# ESTIMATION OF THE PARAMETERS OF A SINGLE EQUATION IN A COMPLETE SYSTEM OF STOCHASTIC EQUATIONS[1,2]

By T. W. Anderson[3] and Herman Rubin[4]

*Columbia University and Institute for Advanced Study*

**1. Summary.** A method is given for estimating the coefficients of a single equation in a complete system of linear stochastic equations (see expression (2.1)), provided that a number of the coefficients of the selected equation are known to be zero. Under the assumption of the knowledge of all variables in the system and the assumption that the disturbances in the equations of the system are normally distributed, point estimates are derived from the regressions of the jointly dependent variables on the predetermined variables (Theorem 1). The vector of the estimates of the coefficients of the jointly dependent variables is the characteristic vector of a matrix involving the regression coefficients and the estimate of the covariance matrix of the residuals from the regression functions. The vector corresponding to the smallest characteristic root is taken. An efficient method of computing these estimates is given in section 7. The asymptotic theory of these estimates is given in a following paper [2].

When the predetermined variables can be considered as fixed, confidence regions for the coefficients can be obtained on the basis of small sample theory (Theorem 3).

A statistical test for the hypothesis of over-identification of the single equation can be based on the characteristic root associated with the vector of point estimates (Theorem 2) or on the expression for the small sample confidence region (Theorem 4). This hypothesis is equivalent to the hypothesis that the coefficients assumed to be zero actually are zero. The asymptotic distribution of the criterion is shown in a following paper [2] to be that of $\chi^2$.

**2. A complete system of linear difference equations.** In many fields of study such as economics, biology, and meteorology the occurrence of values of the observed quantities can be described in terms of a probability model which, as a first approximation, is a set of stochastic equations. Consider a (row) vector $y_t$ of quantities which are observed at time $t$. Suppose that these quantities are *jointly dependent* on a vector $z_t$ of quantities "predetermined" at time $t$ (i.e., known without error at time $t$). Some of the coordinates of $z_t$ may be coordinates

[1] This paper will be included in Cowles Commission Papers, New Series, No. 36.

[2] The results in this paper were presented at meetings of the Institute of Mathematical Statistics in Washington, D. C., April 12, 1946 (Washington Chapter) and in Ithaca, N. Y., August 23, 1946.

[3] Fellow of the John Simon Guggenheim Memorial Foundation; Research Consultant of the Cowles Commission for Research in Economics.

[4] National Research Fellow; Research Consultant of the Cowles Commission for Research in Economics.

of $y_{t-1}$, $y_{t-2}$, etc.; other coordinates of $z_t$ are quantities which are assumed given constants. The set of vectors $y_t(t = 1, 2, \cdots, T)$ are called *endogenous*. The part of the set $z_t$ which does not consist of lagged endogenous variables is called *exogenous*; these are treated as "fixed variates." For convenience we shall think of $t$ as indicating a point of time, although it may in many cases indicate the ordering of a sample in another dimension, or, indeed, the $t$ may indicate simply a numbering of the observations (if $z_t$ is entirely *exogenous*). In a dynamic economic model the endogenous variables are economic quantities such as amount of investment, interest rate, amount of consumption, etc. The exogenous variables are those quantities which are considered to be determined primarily outside the economic system, such as amount of rainfall, amount of government expenditures, time, etc.

A simple probability model may be set up on the assumption that these quantities approximately satisfy certain linear equations. Specifically the model is

$$B_{yy}y'_t + \Gamma_{yz}z'_t = \epsilon'_t \tag{2.1}$$

where $\epsilon_t$ is a (row) vector having a probability distribution with expected value zero and $B_{yy}$ and $\Gamma_{yz}$ are matrices, the former being non-singular. Primes (′) indicate transposition of vectors and matrices. If there are $G$ jointly dependent variables, there are $G$ component equations in (2.1); that is, there are as many equations as there are variables depending on the system. The fact that $y_t$ and $z_t$ do not satisfy linear equations exactly is indicated by setting the linear forms not equal to zero, but equal to random elements, called *disturbances*. We will call the component equations of (2.1) *structural* equations, for they express the structure of the system. For example, one equation involving the amount of goods consumed, the prices of these goods, the size of the national income, etc., might describe the behaviour of the consumers. Another equation involving interest rate might relate to the behaviour of investors.

It has been shown [7], [11], that in general one cannot use ordinary regression methods to estimate the matrices $B_{yy}$ and $\Gamma_{yz}$ and the parameters of an assumed distribution of the disturbances. Mann and Wald [9], for a special class of systems, and Koopmans, Rubin, and Leipnik [11], in a more general case, have obtained maximum likelihood estimates of all of the parameters for the case of the $\epsilon_t$ having a normal multivariate distribution.

Since $B_{yy}$ is non-singular, we can rewrite (2.1) in a different form, called the *reduced form*,

$$y'_t = -B_{yy}^{-1}\Gamma_{yz}z'_t + B_{yy}^{-1}\epsilon'_t, \tag{2.2}$$

or as

$$y'_t = \Pi_{yz}z'_t + \eta'_t \tag{2.3}$$

where

$$\Pi_{yz} = -B_{yy}^{-1}\Gamma_{yz}, \tag{2.4}$$

$$\eta'_t = B_{yy}^{-1}\epsilon'_t. \tag{2.5}$$

If $\epsilon_t$ has a normal distribution, so does $\eta_t$. For a given $t$ then, we can consider the model as specifying a distribution of $y_t$ with conditional expected value $z_t\Pi'_{yz}$.

It is clear that we can multiply (2.1) on the left by any non-singular matrix and obtain a system of equations which defines the same distribution of $y_t$. On the other hand, it has been shown that the only transformations of $(B_{yy}\Gamma_{yz})$ which preserve the linearity of the system of equations are multiplications on the left by non-singular matrices. If there are a priori restrictions on $(B_{yy}\Gamma_{yz})$, the set of matrices which result in new coefficient matrices satisfying these restrictions is correspondingly decreased. If the set of admissible matric multipliers includes only diagonal matrices the system of structural equations is said to be *identified.* In this case only multiplication of all coefficients by a given constant is permitted.

Knowledge of the distribution of $y_t$ given $z_t$ is obviously equivalent to knowledge of $\Pi_{yz}$ in (2.3) and the distribution of $\eta_t$. When the system is identified, the matrix $B_{yy}$ and

$$\Gamma_{yz} = -B_{yy}\Pi_{yz} \tag{2.6}$$

are determined uniquely except for multiplication on the left by a diagonal matrix. Thus identification of a system is equivalent to the possibility of inferring the structural equations from knowledge of the distribution. The estimation of all coefficients of $B_{yy}$ and $\Gamma_{yz}$ has been considered in [11].

