Cowles Foundation Paper 36b

Reprinted from THE ANNALS OF MATHEMATICAL STATISTICS
Vol. 21, No. 4, December, 1950

THE ASYMPTOTIC PROPERTIES OF ESTIMATES 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. In a previous paper [2] the authors have given a method for estimating the coefficients of a single equation in a complete system of linear stochastic equations. In the present paper the consistency of the estimates and the asymptotic distributions of the estimates and the test criteria are studied under conditions more general than those used in the derivation of these estimates and criteria. The point estimates, which can be obtained as maximum likelihood estimates under certain assumptions including that of normality of disturbances, are consistent even if the disturbances are not normally distributed and (a) some predetermined variables are neglected (Theorem 1) or (b) the single equation is in a non-linear system with certain properties (Theorem 2).

Under certain general conditions (normality of the disturbances not being required) the estimates are asymptotically normally distributed (Theorems 3 and 4). The asymptotic covariance matrix is given for several cases. The criteria derived in [2] for testing the hypothesis of over-identification have, asymptotically, χ^2-distributions (Theorem 5). The exact confidence regions developed in [2] for the case that all predetermined variables are exogenous (that is, that the difference equations are of zero order) are shown to be consistent and to hold asymptotically even when this assumption is not true (Theorem 6).

2. Introduction. The complete system of linear stochastic equations considered by the authors in [2] was written

$$\mathrm{B}_{yy}y_t' + \Gamma_{yz}z_t' = \epsilon_t' , \tag{2.1}$$

where y_t is a row vector of G jointly dependent variables at "time" t, z_t is a row vector of K variables predetermined at t, and ϵ_t is a row vector of "disturbances," and B_{yy} and Γ_{yz} are matrices. If B_{yy} is non-singular the distribution of ϵ_t induces the distribution of y_t given z_t.

One component equation of (2.1) was given special treatment. Let β be

---

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

[2] The results of this paper were presented to meetings of the Institute of Mathematical Statistics at Washington, D. C., April 12, 1946 (Washington Chapter) and at Ithaca, New York, August 23, 1946. Most of the research was done at the Cowles Commission for Research in Economics; the authors are indebted to the members of the Cowles Commission staff for many helpful discussions.

[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.

composed of the coefficients of the coordinates of y_t which are not assumed zero in the specified equation, and let x_t be composed of the corresponding components of y_t; similarly let γ be composed of the coefficients of the coordinates of z_t which are not assumed zero, and u_t the corresponding components of z_t; and let ζ_t be the component of ϵ_t associated with the specified equation. Then the single equation is

$$\beta x_t' + \gamma u_t' = \zeta_t . \tag{2.2}$$

Suppose we have a set of observations $x_t, z_t, t = 1, \cdots, T$. For sets of any two vectors a_t and b_t, let the second-order moment matrix be

$$M_{ab} = \frac{1}{T}\sum_{t=1}^{T} a_t' b_t . \tag{2.3}$$

Let s_t be some linear transform of v_t, the set of coordinates of z_t not contained in u_t, chosen so $M_{su} = 0$. Defining

$$W_{xx} = M_{xx} - M_{xz}M_{zz}^{-1}M_{zx} , \tag{2.4}$$

and assuming ϵ_t normally distributed with mean 0, covariance matrix Σ, and independently of $\epsilon_{t'}(t \neq t')$, we find $\hat{\beta}$, the maximum likelihood estimate of β, to be proportional to a vector defined by

$$(M_{xs}M_{ss}^{-1}M_{sx} - \nu W_{xx})b' = 0, \tag{2.5}$$

taking ν as the smallest root of

$$| M_{xs}M_{ss}^{-1}M_{sx} - \nu W_{xx} | = 0. \tag{2.6}$$

The vector is normalized by

$$\hat{\beta}\hat{\Phi}_{xx}\hat{\beta}' = 1, \tag{2.7}$$

where Φ_{xx} may be a function of the estimates of other parameters. The estimate of γ is $\hat{\gamma} = -\hat{\beta}M_{xu}M_{uu}^{-1}$ [2; Theorem 1]. These estimates were derived under the following explicit Assumptions A, B, C, and D:

Assumption A. *The selected structural equation* (2.2) *is one equation of a complete linear system of stochastic equations. It 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 in the system, but missing in* (2.2).

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_{zz} is non-singular with probability one.*

Assumption D. *The disturbance vectors ϵ_t are distributed serially independently and normally with mean zero and covariance matrix Σ_{xx}.*

Under these assumptions it is found that $(1 + \nu)^{-\frac{1}{2}T}$ is the likelihood ratio criterion for testing the hypothesis that the number of components of z_t assumed to have zero coefficients is so great.

If there are no lagged endogenous variables in z_t, we can find confidence regions for β and for β and γ simultaneously as well as an approximate test for the above hypothesis. The assumptions used for these results are A, B, and

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

Assumptions A and B are used in this paper and a number in addition, which will be lettered similarly. It is to be emphasized that the various assumptions are used alternatively, never all at once; in fact many assumptions are mutually exclusive.

3. Consistency of the estimates. The estimates $\hat{\beta}$ and $\hat{\gamma}$ are consistent not only in the case for which they are maximum likelihood estimates, but also in cases in which the disturbances are not normally or even identically distributed. Moreover, for consistency of the estimates it is not necessary that the investigator know all of the components of v_t or use them. Another direction in which the assumptions may be relaxed is to permit the other equations in the system to be non-linear.

