a study of the results achieved by Stone [**15**] and others when applying factor analysis to economic time series. It has usually been found that three general factors were obtained corresponding to a linear trend, the ten year business cycle, and the rate of change of the ten year cycle. Now, if all the variables in the relationship and all the instrumental variables together can be approximately analysed in this way, the maximum number of coefficients that can be simultaneously determined is three, and the addition of instrumental variables after the third will not improve the accuracy. If, however, there are large random effects (of the same order of magnitude as the cyclical movements) such as strikes, wars etc., and if the residual of the relationship can be regarded as independent of the random effects then this might allow the determination of further coefficients. The latter assumption is, however, very rarely realistic.

In practice, when data covering less than twenty years are used, it seems appropriate to use three instrumental variables: a linear trend, a lagged variable that leads in the trade cycle, and a lagged variable that lags with reference to the trade cycle. Analyses of single economic time series indicate that if longer periods of time were studied a factor analysis might disclose more general factors, for example, another factor corresponding to a parabolic trend, and two more factors to represent the building cycle. To some extent, this gain might be cancelled by the need to introduce more complicated trends into the relationship.

Leeds University

REFERENCES

[1] ANDERSON, T. W. AND H. RUBIN: "Estimation of the Parameters of a Single Equation in a Complete Set of Stochastic Equations," *Annals of Mathematical Statistics*, Vol. 20, 1949, pp. 46-65.

[2] ANDERSON AND RUBIN: "The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, Vol. 21, 1950, pp. 570-592.

[3] ANDERSON, T. W.: "Estimation of the Parameters of a Single Equation by the Limited Information Maximum-Likelihood Method," *Cowles Commission Monograph* 10, 1950, pp. 311-322.

[4] ———: "The Asymptotic Distribution of Certain Characteristic Roots and Vectors," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, 1950, pp. 103-130.

[5] CHERNOFF, H. AND H. RUBIN: "Asymptotic Properties of Limited Information Estimates under Generalised Conditions," *Cowles Commission Monograph* 14, 1953, pp. 200-212.

[6] CRAMER, H.: *Mathematical Methods of Statistics*, 1946.

[7] DURBIN, J.: "Errors in Variables," *Review of Institute of International Statistics*, Vol. 22, 1954, pp. 23-54.

[8] FRISCH, R.: *Statistical Confluence Analysis by Means of Complete Regression Systems*, 1934.

[9] GEARY, R. C.: "Studies in Relations between Economics Time Series", *Journal of the Royal Statistical Society, Series B*, Vol. 10, 1949, pp. 158-172.

[10] HOTELLING, H.: "Relations between Two Sets of Variables," *Biometrika* Vol. 28, 1936, pp. 321-335.

[11] HSU, P. L.: "On the Limiting Distribution of Roots of a Determinantal Equation," *Journal of the London Mathematical Society*, Vol. 16, 1941, pp. 183-194.

[12] KOOPMANS, T. C., H. RUBIN, AND R. B. LEIPNIK: "Measuring the Equation Systems of Dynamic Economics," *Cowles Commission Monograph 10*, pp. 52-237.

[13] REIERSØL, O.: "Confluence Analysis by Means of Lag Moments and Other Methods of Confluence Analysis," *Econometrica*, Vol. 9, no. 1, 1941, pp. 1-23.

[14] ———: *Confluence Analysis by Means of Sets of Instrumental Variables*, 1945.

[15] STONE, R.: "On the Interdependence of Blocks of Transactions," *Journal of the Royal Statistical Society Supplement* 8, 1947, pp. 1-13.

[16] TURNBULL, H. W. AND A. C. AITKEN: *An Introduction to the Theory of Canonical Matrices*, 1938.