**3. A single identified equation of a complete system.** In many studies the investigator may be interested only in a specific equation of the system, say,

$$\beta_y y'_t + \gamma_z z'_t = \zeta_t, \tag{3.1}$$

where $\zeta_t$ is a scalar disturbance. The investigator may not be interested in the entire system (2.1) of which (3.1) is one component. Since a considerable amount of computation is necessary to estimate all parameters of a complete system, there arises the problem of estimating only the coefficients of a single equation. It is desirable to do this with the least possible restrictive assumptions about the part of the system which is not the selected structural equation. In order to treat the selected equation at all, we require that it is identified; that is, that there are certain restrictions on $(\beta_y, \gamma_z)$ such that no linear combination of rows of $(B_{yy}\Gamma_{yz})$ satisfies these restrictions other than a constant times $(\beta_y, \gamma_z)$. It is not necessary to assume that every component equation is identified; that is, that the entire system is identified.

We shall suppose that the restrictions imposed are that certain coefficients are zero. We can arrange the components of the vectors so that the restrictions are

$$(\beta_y, \gamma_z) = (\beta, 0, \gamma, 0), \tag{3.2}$$

where

$$\beta = (\beta^1, \cdots, \beta^H) \tag{3.3}$$

has $H$ coefficients not assumed to be zero and

(3.4) $$\gamma = (\gamma^1, \cdots, \gamma^F)$$

has $F$ coefficients not assumed to be zero.

It will be convenient to divide the $G$ components of $y_t$ into two groups (in number $H$ and $G - H$, respectively), and the $K$ components of $z_t$ into two groups (in number $F$ and $D$ respectively) according to whether or not the components enter into (3.1) with coefficients not assumed to be zero. Let

(3.5) $$y_t = (x_t, r_t),$$

(3.6) $$z_t = (u_t, v_t),$$

where

(3.7) $$x_t = (x_{t1}, \cdots, x_{tH}),$$

(3.8) $$r_t = (r_{t1}, \cdots, r_{t,G-H}),$$

(3.9) $$u_t = (u_{t1}, \cdots, u_{tF}),$$

(3.10) $$v_t = (v_{t1}, \cdots, v_{tD}).$$

Then the selected equation is

(3.11) $$\beta x_t' + \gamma u_t' = \zeta_t .$$

Now let us see how the identification is accomplished. Partitioning $\Pi_{yz}$ into $H$ and $G - H$ rows and $F$ and $D$ columns as

$$\Pi_{yz} = \begin{pmatrix} \Pi_{xu}^* & \Pi_{xv} \\ \Pi_{ru} & \Pi_{rv} \end{pmatrix},$$

we can write the reduced form (2.3) as

(3.12) $$x_t' = \Pi_{xu}^* u_t' + \Pi_{xv} v_t' + \delta_t',$$

(3.13) $$r_t' = \Pi_{ru} u_t' + \Pi_{rv} v_t' + \xi_t',$$

where

$$\eta_t = (\delta_t, \xi_t).$$

Multiplying the above equation with $(\beta, 0)$ we obtain

(3.14) $$\beta x_t' = \beta \Pi_{xu}^* u_t' + \beta \Pi_{xv} v_t' + \beta \delta_t' .$$

Since this must be identical to (3.11) we must have

(3.15) $$\gamma = -\beta \Pi_{xu}^* ,$$

(3.16) $$0 = -\beta \Pi_{xv} .$$

The matrices $\Pi_{xu}^*$ and $\Pi_{xv}$ are defined by the distribution of $x_t$ given $u_t$ and $v_t$ (for at least $K = D + F$ linearly independent values of $u_t, v_t$). The equation

(3.11) is identified if and only if the solution of (3.15) and (3.16) for $\beta$ and $\gamma$ is unique except for a constant of proportionality. This depends on the rank of $\Pi_{xv}$ being $H - 1$. Thus a necessary and sufficient condition that (3.11) is identified is that the rank of $x_t$ on $v_t$ be $H - 1$. In particular this implies that the number of coordinates of $v_t$ (the number of zero coefficients in $\gamma_s$) be at least $H - 1$. It can easily be shown that this condition is equivalent to requiring that the rank of the matrix obtained by selecting the $G - H$ columns of $B_{yy}$ and the $D$ columns of $\Gamma_{yz}$ corresponding to the coefficients assumed zero in the selected equation is $G - 1$. This is the condition given by Koopmans and Rubin [11]. Other homogenous linear restrictions can be put in this form.

If the vector $\epsilon_t$ is normally distributed with mean zero the vector $\eta_t$ is normally distributed with mean zero. Let the covariance matrix of $\delta_t$ be $\Omega_{xx}$. Then the variance of $\zeta_t = \beta\delta_t'$ is

$$\sigma^2 = \beta\Omega_{xx}\beta'. \tag{3.17}$$

The constant of proportionality in $\beta$ may be determined by setting the variance of $\zeta_t$, $\sigma^2$, $= 1$; another normalization is

$$\beta^i = 1, \tag{3.18}$$

where $\beta^i$ is the $i$th coordinate of $\beta$. In general the normalization can be written as

$$\beta\Phi_{xx}\beta' = 1, \tag{3.19}$$

where $\Phi_{xx}$ can be either a known constant or can be a known function of unknown parameters.

As an estimation procedure for $\beta$ and $\gamma$ and $D = H - 1$, M. A. Girschick suggested in an unpublished note that one solve equations (3.15) and (3.16) with $(\Pi_{xu}^*, \Pi_{xv})$ replaced by $(P_{xu}^* P_{xv})$, the sample regression of $x$ on $u$ and $v$. By these means Girschick found confidence regions (see section 8) for the parameters of a two equation system. A similar idea lies behind a method of O. Reiersøl [10].

The present paper develops a method for handling the case of $D \geq H$. In this case the rank of $P_{xv}$ is usually $H$, thus giving no admissible estimate of $\beta$. The proposed method follows the approach used in discriminant problems.