3.1. *The linear case.* This case is characterized by Assumption A. We need also to assume:

ASSUMPTION F. *M_{zz} converges to a fixed non-singular limit R in probability.*

Let u_t consist of the part of z_t that enters the selected structural equation (22). The remainder of the components of z_t are divided into two groups as to whether they are known or not. Let c_t be a linear transform of the known components not entering the specified equation such that

$$\operatorname*{plim}_{t\to\infty} M_{uc} = 0, \tag{3.1}$$

and let r_t be a linear transform of the components of z_t not known such that

$$\operatorname*{plim}_{t\to\infty} M_{ur} = 0, \tag{3.2}$$

$$\operatorname*{plim}_{t\to\infty} M_{cr} = 0. \tag{3.3}$$

The relevant part of the "reduced form," obtained from (2.1) by multiplication by B_{yy}^{-1} is

$$x_t' = \bar{\Pi}_{xu}u_t' + \Pi_{xc}c_t' + \Pi_{xr}r_t' + \delta_t'. \tag{3.4}$$

The matrix $(\Pi_{xc}\Pi_{xr})$ is Π_{xs} (defined in [2]) multiplied on the right by a non-singular matrix; hence, $\beta\Pi_{xc} = 0$, and similarly $\beta\bar{\Pi}_{xu} = \gamma$. We shall find it convenient to assume

ASSUMPTION G. *Π_{xc} has rank $H - 1$.*

This means that for T sufficiently large the probability is arbitrarily near 1 that (2.2) is identified.

However, these conditions still do not insure consistency. We need the asymptotic analogue of lack of correlation:

ASSUMPTION H.

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \delta_t' z_t = 0.$$

We do not need to require that the covariance matrices of δ_t are the same or even that they exist. We shall make an assumption about

$$W_{zz}^{*} = M_{zz} - (M_{zu}\, M_{zc}) \begin{pmatrix} M_{uu} & M_{uc} \\ M_{cu} & M_{cc} \end{pmatrix}^{-1} \begin{pmatrix} M_{uz} \\ M_{cz} \end{pmatrix}. \tag{3.5}$$

ASSUMPTION I. *The ratio of the largest to the smallest characteristic roots of* W_{zz}^{*} *is bounded in probability.*

This means that for a suitable constant K

$$\lim_{T\to\infty} P\left(\frac{l(W_{zz}^{*})}{s(W_{zz}^{*})} > K\right) = 0, \tag{3.6}$$

where $P(E)$ denotes the probability of event E and $s(A)$ and $l(A)$ are the smallest and largest roots of the matrix A, respectively.

Assumptions F and H imply that $P_{xu} \to \bar{\Pi}_{xu}$ and $P_{xc} \to \Pi_{xc}$ in probability, where $P_{xu} = M_{xu}M_{uu}^{-1}$ and P_{xc} is the part of

$$(M_{xu}\, M_{xc}) \begin{pmatrix} M_{uu} & M_{uc} \\ M_{cu} & M_{cc} \end{pmatrix}^{-1} \tag{3.7}$$

corresponding to the vector[5] c_t. The first assertion follows because $M_{xu}M_{uu}^{-1} = (\Pi_{xu}M_{uu} + \Pi_{xc}M_{xc} + \Pi_{xr}M_{ru} + M_{\delta u})M_{uu}^{-1}$ and $M_{xc} \to 0$, $M_{ru} \to 0$, and $M_{\delta u} \to 0$ in probability by (3.1), (3.3) and Assumption H; the second assertion follows similarly. Since matrix multiplication is continuous, and the characteristic roots of a matrix are continuous functions of the matrix,[6]

$$\operatorname*{plim}_{T\to\infty} s[P_{xc}M_{ss}P_{xc}'] = 0, \tag{3.8}$$

where $M_{ss} = (M_{cc} - M_{cu}M_{uu}^{-1}M_{uc})$. This follows from the well-known theorem (a proof of which is given in [4]) that if a random vector X_T converges stochastically to X, then $f(X_T)$ converges stochastically to $f(X)$ if $f(y)$ is continuous at X.

We shall find the following lemmas convenient. The proofs are simple and have been given in [1].

LEMMA 1. *Let B be positive definite, A positive semi-definite. Then the smallest root ν of $| A - xB | = 0$ is less than or equal to $s(A)/s(B)$.*

[5] See Section 4 of [2].

[6] Because of the assertion above and Assumptions F and G only one characteristic root of the matrix approaches zero in probability.

LEMMA 2. *Each element of a positive definite matrix is less in absolute value than the largest characteristic root.*

Let ν be the smallest root of

$$|P_{xc}M_{ss}P'_{xc} - \nu W^*_{xx}| = 0. \tag{3.9}$$

Then $\operatorname{plim}_{T\to\infty} \nu W^*_{xx} = 0$. This statement follows from (3.8) and Lemmas 1 and 2.

Since 0 is a simple characteristic root of $\Pi_{xc} \operatorname{plim}_{T\to\infty} M_{ss}\Pi'_{xc}$, it follows from (3.9) and the consistency of P_{xu} and P_{xc} that $\hat{\beta}$ approaches β apart from normalization. The following theorem results directly:

THEOREM 1. *Under Assumptions* A, F, G, H, *and* I, *and if* $\operatorname{plim}_{T\to\infty} \beta\hat{\Phi}_{xx}\beta' = 1$,

$$\operatorname{plim}_{T\to\infty} \hat{\beta} = \beta, \tag{3.10}$$

$$\operatorname{plim}_{T\to\infty} \hat{\gamma} = \gamma, \tag{3.11}$$

where $\hat{\beta}$ *and* $\hat{\gamma}$ *are calculated as if* $r_t = 0$ *and as if the remainder of* A, B, C, *and* D *were satisfied.*[7]

3.2. *The non-linear case.* In this section we apply the estimates obtained in [2] to an equation of a complete system in which the remaining equations may be non-linear. We replace Assumption A by the following assumption:

ASSUMPTION J. *The selected structural equation* (2.2) *is one equation of a complete system of stochastic equations:*

$$F_i(y_t, z_t) = \epsilon_{ti} \qquad (i = 1, \cdots, G). \tag{3.11}$$

Let us solve the complete system (3.11) for the components of y_t. We obtain

$$y_{tj} = h_j(z_t, \epsilon_t). \tag{3.12}$$

Let u_t be the subvector of z_t occurring in the selected structural equation. Let c_t be a vector function of z_t such that $\operatorname{plim}_{T\to\infty} M_{cu} = 0$. We may write (3.12) for those y's occurring in the selected structural equation as

$$x'_t = \bar{\Pi}_{xu}u'_t + \Pi_{xc}c'_t + \varphi'(z_t, \epsilon_t), \tag{3.13}$$

where the components of $\varphi(z_t, \epsilon_t)$ are the residuals from the formal limiting regression of x_t on u_t and c_t. The proof of Theorem 1 can be used to prove the following:

THEOREM 2. *If Assumptions* F, G, H, I, *and* J *are satisfied with* z_t *replaced by* (u_t, c_t) *and* δ_t *replaced by* $\varphi(z_t, \epsilon_t)$, *and* $r_t = 0$, *and if* $\operatorname{plim}_{T\to\infty} \beta\hat{\Phi}_{xx}\beta' = 1$, *then*

$$\operatorname{plim}_{T\to\infty} \hat{\beta} = \beta, \tag{3.14}$$

$$\operatorname{plim}_{T\to\infty} \hat{\gamma} = \gamma. \tag{3.15}$$

---

[7] This follows from the above statements because $\hat{\beta}$ and $\hat{\gamma}$ are (vector-valued) rational functions of M_{zz}, P_{xs}, W^*_{xx} and Φ_{xx} which approach limits in probability.

4. The asymptotic distribution of the estimates.

4.1. *The asymptotic distribution of P_{zs} and P_{zu}.* To obtain the asymptotic distribution of the estimates we need stronger assumptions. Throughout Sections 4.1 and 4.2 we use Assumptions A, B, F, H, I, and the following:

ASSUMPTION K. *The exogenous variables are bounded; the vector of disturbances of the complete system has mean zero, and is serially independent; for some $\lambda > 0$ and some M, $\mathcal{E}(|\delta_{ti}|^{4+\lambda}) < M$; the coordinates of z_t may be linear combinations of lagged endogenous variables. If the endogenous part of a coordinate is*

$$\sum_{\tau=1}^{\infty} \sum_{i=1}^{G} g_{\tau i} y_{t-\tau,i},$$

then

$$\sum_{\tau=1}^{\infty} \sum_{i=1}^{G} | g_{\tau i} | < \infty$$

and

$$\sum_{\tau=t}^{\infty} \sum_{i=1}^{G} g_{\tau i} y_{t-\tau,i}$$

is bounded.

ASSUMPTION L. *The matrix Φ_{zz} is known and constant.*

ASSUMPTION M. *For each i, j, k, l, $1 \leq i, j \leq H$, $1 \leq k, l \leq K$,*

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathcal{E}(\delta_{ti} \delta_{tj} z_{tk} z_{tl}) = \kappa_{ijkl}$$

exists.

Let the components of M_{yy}, M_{yz}, M_{zz} be arranged as a vector $m(T)$ with mean value $\mu(T)$. It has been shown [3] that $\sqrt{T}(m(T) - \mu(T))$ is asymptotically distributed according to $N(0, \Sigma)$, the normal distribution with mean 0 and covariance matrix Σ composed of elements

$$\sigma_{ij} = \lim_{T \to \infty} \mathcal{E}(T[m_i(T) - \mu_i(T)]\,[m_j(T) - \mu_j(T)]).$$

In conjunction with this result we make repeated use of a special case of Theorem 6 of [4]:

Suppose $\sqrt{T}(x_{jT} - \xi_{jT})$ $(j = 1, \cdots, n)$ have the joint asymptotic distribution $N(0, \Psi)$ with ξ_{jT} being functions of T such that $\lim_{T \to \infty} \xi_{jT} = \xi_j$. Let $f_{kT}(z_1, \cdots, z_n)$ be random Borel-measurable functions of n real variables such that $\frac{\partial f_{kT}}{\partial z_j} = \alpha_{kjT}(z)$ exists with probability one for T sufficiently large and z in a fixed neighborhood of ξ, and suppose that there exist numbers α_{kj} such that for any $\epsilon > 0$, and $\lambda > 0$, $P(\sup_{(z-\xi_T)(z-\xi_T)' \leq (\lambda/T)} | \alpha_{kjT}(z) - \alpha_{kj} | > \epsilon)$ approaches zero. Then if $y_{kT} = f_{kT}(x_{1T}, \cdots, x_{nT})$ and $\eta_{kT} = f_{kT}(\xi_{1T}, \cdots, \xi_{nT})$, the random variables $\sqrt{T}(y_{kT} - \eta_{kT})$ have the joint asymptotic distribution $N(0, A\Psi A')$, where $A = (\alpha_{ij})$.