In a second paper [2] the present authors shall give asymptotic properties of these estimates that give a certain justification for the use of them. Under very general assumptions concerning the $v_t$ and the $\epsilon_t$ we prove that these estimates are consistent. These hypotheses permit the investigator to neglect some predetermined variables absent from his particular equation. Alternative assumptions include the case of the other $G - 1$ equations being non-linear. Finally, it is shown that the estimates are asymptotically normally distributed. For this result it is not necessary to assume that the disturbances are normally distributed, or even that they have identical distributions.

**4. A description of the estimation procedure.** In a sense the dependence of the endogenous variables $x_t$ on the predetermined variables $u_t$ and $v_t$ is given by the matrix $(\Pi^*_{xu}\ \Pi_{xv})$ of regression coefficients of $x_t$ on $u_t$ and $v_t$. The interdependence of the coordinates of $x_t$ indicated by the selected equation nullifies the dependence on $v_t$; that is,

$$\beta\Pi_{xv} = 0. \tag{4.1}$$

Suppose we wish to estimate $\beta$ and $\gamma$ from a sample of $T$ observations: $(x_1, z_1), (x_2, z_2), \cdots (x_T, z_T)$. The information we need can be summarized in the second order moment matrices

$$M_{xx} = \frac{1}{T}\sum_{t=1}^{T} x'_t x_t, \tag{4.2}$$

$$M_{xz} = (M_{xu} M_{xv}) = \frac{1}{T}\left(\sum_{t=1}^{T} x'_t u_t \sum_{t=1}^{T} x'_t v_t\right), \tag{4.3}$$

$$M_{zz} = \begin{pmatrix} M_{uu} & M_{uv} \\ M_{vu} & M_{vv}\end{pmatrix} = \frac{1}{T}\begin{pmatrix} \sum_{t=1}^{T} u'_t u_t & \sum_{t=1}^{T} u'_t v_t \\ \sum_{t=1}^{T} v'_t u_t & \sum_{t=1}^{T} v'_t v_t\end{pmatrix}. \tag{4.4}$$

Since one coordinate of $u_t$ may be unity there is no advantage in taking these moments about the mean. We shall find it more convenient to use instead of $v_t$ the part of $v_t$ that is orthogonal to $u_t$; that is, we shall use

$$s'_t = v'_t - M_{vu}M^{-1}_{uu}u'_t. \tag{4.5}$$

The moments are then $M_{xx}$, $M_{xu}$, $M_{uu}$,

$$M_{xs} = M_{xv} - M_{xu}M^{-1}_{uu}M_{uv}, \tag{4.6}$$

and

$$M_{ss} = M_{vv} - M_{vu}M^{-1}_{uu}M_{uv}. \tag{4.7}$$

We can express the reduced form as

$$x'_t = \Pi_{xu}u'_t + \Pi_{xs}s'_t + \delta'_t, \tag{4.8}$$

where

$$\begin{aligned} \Pi_{xu} &= \Pi^*_{xu} + \Pi_{xv}M_{vu}M^{-1}_{uu}, \\ \Pi_{xs} &= \Pi_{xv}. \end{aligned} \tag{4.9}$$

An estimate of $\Pi_{xs}$ is the regression of $x$ on $s$,

$$P_{xs} = M_{xs}M^{-1}_{ss}. \tag{4.10}$$

To estimate $\beta$ we take the $\beta$ that makes $\beta P_{xs}$ smallest in the metric determined by the moment matrix of the residuals

$$W_{xx} = M_{xx} - P_{xs}M_{ss}P'_{xs} - P_{xu}M_{uu}P'_{xu}, \tag{4.11}$$

where

$$P_{xu} = M_{xu}M_{uu}^{-1}. \tag{4.12}$$

This is the natural generalization of least squares; the greatest weight is given to the component with least variance. This estimate is the vector satisfying

$$(P_{xs}M_{ss}P'_{xs} - \nu W_{xx})b' = 0 \tag{4.13}$$

which is associated with the smallest root of

$$| P_{xs}M_{ss}P'_{xs} - \nu W_{xx} | = 0. \tag{4.14}$$

This is normalized and the estimate of $\gamma$ is $-bP_{xu}$.

In section 5 we derive these estimates by the method of maximum likelihood under certain assumptions. Although it is assumed that the disturbances are normally distributed for this derivation, the estimates can be used in more general situations. This theory is in one sense a special case of the theory of estimating a matrix of means of a given dimensionality which is an extension of the discriminant function theory [5]. For an application of this method of estimation see [6].

**5. Derivation of maximum likelihood estimates.** We derive the estimates of $\beta$, $\gamma$, and $\sigma^2$ under the following assumptions:

ASSUMPTION A. *The selected structural equation*

$$\beta x'_t + \gamma u'_t = \zeta_t \tag{3.11}$$

*is one equation of a complete linear system of G stochastic equations. The equation is identified by the fact that if H is the number of coordinates in $x_t$ there are at least $H - 1$ coordinates in $v_t$, the vector of predetermined variables not in* (3.11) *but in the system.*

ASSUMPTION B. *At time t all of the coordinates of $z_t = (u_t, v_t)$ are given.*

ASSUMPTION C. *The coordinates of $z_t$ are given functions of exogenous variables and of coordinates of $y_{t-1}, y_{t-2}, \cdots$. If coordinates of $y_0, y_{-1}, \cdots$ are involved in $z_t$, they will be considered as given numbers. The moment matrix $M_{ss}$ is non-singular with probability one.*

ASSUMPTION D. *The disturbance vectors $\delta_t$ are distributed serially independently and normally with mean zero and covariance matrix $\Omega_{xx}$.*

We shall consider normalizations (3.19) where $\Phi_{xx}$ may be a function of other parameters, but

$$\partial\Phi_{xx}/\partial\beta = 0. \tag{5.1}$$

We can state the results in a theorem:

THEOREM 1. *Under assumptions A, B, C, and D the maximum likelihood estimate of $\beta$ is*

$$\hat{\beta} = b/\sqrt{b\hat{\Phi}_{xx}b'}, \tag{5.2}$$

*where b is the solution of*

$$(4.13)\qquad (P_{xs}M_{ss}P'_{xs} - \nu W_{xx})b' = 0$$

*corresponding to the smallest value of* $\nu$ *and* $P_{xs}$ *is defined by* (4.10), $M_{ss}$ *by* (4.6), *and* $W_{xx}$ *by* (4.11). *An estimate of* $\gamma$ *based on the maximum likelihood estimate* $\hat{\Pi}_{xu}$ *is given by*

$$(5.3)\qquad \hat{\gamma} = -\hat{\beta}P_{xu},$$

*where* $P_{xu}$ *is given by* (4.12). *The estimate of* $\sigma^2$ *is*

$$(5.4)\qquad \hat{\sigma}^2 = (1 + \nu)/b\hat{\Phi}_{xx}b'$$

*if*

$$(5.5)\qquad bWb' = 1.$$

We apply the method of maximum likelihood to

$$(5.6)\qquad L = (2\pi)^{-\frac{1}{2}TH}\,|\,\Omega_{xx}^{-1}\,|^{\frac{1}{2}T}\exp\left\{-\tfrac{1}{2}\sum_{t=1}^{T}(x_t - z_t\Pi'_{xz})\Omega_{xx}^{-1}(x'_t - \Pi_{xz}z'_t)\right.$$

under the restrictions (4.1) and (3.19). Replacing $v_t$ by $s_t$ and adding (4.1) and (3.19) multiplied by Lagrange multipliers $\lambda$ (a vector of $D$ coordinates) and $\phi$ respectively to the logarithm of $L$ we obtain after division by $T$

$$(5.7)\qquad \begin{aligned} A = {} & -\tfrac{1}{2}H\log 2\pi + \tfrac{1}{2}\log|\,\Omega_{xx}^{-1}\,| + \beta\Pi_{xs}\lambda' + \phi(\beta\Phi_{xx}\beta' - 1) \\ & - \frac{1}{2T}\sum_{t=1}^{T}(x_t - u_t\Pi'_{xu} - s_t\Pi'_{xs})\Omega_{xx}^{-1}(x'_t - \Pi_{xu}u'_t - \Pi_{xs}s'_t). \end{aligned}$$

Differentiating (5.7) with respect to $\beta$, we obtain

$$(5.8)\qquad \frac{\partial A}{\partial\beta} = \Pi_{xs}\lambda' + 2\phi\Phi_{xx}\beta'.$$

Setting this equal to zero and multiplying by $\beta$, we have

$$\beta\Pi_{xs}\lambda' + 2\phi\beta\Phi_{xx}\beta' = 0.$$

By virtue of (4.1) and (3.19), the Lagrange multiplier $\phi$ must be zero. Hence, as far as the derivarives of (5.7) are concerned the restriction (3.19) does not enter. The setting of the derivatives of (5.7) equal to zero and (4.1) will define $\hat{\beta}$ except for a constant of proportionality which is finally determined by (3.19). For convenience in deriving the estimates we shall use the normalization

$$(5.9)\qquad \beta\Omega_{xx}\beta' = 1.$$

The derivatives of (5.7) with respect to the coordinates of $\Omega_{xx}$, $\Pi_{xu}$, $\Pi_{xs}$, and $\beta$ are set equal to zero, resulting in

$$(5.10)\qquad \begin{aligned} \hat{\Omega}_{xx} = {} & M_{xx} - M_{xs}\hat{\Pi}'_{xs} - M_{xu}\hat{\Pi}'_{xu} - \hat{\Pi}_{xs}M_{sx} \\ & - \hat{\Pi}_{xu}M_{ux} + \hat{\Pi}_{xu}M_{uu}\hat{\Pi}'_{xu} + \hat{\Pi}_{xs}M_{ss}\hat{\Pi}'_{xs}, \end{aligned}$$

(5.11) $$\hat{\Omega}_{xx}^{-1}(M_{xs} - \hat{\Pi}_{xs}M_{ss}) + \hat{\beta}'\hat{\lambda} = 0,$$

(5.12) $$\hat{\Omega}_{xx}^{-1}(M_{xu} - \hat{\Pi}_{xu}M_{uu}) = 0,$$

(5.13) $$\hat{\Pi}_{xs}\hat{\lambda}' = 0.$$

Solving (5.12) for $\hat{\Pi}_{xu}$, we obtain

(5.14) $$\hat{\Pi}_{xu} = P_{xu}$$

defined by (4.12). Solving (5.11) for $\hat{\Pi}_{xs}$, we obtain

(5.15) $$\hat{\Pi}_{xs} = P_{xs} + \hat{\Omega}_{xx}\hat{\beta}'\hat{\lambda}M_{ss}^{-1}.$$

Multiplying (5.15) by $\hat{\beta}$ and solving for $\hat{\lambda}$, we obtain

(5.16) $$\hat{\lambda} = -\hat{\beta}P_{xs}M_{ss}.$$

Substitution into (5.15) gives

(5.17) $$\hat{\Pi}_{xs} = (I - \hat{\Omega}_{xx}\hat{\beta}'\hat{\beta})P_{xs}.$$

In view of (5.14) and (5.17) we can write (5.10) as

(5.18) $$\hat{\Omega}_{xx} = W_{xx} + \hat{\Omega}_{xx}\hat{\beta}'\hat{\beta}P_{xs}M_{ss}P'_{xs}\hat{\beta}'\hat{\beta}\hat{\Omega}_{xx}.$$

Let

(5.19) $$\hat{\beta}P_{xs}M_{ss}P'_{xs}\hat{\beta}' = \mu.$$

Then multiplication of (5.18) on the right by $\hat{\beta}'$ with use of (5.9) gives

$$\begin{aligned}\hat{\Omega}_{xx}\hat{\beta}' &= W_{xx}\hat{\beta}' + \hat{\Omega}_{xx}\hat{\beta}'\hat{\beta}P_{xs}M_{ss}P'_{xs}\hat{\beta}' \\ &= W_{xx}\hat{\beta}' + \mu\hat{\Omega}_{xx}\hat{\beta}',\end{aligned}$$

that is,

(5.20) $$\hat{\Omega}_{xx}\hat{\beta}' = \frac{1}{1-\mu} W_{xx}\hat{\beta}'.$$

Equation (5.13) can be written as

(5.21) $$P_{xs}M_{ss}P'_{xs}\hat{\beta}' - \mu\hat{\Omega}_{xx}\hat{\beta}' = 0$$

by substitution from (5.16), (5.17) and (5.19). Combining (5.20) and (5.21) we obtain

(5.22) $$(P_{xs}M_{ss}P'_{xs} - \nu W_{xx})\hat{\beta}' = 0,$$

where

(5.23) $$\nu = \mu/(1 - \mu).$$

For (5.22) to have a solution, $\nu$ must be a root of

(4.14) $$| P_{xs}M_{ss}P'_{xs} - \nu W_{xx} | = 0.$$

Substituting from (5.20) into (5.18) we obtain

(5.24) $$\hat{\Omega}_{xx} = W_{xx} + \mu\left(\frac{1}{1-\mu}\right)^2 W_{xx}\hat{\beta}'\hat{\beta}W_{xx} = W_{xx} + \nu(1+\nu)W_{xx}\hat{\beta}'\hat{\beta}W_{xx}.$$