To obtain the asymptotic distributions we have only to verify that the assumptions of this statement are satisfied, and compute A, since the asymptotic distribution is characterized completely by $A\Psi A'$. We shall denote the element in the k-th row and l-th column of $A\Psi A'$ by $\sigma(f_k, f_l)$. We shall find it convenient to use the notation $df = Adx$; that is, the differential df is defined in terms of the limit matrix A.

Let

$$A = M_{\delta u}, \tag{4.1}$$

$$B = M_{\delta s}, \tag{4.2}$$

$$C = \operatorname*{plim}_{T\to\infty} M_{uu}, \tag{4.3}$$

$$E = \operatorname*{plim}_{T\to\infty} M_{ss}, \tag{4.4}$$

$$L = P_{xu}, \tag{4.5}$$

$$P = P_{xs} = M_{xs}M_{ss}^{-1}, \tag{4.6}$$

$$\Lambda = \overline{\Pi}_{xu}, \tag{4.7}$$

$$\Pi = \Pi_{xs}. \tag{4.8}$$

The matrix L is the random function $AM_{uu}^{-1} + \Pi_{xc}M_{cu}M_{uu}^{-1} + \Lambda$ of A, P is the random function $BM_{ss}^{-1} + \Pi$ of B. Then

$$dL = (dA)C^{-1}, \tag{4.9}$$

$$dP = (dB)E^{-1}. \tag{4.10}$$

However

$$\sigma(a_{ik}, a_{jl}) = \alpha_{ijkl}, \tag{4.11}$$

$$\sigma(a_{ik}, b_{jl}) = \beta_{ijkl}, \tag{4.12}$$

$$\sigma(b_{ik}, b_{jl}) = \gamma_{ijkl}, \tag{4.13}$$

where α_{ijkl}, β_{ijkl}, γ_{ijkl} are the appropriate quantities κ_{abcd}, respectively. From these we may compute $\sigma(l_{ij}, l_{kl})$, $\sigma(l_{ij}, p_{kl})$, and $\sigma(p_{ij}, p_{kl})$, the elements of the asymptotic covariance matrix of the elements of L and P (which are asymptotically normally distributed by the above). These elements can be estimated consistently from the sample (the proof follows from Theorem 1).

4.2. *The asymptotic distribution of $\hat{\beta}$ and $\hat{\gamma}$ for constant normalization.* In this section we shall show that $\hat{\beta}$ and $\hat{\gamma}$ are asymptotically normally distributed (Theorem 3). In view of the above theorem on asymptotic distributions the intricate part of the proof is in obtaining the covariance matrix. First we shall demonstrate that the elements of νW are $o(1/\sqrt{T})$ in probability. Since Assumption I holds, it is sufficient to show that $s(P_{xs}M_{ss}P'_{xs})$ is $o(1/\sqrt{T})$ in probability. This means $d \mid P_{xs}M_{ss}P'_{xs} \mid = 0$, since each of the characteristic roots of $P_{xs}M_{ss}P'_{xs}$ except the smallest approaches a non-zero limit in probability.

For any matrix A, A_{ij} denotes the matrix obtained by deleting the i-th row and j-th column from A, and $A_{ik,jl}$ is the matrix obtained by deleting the i-th and k-th rows and the j-th and l-th columns. Let

$$A^{ij} = (-1)^{i+j} | A_{ij} |,$$
$$A^{ij,kl} = (-1)^{i+j+k+l+\epsilon} | A_{ik,jl} |,$$

where $\epsilon = 0$ if $(i - k)(j - l) > 0$, 1 otherwise when $i \neq k$, $j \neq l$. $A^{ij,kl} = 0$ if $i = k$ or $j = l$. In the rest of the paper we use the summation convention of tensor calculus for lower case indices; namely, that whenever a lower case letter appears as a superscript and a subscript in an expression, the corresponding terms are to be summed on that index.

In general

$$d | A | = A^{ij} da_{ij} . \tag{4.14}$$

We may consider $P_{zs}M_{ss}P'_{zs}$ as a random function of P_{zs}. Then

$$d(i,j\text{-th element of } P_{zs}M_{ss}P'_{zs}) = \pi^k_i e_{kl}\, dp^l_j + \pi^l_j e_{kl}\, dp^k_i . \tag{4.15}$$

However

$$(\Pi_{zs}E\Pi'_{zs})^{ij} = \rho^j \beta^i = \rho^i \beta^j, \tag{4.16}$$

where ρ^i is a factor of proportionality. Since $\beta\Pi_{zs} = 0$, we have $d | P_{zs}M_{ss}P'_{zs} | = 0$. Then it can be shown that $d(\hat{\Pi}_{zs}M_{ss}\hat{\Pi}'_{zs} - P_{zs}M_{ss}P'_{zs}) = 0$, where $\hat{\Pi}_{zs} = \left(I - \frac{W_{zz}\hat{\beta}'\hat{\beta}}{\hat{\beta}W_{zz}\hat{\beta}'}\right)P_{zs}$.