To determine which root of (4.14) to use we shall compute the value of the likelihood function when these estimates are used. It will be convenient to use the solution $b$ of (4.13) with normalization (5.5). Thus $b$ is proportional to $\hat{\beta}$; in fact, since

$$\hat{\beta}\hat{\Omega}_{xx}\hat{\beta}' = \frac{1}{1-\mu}\hat{\beta}W_{xx}\hat{\beta}'$$

from (5.20), we see that

$$\hat{\beta} = b\sqrt{1-\mu} = b/\sqrt{1+\nu}.$$

Let the other solutions of (4.13) be $b_2, \cdots, b_H$, with corresponding roots $\nu_2, \cdots, \nu_H$, and

$$B^* = \begin{pmatrix} b \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_H \end{pmatrix}.$$

Since

(5.25) $$|\hat{\Omega}_{xx}| = |W_{xx} + \nu W_{xx}b'bW_{xx}|,$$

we have

(5.26) $$|B^*||\hat{\Omega}_{xx}||B^{*\prime}| = |I + \nu B^*W_{xx}b'bW_{xx}B^{*\prime}|.$$

Since

$$bW_{xx}B^{*\prime} = (1, 0, \cdots, 0),$$

and since

$$|B^*|^2 = |W_{xx}|^{-1},$$

we deduce from (5.26)

$$|\hat{\Omega}_{xx}| = |W_{xx}|(1+\nu).$$

Multiplying (5.10) by $\hat{\Omega}_{xx}^{-1}$, taking the trace, and substituting in (5.6) we obtain

(5.27) $$\hat{L} = (2\pi e)^{-\frac{1}{2}TH}|W_{xx}|^{-\frac{1}{2}T}(1+\nu)^{-\frac{1}{2}T}.$$

This is a maximum if $\nu$ is the smallest root of (4.14).

The theorem now results. The expression for $\hat{\sigma}^2$ follows from

$$\hat{\sigma}^2 = \hat{\beta}\hat{\Omega}_{xx}\hat{\beta}' = b\hat{\Omega}_{xx}b'/b\hat{\Phi}_{xx}b'.$$

If $\Phi_{xx}$ is a known constant matrix, $\hat{\Phi}_{xx} = \Phi_{xx}$; if $\Phi_{xx}$ is a function of the parameters, $\hat{\Phi}_{xx}$ is the same function of the estimates.

If we define

$$\hat{\gamma} = -\hat{\beta}\hat{\Pi}^*_{xu}, \tag{5.28}$$

we have by (4.9)

$$\hat{\gamma} = -\hat{\beta}(\hat{\Pi}_{xu} - \hat{\Pi}_{xs}M_{su}M_{uu}^{-1}). \tag{5.29}$$

Since $\hat{\beta}$ annihilates $\hat{\Pi}_{xs}$, (5.3) results.
The estimate of $\Pi_{xv}$ is given by (5.17) and the estimate of $\Omega_{xx}$ is

$$\hat{\Omega}_{xx} = W_{xx} + \nu W_{xx}b'bW_{xx}. \tag{5.30}$$

**6. The likelihood ratio test of restrictions.** It has been assumed that the selected structural equation is identified by imposing the restrictions that certain coefficients are zero. It was noted in Section 3 that at least $G - 1$ such restrictions are necessary. If $D$, the number of restrictions on the predetermined variables, is more than $H - 1$, we can test the hypothesis that these $D$ coefficients are zero against the alternative that only a smaller number are zero. This is equivalent to a test that $\Pi_{xv}$ is of rank $H - 1$ against the alternative that the rank is $H$.

It can be seen intuitively that the smallest root $\nu$ of (4.14) indicates how near $P_{xs}$ is to being singular. This statistic can be used to test the hypothesis that $\Pi_{xv}$ is of rank $H - 1$. The test is similar to the test of rank suggested by P. L. Hsu [8]. The test is stated precisely in the following theorem:

THEOREM 2. *Under assumptions A, B, C, and D the likelihood ratio criterion for testing the hypothesis that $\Pi_{xv}$ is of rank $H - 1$ against the alternative that it is of rank $H$ is*

$$(1 + \nu)^{-\frac{1}{2}T}, \tag{6.1}$$

*where $\nu$ is the smallest root of* (4.14).

PROOF. If there is no restriction on $\Pi_{xs}$, the maximum likelihood estimate of $\Pi_{xs}$ is $P_{xs}$, of $\Pi_{xu}$ is $P_{xu}$, and of $\Omega_{xx}$ is $W_{xx}$. Then the likelihood function is

$$(2\pi e)^{-\frac{1}{2}TH} \mid W_{xx} \mid^{-\frac{1}{2}T}. \tag{6.2}$$

The ratio between this and the likelihood function (5.27) maximized under the hypothesis that the rank of $\Pi_{xv}$ is $H - 1$ is (6.1).

It is proved in the paper following the present one that under certain conditions (more general than those of Theorem 2)

$$-2 \log [(1 + \nu)^{-\frac{1}{2}T}] = T \log (1 + \nu) \tag{6.3}$$

is distributed asymptotically as $\chi^2$ with $D - H + 1$ degrees of freedom. Thus an approximate test of significance is given by comparing (6.3) with a significance point of the $\chi^2$-distribution with degrees of freedom equal to the excess number of coefficients required to be zero (i.e., the number beyond the minimum required for identification).

**7. Computational procedure.** The estimation procedure in sections 4 and 5 does not indicate the most efficient method for computing those estimates. The procedure given here is believed to be efficient for ordinary computational equipment and can easily be adapted for sequence-controlled computing machines.