Let $\Theta = \Pi_{zs}E\Pi'_{zs}$ and $F = P_{zs}M_{ss}P'_{sz}$. We know that $\hat{\beta}_i = \hat{\rho}_J\hat{\Theta}^{iJ}$, where $\rho_J = 1/\rho^J$ (and the capital letter J indicates that there is not to be a sum on that index), and $\hat{\Theta} = \hat{\Pi}_{zs}M_{ss}\hat{\Pi}'_{zs}$. Hence

$$d\hat{\beta}^i = \rho_J d\hat{\Theta}^{iJ} + \Theta^{iJ} d\hat{\rho}_J . \tag{4.17}$$

However $\hat{\beta}^i\hat{\beta}^j\varphi_{ij} = 1$; therefore $\hat{\rho}_J = (\hat{\Theta}^{iJ}\hat{\Theta}^{kJ}\varphi_{ik})^{-\frac{1}{2}}$. From this it follows that

$$d\hat{\rho}_J = -(\rho_J)^3\Theta^{iJ}\varphi_{ik}\, d\hat{\Theta}^{kJ}. \tag{4.18}$$

From (4.14) we see $d\hat{\Theta}^{kJ} = \Theta^{kJ,\alpha\beta}d\hat{\theta}_{\alpha\beta}$. Therefore

$$d\hat{\beta}^i = \rho_J[\Theta^{iJ,\alpha\beta} - \beta^i\beta^l\Theta^{kJ,\alpha\beta}\varphi_{kl}]d\hat{\theta}_{\alpha\beta} . \tag{4.19}$$

Let us define $\psi_j = \beta^i\varphi_{ij}$. Let us multiply (4.19) by $\theta_{\gamma i}$ and ψ_i. We obtain

$$\begin{aligned} \theta_{\gamma i}d\hat{\beta}^i &= \rho_J\theta_{\gamma i}\Theta^{iJ,\alpha\beta}d\hat{\theta}_{\alpha\beta} \\ &= \rho_J\delta_\gamma{}^J\Theta^{\alpha\beta}d\hat{\theta}_{\alpha\beta} - \rho_J\Theta^{J\beta}d\hat{\theta}_{\gamma\beta} = -\beta^\alpha d\hat{\theta}_{\gamma\alpha} , \end{aligned} \tag{4.20}$$

$$\psi_i d\hat{\beta}^i = 0. \tag{4.21}$$

Let us simplify (4.20). We see that

$$\beta^\alpha d\hat{\theta}_{\gamma\alpha} = \beta^\alpha\pi^k_\gamma e_{kl}dp^l_\alpha . \tag{4.22}$$

Hence

$$\begin{aligned}\sigma(\beta^{\alpha} d\hat{\theta}_{\gamma\alpha}, \beta^{\mu} d\hat{\theta}_{\gamma\mu}) &= \beta^{\alpha}\pi^{k}_{\gamma} e_{kl}\beta^{\mu}\pi^{h}_{\gamma} e_{hj} e^{lm} e^{ji}\gamma_{\alpha\mu m i} \\ &= \beta^{\alpha}\beta^{\mu}\pi^{m}_{\gamma}\pi^{i}_{\nu}\gamma_{\alpha\mu m i} = r_{1\gamma\nu},\end{aligned} \tag{4.23}$$

say. Let $\sigma(\hat{\beta}^{i}, \hat{\beta}^{j}) = q_1^{ij}$, and let $Q_1 = (q_1^{ij})$. Then from (4.20) and (4.23) we obtain

$$\Theta Q_1 \Theta = R_1, \tag{4.24}$$

and (4.21) is

$$\psi Q_1 = 0. \tag{4.25}$$

It may be shown (see [1], for example) that the solution is

$$Q_1 = (I - \beta'\psi)_{\cdot k}(\Theta_{kk})^{-1}(R_1)_{kk}(\Theta_{kk})^{-1}(I - \psi'\beta)_{k\cdot}, \tag{4.26}$$

where $k(1 \leq k \leq H)$ is arbitrary except that $\beta^{k} \neq 0$, and $A_{\cdot k}$ denotes A with the k-th column deleted, etc. If the normalization is $\beta^{i} = 1$, $k = i$ is a convenient choice.

Since $\hat{\gamma} = -\hat{\beta}L$,

$$d\hat{\gamma}^{m} = -d\hat{\beta}^{i}\lambda_{i}^{m} - \beta^{i} dl_{i}^{m}. \tag{4.27}$$

Hence

$$\sigma(\hat{\beta}^{j}, \hat{\gamma}^{m}) = -\sigma(\hat{\beta}^{j}, \hat{\beta}^{i})\lambda_{i}^{m} - \sigma(\hat{\beta}^{j}, l_{i}^{m})\beta^{i}, \tag{4.28}$$

$$\sigma(\hat{\gamma}^{m}, \hat{\gamma}^{n}) = \sigma(\hat{\beta}^{j}, \hat{\beta}^{i})\lambda_{i}^{m}\lambda_{j}^{n} + \sigma(\hat{\beta}^{j}, l_{i}^{m})\beta^{i}\lambda_{j}^{n} + \sigma(\hat{\beta}^{j}, l_{i}^{n})\beta^{i}\lambda_{j}^{m} + \sigma(l_{i}^{m}, l_{j}^{n})\beta^{i}\beta^{j}. \tag{4.29}$$

We, therefore, see that we must compute $\sigma(\hat{\beta}^{j}, l_{i}^{m})\beta^{i}$ and $\sigma(l_{i}^{m}, l_{j}^{n})\beta^{i}\beta^{j}$. We find, from (4.20), (4.21), and (4.22) that

$$\theta_{\gamma j}\beta^{i}\sigma(\hat{\beta}^{j}, l_{i}^{m}) = -\beta^{i}\beta^{j}\pi_{j}^{k}c^{mp}\beta_{ijpk} = r_{2\gamma}^{m}, \tag{4.30}$$

say. Let $(\sigma(\hat{\beta}^{j}, l_{i}^{m})\beta^{i}) = Q_2$, and let $R_2 = (r_{2\gamma}^{m})$. Then, from (4.30) and (4.21) we obtain

$$\Theta Q_2 = R_2, \tag{4.31}$$

$$\psi Q_2 = 0. \tag{4.32}$$

The solution is

$$Q_2 = (I - \beta'\psi)_{\cdot k}(\Theta_{kk})^{-1}(R_2)_{k\cdot}. \tag{4.33}$$

We find, readily, that

$$\beta^{i}\beta^{j}\sigma(l_{i}^{m}, l_{j}^{n}) = \beta^{i}\beta^{j}c^{mp}c^{nq}\alpha_{ijpq} = q_3^{mn}, \tag{4.34}$$

say, where $(c^{mp}) = C^{-1}$. Let $Q_3 = (q_3^{mn})$. This concludes the proof of Theorem 3.

THEOREM 3. *If Assumptions* A, B, F, H, I, K, L, *and* M *are satsfied*, $\sqrt{T}(\hat{\beta} - \beta)$ *and* $\sqrt{T}(\hat{\gamma} - \gamma)$ *are asymptotically jointly normally distributed with means zero and covariance matrix*

$$\sigma(\hat{\beta}', \hat{\beta}) = Q_1, \tag{4.35}$$

$$\sigma(\hat{\beta}', \hat{\gamma}) = -Q_2\bar{\Pi}_{xu} - Q_2 , \tag{4.36}$$

$$\sigma(\hat{\gamma}', \hat{\gamma}) = \bar{\Pi}'_{xu}Q_1\bar{\Pi}_{xu} + \bar{\Pi}'_{xu}Q_2 + Q_2'\bar{\Pi}_{xu} + Q_3 , \tag{4.37}$$

where Q_1 *is given by* (4.26), Q_2 *by* (4.33), *and* Q_3 *by* (4.34).

If there is a kind of asymptotic independence of ζ_t and z_t, then the above expressions may be simplified. Corollary 1 results from Theorem 3 and the following assumption:

ASSUMPTION N. $\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathcal{E}(\zeta_t^2 z_t' z_t) = \sigma^2 R$, *where* R *is defined in Assumption* F.

COROLLARY 1. *If Assumptions* A, B, F, H, I, K, L, M, *and* N *are satisfied,* $\sqrt{T}(\hat{\beta} - \beta)$ *and* $\sqrt{T}(\hat{\gamma} - \gamma)$ *are asymptotically jointly normally distributed with means zero and covariance matrix*

$$\sigma(\hat{\beta}', \hat{\beta}) = \sigma^2(I - \beta'\psi)_{\cdot k}(\Theta_{kk})^{-1}(I - \psi'\beta)_{k\cdot} , \tag{4.38}$$

$$\sigma(\hat{\beta}', \hat{\gamma}) = -\sigma^2(I - \beta'\psi)_{\cdot k}(\Theta_{kk})^{-1}(\bar{\Pi}_{xu} + \psi'\gamma)_{k\cdot} , \tag{4.39}$$

$$\sigma(\hat{\gamma}', \hat{\gamma}) = \sigma^2[(\bar{\Pi}'_{xu} + \gamma'\psi)_{\cdot k}(\Theta_{kk})^{-1}(\bar{\Pi}_{xu} + \psi'\gamma)_{k\cdot} + C^{-1}]. \tag{4.40}$$

4.3. ***Asymptotic** distribution of the estimates of the parameters* β *and* γ *with normalization a function of* Ω_{xx} .

If we relax Assumption L that Φ_{xx} is constant, we obtain a more general result. Since the proof, however, is more involved, we shall not give it here; the reader is referred to [1]. In the derivation of the estimates Ω_{xx} was defined as $\mathcal{E}(\delta_t'\delta_t)$. In the asymptotic theory we do not assume that this is the same for each t. We use the following assumption:

ASSUMPTION O.

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathcal{E}(\delta_{ti}\delta_{tj}\delta_{tk}z_{tl}) = n_{ijkl} \text{ exists;}$$

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathcal{E}(\delta_{ti}\delta_{tj}) = \bar{\omega}_{ij} \text{ exists;}$$

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathcal{E}(\delta_{ti}\delta_{tj}\delta_{tk}\delta_{tl}) = \bar{\omega}_{ijkl} + \bar{\omega}_{ij}\bar{\omega}_{kl} \text{ exists.}$$

Let δ_{ijkl} be the quantities n_{ijkl} corresponding to the u's, ϵ_{ijkl}, the quantities corresponding to the c's. Define

$$\chi^{ij} = \tfrac{1}{2}\beta^k\beta^l \frac{\partial\varphi_{kl}}{\partial\omega_{ij}}, \tag{4.41}$$

$$r_{4\gamma} = \beta^k\pi_\gamma^l\chi^{ij}\epsilon_{ijkl} , \tag{4.42}$$

$$q_4' = (I - \beta'\psi)_{\cdot k}(\Theta_{kk})^{-1}(r_4')_{k\cdot} , \tag{4.43}$$

$$q_5 = \chi^{ij}\chi^{kl}\bar{\omega}_{ijkl} , \tag{4.44}$$

$$q_6^k = \chi^{ij}\beta^m\delta_{ijml}c^{kl}. \tag{4.45}$$

With the aid of the matrices Q_1 , Q_2 , and Q_3 , the vectors q_4 and q_6 , and the scalar q_5, we may express the asymptotic covariance matrix of the estimates. We obtain