Let us see what expressions occur in the estimation procedure for $\beta$ and $\gamma$. We find that we must first know $P_{xs}M_{ss}P'_{xs}$, $W_{xx}$, and $P_{xu}$; these will suffice if $\Phi_{xx}$ is constant or $\Omega_{xx}$ to estimate $\beta$, $\gamma$, and $\sigma^2$. In what follows, we shall assume the normalization is $\beta^1 = 1$, as the results for other normalizations follow immediately. Examining the estimation equations, we see that we may use any matrices proportional to the moment matrices. If equation (3.11) has a constant term, it is better to use moments about the mean and estimate the constant term by setting the calculated mean of the disturbances equal to zero. One possible method of correcting for the mean is to calculate

$$m^*_{pq} = T \sum_{t=1}^{T} p_t q_t - \left(\sum_{t=1}^{T} p_t\right)\left(\sum_{t=1}^{T} q_t\right) . \tag{7.1}$$

The estimation procedure for $\beta$, $\sigma^2$, and the remainder of $\gamma$ is not affected by correcting for the mean. The computational procedure indicated here is unchanged except for a factor of proportionality in the equation for $\sigma^2$ if a different form of correction for the mean is used.

7.1. *Calculation of $M_{xz}M_{zz}^{-1}M_{zx}$ and $W_{xx}$.* It is known that

$$W_{xx} = M_{xx} - M_{xz}M_{zz}^{-1}M_{zx} . \tag{7.2}$$

We shall use (7.2) to compute $W_{xx}$. We shall compute $M_{xz}M_{zz}^{-1}M_{zx}$ by the method given by Dwyer [4]. Let us denote the element in the $i$th row and $j$th column of $M_{zz}$ by $a_{ij}$, and the element in the $i$th row and $j$th column of $M_{zx}$ by $b_{ij}$. Let us construct the following array

$$\begin{array}{llllll}
c_{11}c_{12} & \cdots & c_{1K} & e_{11}\; e_{12} & \cdots & e_{1H} \\
d_{11}d_{12} & \cdots & d_{1K} & f_{11}\; f_{12} & \cdots & f_{1H} \\
\quad c_{22} & \cdots & c_{2K} & e_{21}\; e_{22} & \cdots & e_{2H} \\
\quad d_{22} & \cdots & d_{2K} & f_{21}\; f_{22} & \cdots & f_{2H} \\
 & \cdots & & & \cdots & \\
 & & c_{KK} & e_{K1}e_{K2} & \cdots & e_{KH} \\
 & & d_{KK} & f_{K1}f_{K2} & \cdots & f_{KH}
\end{array}$$

where

$$c_{ij} = a_{ij} - \sum_{k<i} d_{ki}c_{kj}, \qquad 1 \le i \le j \le k,$$

$$e_{ij} = b_{ij} - \sum_{k<i} d_{ki}e_{ki}, \qquad 1 \le i \le k, 1 \le j \le H$$

$$d_{ij} = \frac{c_{ij}}{c_{ii}}, \qquad 1 \le i \le j \le K,$$

$$f_{ij} = \frac{e_{ij}}{c_{ii}}, \qquad 1 \le i \le K, 1 \le j \le H.$$

Then the element in the $i$th row and $j$th column of the symmetric matrix $M_{xz}M_{zz}^{-1}M_{zx}$ is

$$\sum_{k=1}^{K} e_{ki} f_{kj} .$$

If we wish to estimate several equations in the system by this method, this step need only be done once, as $M_{xz}M_{zz}^{-1}M_{zx}$ and $W_{xx}$ do not depend upon the equation (except that $x$ would be enlarged).

7.2. *Computation of* $P_{xu}$. We shall compute $P_{xu}$ by the abbreviated Doolittle method. Let us now denote the element in the $i$th row and $j$th column of $M_{uu}$ by $a_{ij}$, of $M_{ux}$ by $b_{ij}$. Then let us perform the previous operations, not including the last step. We may arrange the work, if only one equation is to be estimated, so that this is already done. Then define

$$g_{ij} = f_{ij} - \sum_{i<k\leq F} d_{ik} g_{kj}, \qquad 1 \leq i \leq F, 1 \leq j \leq H.$$

Then the element in the $i$th row and $j$th column of $P_{xu}$ is $g_{ji}$.

7.3. *Computation of* $P_{xs}M_{ss}P'_{xs}$. We know that

$$P_{xs}M_{ss}P'_{xs} = M_{xz}M_{zz}^{-1}M_{zx} - M_{xu}M_{uu}^{-1}M_{ux} . \tag{7.3}$$

Let us compute $P_{xs}M_{ss}P'_{xs}$, using (7.3). We must first calculate $M_{xu}M_{uu}^{-1}M_{ux}$. We may do this either by the method of section 7.1, or as $P_{xu}M_{ux}$.

7.4. *Computation of* $\nu$, $\hat{\beta}$, *and* $\hat{\gamma}$. We shall use

$$\hat{\gamma} = -\hat{\beta} P_{xu} \tag{5.3}$$

to compute $\hat{\gamma}$ after has$\hat{\beta}$ been computed.

Case 1) $H = 1$. In this case the vector $\hat{\beta} = (1)$, $\nu = P_{xs}M_{ss}P'_{xs}/W_{xx}$.

Case 2) $H = 2, D > 1$. Let $a_{ij}$ denote the element in the $i$th row and $j$th column of $P_{xs}M_{ss}P'_{xs}$, $w_{ij}$ the element in the $i$th row and $j$th column of $W_{xx}$. Define

$$k_0 = | P_{xs}M_{ss}P'_{xs} | ,$$

$$k_1 = | W_{xx} |$$

$$k_2 = \tfrac{1}{2}(a_{11}w_{22} + a_{22}w_{11} - 2a_{12}w_{12}).$$

Then

$$\nu = \frac{k_2 - \sqrt{k_2^2 - k_0 k_1}}{k_1} .$$

Let $\hat{\Theta} = P_{xs}M_{ss}P'_{xs} - \nu W_{xx}$. Then

$$\hat{\beta}^1 = 1,$$

$$\hat{\beta}^2 = -\frac{\hat{\theta}_{11}}{\hat{\theta}_{12}} = -\frac{\hat{\theta}_{21}}{\hat{\theta}_{22}} .$$

Case 3) $H = 2, D = 1$. In this case $\nu = 0$. Then $\hat{\Theta} = P_{xs}M_{ss}P'_{xs}$, and $\hat{\beta}$ may be computed as before.