THEOREM 4. *If Assumptions* A, B, F, H, I, K, M, *and* O *are satisfied, and* Φ_{xx} *is a function of* Ω_{xx}, $\sqrt{T}(\hat{\beta} - \beta)$ *and* $\sqrt{T}(\hat{\gamma} - \gamma)$ *are asymptotically jointly normally distributed with means zero and covariance matrix*

$$\sigma(\hat{\beta}', \hat{\beta}) = Q_1 + q_4'\beta + \beta' q_4 + q_5\beta'\beta, \tag{4.46}$$

$$\sigma(\hat{\beta}', \hat{\gamma}) = -Q_1\bar{\Pi}_{xu} + q_4'\gamma - \beta' q_4\bar{\Pi}_{xu} + q_5\beta'\gamma - Q_2 - \beta' q_6, \tag{4.47}$$

$$\begin{aligned}\sigma(\hat{\gamma}', \hat{\gamma}) = {} & \bar{\Pi}_{xu}' Q_1 \bar{\Pi}_{xu} - \bar{\Pi}_{xu}' q_4'\gamma - \gamma' q_4 \bar{\Pi}_{xu} + q_5\gamma'\gamma \\ & + \bar{\Pi}_{xu}' Q_2 + Q_2'\bar{\Pi}_{xu} - \gamma' q_6 - q_6'\gamma + Q_3,\end{aligned} \tag{4.48}$$

where Q_1, Q_2, Q_3, q_4, q_5, *and* q_6 *are given by* (4.26), (4.33), (4.34), (4.43), (4.44), *and* (4.45) *respectively.*

COROLLARY 2. *If Assumptions* A, B, D, F, H *and* K *are satisfied, and* $\Phi_{xx} = \Omega_{xx}$, $\sqrt{T}(\hat{\beta} - \beta)$ *and* $\sqrt{T}(\hat{\gamma} - \gamma)$ *are asymptotically jointly normally distributed with means zero and covariance matrix*

$$\sigma(\hat{\beta}', \hat{\beta}) = (I - \beta'\psi)_{\cdot k}(\Theta_{kk})^{-1}(I - \psi'\beta)_{k\cdot} + \tfrac{1}{2}\beta'\beta, \tag{4.49}$$

$$\sigma(\hat{\beta}', \hat{\gamma}) = -(I - \beta'\psi)_{\cdot k}(\Theta_{kk})^{-1}(\bar{\Pi}_{xu} + \psi'\gamma)_{k\cdot} + \tfrac{1}{2}\beta'\gamma, \tag{4.50}$$

$$\sigma(\hat{\gamma}', \hat{\gamma}) = (\bar{\Pi}_{xu}' + \gamma'\psi)_{\cdot k}(\Theta_{kk})^{-1}(\bar{\Pi}_{xu} + \psi'\gamma)_{k\cdot} + C^{-1} + \tfrac{1}{2}\gamma'\gamma. \tag{4.51}$$

5. Asymptotic distribution of the likelihood ratio criterion and the small sample criterion for testing a certain hypothesis. The likelihood ratio criterion for testing the hypothesis that the number of coordinates of z_t with zero coefficients in the selected structural equation is as great as it is assumed to be is $(1 + \nu)^{-\frac{1}{2}T}$ [2, Theorem 2], where ν is the smallest root of

$$| P_{xs} M_{ss} P_{xs}' - \nu W_{xx} | = 0. \tag{5.1}$$

Then

$$T\nu = T\frac{\hat{\beta} P_{xs} M_{ss} P_{xs}' \hat{\beta}'}{\hat{\beta} W_{xx} \hat{\beta}'} = (\sqrt{T}\hat{\beta} P_{xs}) \frac{M_{ss}}{\hat{\beta} W_{xx}\hat{\beta}'} (\sqrt{T}\hat{\beta} P_{xs})'. \tag{5.2}$$

From Theorem 5 of [4] it follows that the asymptotic distribution of $T\nu$ is the same as that of the quadratic form $x\frac{E}{\sigma^2}x'$, where x has the limiting distribution of $\sqrt{T}\hat{\beta}P_{xs}$, use being made of $\operatorname{plim}_{T\to\infty} \hat{\beta} W_{xx}\hat{\beta}' = \sigma^2$. We have

$$dx^i = \beta^j dp_j^i + d\hat{\beta}^j \pi_j^i. \tag{5.3}$$

Let $\Upsilon = (I - \beta'\psi)_{\cdot k}(\Theta_{kk})^{-1}(I - \psi'\beta)_{k\cdot}$. Then

$$d\hat{\beta}^j = -\upsilon^{jk}\beta^l \pi_k^m e_{mn} dp_l^n. \tag{5.4}$$

Substituting in (5.3), we obtain

$$dx^i = \beta^j dp_j^i - \upsilon^{lk}\beta^j \pi_k^m e_{mn} dp_j^n \pi_l^i. \tag{5.5}$$

Then

$$\sigma(x^i, x^g) = \sigma^2(e^{ig} - \pi_k^i \pi_q^g v^{kq}) = \sigma^2 \xi^{ig} \tag{5.6}$$

say, and $(\xi^{ig}) = \Xi$.

Let F be chosen so $E = FF'$ and $F'\Xi F = \Psi$ is diagonal. Since $E\Xi E\Xi E = E\Xi E$, the diagonal elements of Ψ are 1 and 0. The number of elements that are 1 is the rank of $E\Xi E$, namely, $D - H + 1$, where D is the number of coordinates of v_t (the number of coordinates whose coefficients in the selected equation are assumed to be zero). Let $z = \frac{1}{\sigma} xF$. Then the asymptotic distribution of $T\nu$ is the distribution of zz' where z is normally distributed with mean zero and covariance matrix Ψ. It is the χ^2-distribution with $D - H + 1$ degrees of freedom. We observe that $T \log (1 + \nu)$ and $TD\lambda$ are asymptotically equal to $T\nu$, where λ is the criterion based on small sample theory [2, Theorem 4]. Finally, we note that ν is independent of the normalization of β.

THEOREM 5. *If Assumptions* A, B, F, H, I, K, M, *and* N *are satisfied,* -2 *times the logarithm of the likelihood ratio criterion,* $-T/2 \log (1 + \nu)$, *the asymptotically equivalent* $T\nu$ *and* TD *times the small sample criterion,* λ, *for testing the hypothesis that the number of coordinates with zero coefficients is* D *are asymptotically distributed as* χ^2 *with* $D - H + 1$ *degrees of freedom.*

This theorem indicates how conservative the small sample test is asymptotically, for that test asymptotically is equivalent to using $T\nu$ as having an asymptotic χ^2-distribution with D degrees of freedom.

6. Asymptotic behavior of confidence regions based on small sample theory. In [2] we deduced confidence regions for β and for β and γ when Assumption E holds. If the normalization of β is

$$\beta \Phi_{xx} \beta' = 1, \tag{6.1}$$

where Φ_{xx} is a given matrix, then a confidence region (a) for β of confidence ϵ consists of all β^* satisfying (6.1) and

$$\frac{\beta^* M_{xs} M_{ss}^{-1} M_{sx} \beta^{*\prime}}{\beta^* W_{xx} \beta^{*\prime}} \leq \frac{D}{T - K} F_{D, T-K}(\epsilon), \tag{6.2}$$

where $F_{D, T-K}(\epsilon)$ is chosen so the probability of (6.2) for $\beta^* = \beta$ is ϵ and K is the number of coordinates of z_t and D is the number of coordinates of v_t. A region (b) for β and γ simultaneously consists of β^* and γ^* satisfying (6.1) and

$$\frac{\beta^* M_{xu} M_{uu}^{-1} M_{ux} \beta^{*\prime} + \beta^* M_{xu} \gamma^{*\prime} + \gamma^* M_{ux} \beta^{*\prime} + \gamma^* M_{uu} \gamma^{*\prime} + \beta^* M_{xs} M_{ss}^{-1} M_{sx} \beta^{*\prime}}{\beta^* W_{xx} \beta^{*\prime}} \leq \frac{K}{T - K} F_{K, T-K}(\epsilon). \tag{6.3}$$

We shall now show that even if Assumption E does not hold the regions have asymptotically confidence coefficients ϵ and they are consistent under general conditions.

Let $c = \beta M_{xu} M_{uu}^{-1} + \gamma$, $e = \beta M_{xs} M_{ss}^{-1}$. We observe from Section 4 that if Assumptions A, B, F, H, K, L, M and N are satisfied, the vectors $\sqrt{T}c$ and $\sqrt{T}e$ have asymptotic independent distributions $N(0, \sigma^2 C^{-1})$ and $N(0, \sigma^2 E^{-1})$, respectively. Then $TcM_{uu}c'/\sigma^2$ and $TeM_{ss}e'/\sigma^2$ will have asymptotic independent χ^2-distributions with $F(= K - D)$ and D degrees of freedom, respectively. Also $\beta W_{xx}\beta'$ approaches σ^2 stochastically. By Theorems 5 and 6 of [4], the left-hand sides of (6.2) and (6.3) have asymptotic F-distributions with D and $T - K$ degrees of freedom and K and $T - K$ degrees of freedom, respectively.

We shall prove that (a) is consistent for β; the proof is similar for (b) as a region for β and γ. If we replace β by b in the definition of e, $eM_{ss}e' = bM_{xs}M_{ss}^{-1}M_{sx}b'$. For $b \neq \beta$ we must show that the probability that b will fall in the confidence region for β approaches zero. The above form approaches $b\Pi_{xs}E\Pi_{xs}'b'$ in probability. If $b \neq \beta$ and satisfies (6.1) then $b\Pi_{xs} \neq 0$ and $eM_{ss}e'$ has a non-zero limit in probability since E is positive definite. Thus b is not in the limiting confidence region.

THEOREM 6. *If Assumptions* A, B, F, H, I, K, M, *and* N *are satisfied, the confidence regions of Theorem* 3 *of* [2] (*including* (a) *and* (b) *above*) *are consistent, and the regions* (a) *and* (b) *have asymptotically the confidence levels* ϵ.

REFERENCES

[1] T. W. ANDERSON AND HERMAN RUBIN, "Estimation of the parameters of a single stochastic difference equation in a complete system," Cowles Commission for Research in Economics, 1947, dittoed.

[2] T. W. ANDERSON AND HERMAN RUBIN, "Estimation of the parameters of a single equation in a complete system of stochastic equations," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 46–63.

[3] H. RUBIN, "Consistency and asymptotic normality in stable linear stochastic difference systems," to be published.

[4] H. RUBIN, "Topological properties of measures on topological spaces," *Duke Math. Journ.*, to be published.