Case 4) $H > 2, D > H - 1$. Using the procedure of section 7.2, compute $A = (P_{xs}M_{ss}P)'^{-1}_{xs}W_{xx}$. Let us multiply equation (5.22) by $-\frac{1}{\nu}(P_{xs}M_{ss}P'_{xs})^{-1}$, and set $1/\nu = \lambda$. We obtain

$$(A - \lambda I)\hat{\beta}' = 0, \tag{7.4}$$

where $\lambda$ is the largest characteristic root of $A$. Then we may employ the method of Aitken [1] to estimate $\lambda$ and $\hat{\beta}$. Let $q_0$ be an approximation to $\hat{\beta}$. The column of $A$ with largest absolute values is generally a satisfactory approximation. Define

$$q'_i = Aq'_{i-1},$$

$$\lambda_{ij} = \frac{q_i^j}{q_{i-1}^j}.$$

The quantities $\lambda_{ij}$ approach $\lambda$ as $i$ increases, and the normalized vectors $q_i$ approach $\hat{\beta}$. The convergence may be accelerated by the methods given by Aitken. The normalization should not be carried out until the $\lambda_{ij}$ are sufficiently close for different $j$.

Case 5) $H > 2, D = H - 1$. Let us go through the procedure of section 7.2 with $A = P_{xs}M_{ss}P'_{xs}$, and with no matrix $B$. Then $c_{HH} = 0$. Set $g_H = 1$, and compute

$$g_i = -\sum_{i<k\leq H} d_{ik}\, g_k,$$

Then

$$\beta^i = \frac{g_i}{g_1}, \qquad \nu = 0.$$

7.5. *Computation of* $\hat{\sigma}^2$. We have

$$\hat{\sigma}^2 = \hat{\beta}\Omega_{xx}\hat{\beta}' = (1 + \nu)\hat{\beta}W_{xx}\hat{\beta}'. \tag{7.5}$$

If we use the $m^*$'s instead of the $m$'s, we must divide by $T^2$, and if other factors of proportionality are used, we must divide by them. $\sigma^2$ is in general biased, but the bias depends upon the nature of the complete system, and is not easy to calculate. The bias is of the order of $1/T$.

**8. Confidence regions based on small sample theory.**[5] If all of the predetermined variables in the system are exogenous (i.e., "fixed"), we can obtain confidence regions for the coefficients of one equation on the basis of small sample theory. To do this we require only that the disturbance of the selected equation be normally distributed; that is, the linear form in the observations $\beta x'_t + \gamma u'_t$

[5] We are indebted to Professor A. Wald for assistance in simplifying our approach to this problem.

is normally distributed with mean zero and variance $\sigma^2$. The regression of this on fixed variates is normally distributed and certain quadratic forms in these linear forms have $\chi^2$-distributions. On the basis of this we can set up confidence regions for the coefficients.

In addition to assumptions A and B we use the following:

ASSUMPTION E. *All of the coordinates of $z_t = (u_t\ v_t)$ are exogenous. The moment matrix $M_{ss}$ is non-singular. The disturbances of the selected equation are distributed independently and normally with mean 0 and variance $\sigma^2$.*

Suppose we have a set of observations $(x_1, u_1, v_1), \cdots (x_T, u_T, v_T)$. If we know $\beta$ and $\gamma$ we can obtain $T$ values of

$$w_t = \beta x_t' + \gamma u_t', \qquad t = 1, \cdots, T. \tag{8.1}$$

The sample regression coefficients of $w_t$ on $u_t$ and $s_t$ are

$$c = \frac{1}{T}\sum_{t=1}^{T} w_t u_t M_{uu}^{-1} = \beta M_{xu} M_{uu}^{-1} + \gamma, \tag{8.1}$$

$$e = \frac{1}{T}\sum_{c=1}^{T} w_t s_t M_{ss}^{-1} = \beta M_{xs} M_{ss}^{-1}. \tag{8.3}$$

The two vectors $c$ and $e$ are distributed independently and normally with mean 0 and covariance matrices

$$\mathcal{E}(c'c) = \sigma^2 M_{uu}^{-1}, \tag{8.4}$$

$$\mathcal{E}(e'e) = \sigma^2 M_{ss}^{-1}, \tag{8.5}$$

Hence (by usual regression theory)

$$C = \frac{1}{\sigma^2} cM_{uu}c' = \frac{1}{\sigma^2}[\beta M_{xu}M_{uu}^{-1}M_{ux}\beta' + \beta M_{xu}\gamma' + \gamma M_{ux}\beta' + \gamma M_{uu}\gamma], \tag{8.6}$$

$$E = \frac{1}{\sigma^2} eM_{ss}e' = \frac{1}{\sigma^2}\beta M_{xs}M_{ss}^{-1}M_{sx}\beta' \tag{8.7}$$

$$= \frac{1}{\sigma^2}\beta(M_{xv} - M_{xu}M_{uu}^{-1}M_{uv})(M_{vv} - M_{vu}M_{uu}^{-1}M_{uv})^{-1}(M_{vx} - M_{vu}M_{uu}^{-1}M_{ux})\beta',$$

$$A = \frac{1}{\sigma^2}\left(\frac{1}{T}\sum_{t=1}^{T} w_t^2 - C - E\right) = \frac{1}{\sigma^2}\beta \dot{W}_{xx}\beta', \tag{8.8}$$

are distributed independently as $\chi^2$ with $F$, $D$, and $T - K$ degrees of freedom, respectively. The ratio of any two has an $F$-distribution.

On the basis of these considerations we can obtain the desired confidence regions.

THEOREM 3. *Suppose assumptions A, B, and E are true. If the normalization is*

$$\beta\Phi_{xx}\beta' = 1. \tag{8.9}$$

*where $\Phi_{xx}$ is a given matrix, (a) a confidence region for $\beta$ of confidence $\epsilon$ consists of all $\beta^*$ satisfying* (8.9) *and*

$$\frac{\beta^* M_{xs} M_{ss}^{-1} M_{sx} \beta^{*\prime}}{\beta^* W_{xx} \beta^{*\prime}} \cdot \frac{T-K}{D} \le F_{D,T-K}(\epsilon), \tag{8.10}$$

*where $F_{D,T-K}(\epsilon)$ is chosen so the probability of* (8.10) *for $\beta^* = \beta$ is $\epsilon$. (b). A confidence region for $\beta$ and $\gamma$ simultaneously consists of all $\beta^*$ and $\gamma^*$ satisfying* (8.9) *and*

$$\frac{\beta^* M_{xu} M_{uu}^{-1} M_{ux} \beta^{*\prime} + \beta^* M_{xu} \gamma^{*\prime} + \gamma^* M_{xu} \beta^{*\prime} + \gamma^* M_{uu} \gamma^* + \beta^* M_{xs} M_{ss}^{-1} M_{sx} \beta^{*\prime}}{\beta^* W_{xx} \beta^{*\prime}} \cdot \frac{T-K}{K} \le F_{K,T-K}(\epsilon). \tag{8.11}$$

*(c) If the normalization is $\sigma^2 = 1$, then a confidence region for $\beta$ of confidence $\epsilon_1 \epsilon_2$ consists of all $\beta^*$ satisfying*

$$\beta^* M_{xs} M_{ss}^{-1} M_{sx} \beta^{*\prime} \le \chi_D^2(\epsilon_1), \tag{8.12}$$

$$\underline{\chi}^2_{T-K}(\epsilon_2) \le \beta^* W_{xx} \beta^{*\prime} \le \bar{\chi}^2_{T-K}(\epsilon_2), \tag{8.13}$$

*where $\chi_D^2(\epsilon_1)$ is chosen so that the probability of* (8.12) *is $\epsilon_1$ when $\beta^* = \beta$ and $\underline{\chi}^2_{T-K}(\epsilon_2)$ and $\bar{\chi}^2_{T-K}(\epsilon_2)$ are chosen so that the probability of* (8.13) *is $\epsilon_2$ when $\beta^* = \beta$ and*

$$\underline{\chi}^2(\epsilon_2) \le 1 \le \bar{\chi}^2(\epsilon_s). \tag{8.14}$$

*(d) A confidence region for $\beta$ an $\gamma$ simultaneously consists of all $\beta^*$ and $\gamma^*$ satisfying* (8.13) *and*

$$\beta^* M_{xu} M_{uu}^{-1} M_{ux} \beta^{*\prime} + \beta^* M_{xu} \gamma^{*\prime} + \gamma^* M_{ux} \beta^{*\prime} + \gamma^* M_{uu} \gamma^{*\prime} + \beta^* M_{xs} M_{ss}^{-1} M_{sx} \beta^{*\prime} \le \chi_K^2(\epsilon_1). \tag{8.15}$$

Region (c) is the interior of an ellipsoid and an ellipsoidal shell in the $\beta^*$-space; region (d) is similar in the $\beta^*$, $\gamma^*$-space. Region (a) consists of the intersection of the quadric surface (8.9) nd the interior of a cone in the $\beta^*$-space; region (b) is similar in the $\beta^*$, $\gamma^*$-space.

It is clear that there are many other ways of constructing confidence regions by taking regression on other fixed variates. Of these the best seem to be those of theorem 3. It has been proved [2] that the regions of theorem 3 are consistent in the sense that for sufficiently large $T$ the probability is arbitrarily near 1 that all of the confidence region is within a certain distance of $\beta$ or $\beta$, $\gamma$. For an application of this technique to economic data see a paper by Bartlett [3] who suggested this method independently.

**9. An approximate small sample test of restrictions.** When $\beta^* = \beta$, the probability of (8.10) is $\epsilon$. If $\beta^*$ is replaced by $\hat{\beta}$ which minimizes the expression on the left, the probability is at least as great; it is, say, $1 - \delta$. This ratio is $\lambda$, the smallest root of

$$\left| \frac{1}{D} M_{xz} M_{zz}^{-1} M_{zx} - \lambda \frac{T}{T - K} W_{xx} \right| = 0, \tag{9.1}$$

Since

$$\lambda = \frac{T - K}{TD} \nu, \tag{9.2}$$

where $\nu$ is the smallest root of (4.14), the probability of

$$\nu \geq \frac{TD}{T - K} F_{D,T-K}(\epsilon) \tag{9.3}$$

is $\delta \leq (1 - \epsilon)$. We summarize this as follows:

THEOREM 4. *Under assumptions A, B, and E, the inequality* (9.3), *where* $\nu$ *is the smallest root of* (4.14), *constitutes a test of the hypothesis that the coefficients of* $v_t$ *in the selected structural equation are zero of significance less than* $1 - \epsilon$.

This test is simply an approximation to the test given in section 6. The exact probability, $\delta$, of (9.3) is unknown; in fact the distribution of $\nu$ depends on $\Pi_{xv}$ and the distribution of $\delta_t$. However, since $\delta$ lies between 0 and $1 - \epsilon$, we know that if the test is used as though the level were $1 - \epsilon$, the test will be "conservative."

Another approximate test of the restrictions can be obtained from the inequality (8.11). If the hypothesis is rejected on the basis of one of these tests, the corresponding confidence region (for $\beta$ or for $\beta$ and $\gamma$) is imaginary, for all $\beta$ or $\beta$ and $\gamma$ are excluded. It should be noticed that the use of a given ratio to test the hypothesis at significance level $\delta(\leq 1 - \epsilon)$ does not affect the confidence coefficient $\epsilon$ of the confidence region when the hypothesis is true.

## REFERENCES

[1] A. C. AITKEN, "Studies in practical mathematics II. The evaluation of the latent roots and latent vectors of a matrix," *Edinb. Math. Soc. Proc.*, Vol. 57 (1936–7), pp. 269–305.

[2] T. W. ANDERSON AND HERMAN RUBIN, "The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equations," to be published.

[3] M. S. BARTLETT, "A note on the statistical estimation of demand and supply relations from time series," *Econometrica*, Vol. 16 (1948), pp. 323–329.

[4] P. S. DWYER, "Evaluation of linear forms," *Psychometrika*, Vol. 6 (1941), pp. 355–365.

[5] R. A. FISHER, "The statistical utilization of multiple measurements," *Annals of Eugenics*, Vol. 8 (1938), pp. 376–386.

[6] M. A. GIRSHICK AND T. HAAVELMO, "Statistical analysis of the demand for food: examples of simultaneous estimation of structural equations," *Econometrica*, Vol. 15 (1947), pp. 79–110.

[7] T. HAAVELMO, "Statistical implications of a system of simultaneous equations," *Econometrica*, Vol. 11 (1943), pp. 1–12.

[8] P. L. Hsu, "On the problem of rank and the limiting distribution of Fisher's test function," *Annals of Eugenics*, Vol. 11 (1941), pp. 39–41.
[9] H. B. Mann and A. Wald, "On the statistical treatment of linear stochastic difference equations," *Econometrica*, Vol. 11 (1943), pp. 173–220.
[10] Olav Reiersøl, "Confluence analysis by means of lag moments and other methods of confluence analysis," *Econometrica*, Vol. 9 (1941), pp. 1–24.
[11] *Statistical Inference in Dynamic Economic Systems*, to be published as Cowles Commission Monograph No. 10.