The Annals of Mathematical Statistics
1969, Vol. 40, No. 2, 583–594

ASYMPTOTIC DISTRIBUTION OF MAXIMUM LIKELIHOOD ESTIMATORS IN A LINEAR MODEL WITH AUTOREGRESSIVE DISTURBANCES[1]

By Clifford Hildreth

University of Minnesota

1. Summary. It is shown that maximum likelihood estimators of parameters of a linear model with autoregressive disturbances have an asymptotic multivariate normal distribution with mean vector equal to the true parameter values. Inspection of the variance matrix shows that the estimators are asymptotically efficient and that the estimates of coefficients of the independent variables have the same variance matrix as the best unbiased estimates for a modified model in which the autocorrelation parameter is known.

It is conjectured that the asymptotic distribution of the estimates of coefficients of independent variables may be a useful approximation for moderate sized samples. Alternative approximations for the estimates of the autoregression coefficient and the variance are suggested for further study.

2. The model and principal result. An observed vector y of T components is assumed to be a random drawing from a multivariate normal population with mean vector

$$\mu = Z\gamma, \tag{1}$$

and variance matrix

$$\Omega = \nu A. \tag{2}$$

Z is a known $T \times K$ matrix of rank K; γ is a vector of unknown coefficients to be estimated. ν is an unknown positive constant and A is a $T \times T$ matrix with typical element

$$a_{st} = (1/1 - \rho^2)\rho^{|t-s|}, \tag{3}$$

for $s = 1, 2, \cdots, T$; $t = 1, 2, \cdots, T$, where ρ is an unknown parameter with $|\rho| < 1$.

In the usual contexts for applying this model, the components of y are observed values of a dependent variable, and the elements of Z are observations on K independent variables. A particular observation of the dependent variable

Received 19 September 1966; revised 17 October 1968.

[1] This paper is a revision of RAND Memorandum RM-5059-PR of the same title prepared in July 1966. Suggestions by one of the editors and by Thomas Muench and Leonid Hurwicz were very helpful in preparation of the revision which was completed under National Science Foundation Project 0620-5259 of the University of Minnesota. Warren Dent read an earlier draft and detected several slips. The author is a consultant for the RAND Corporation.

is related to corresponding observations of the independent variables by

(4) $$y_t = \sum_{k=1}^{K} z_{tk}\gamma_k + u_t, \qquad t = 1, 2, \cdots, T,$$

where the u_t are unobserved random disturbances generated by a first-order autoregressive relation

(5) $$u_t = \rho u_{t-1} + v_t, \qquad t = 2, \cdots, T,$$

and the v_t are normal, independent, with zero means and common variance ν. If, in addition, it is assumed that u_t has a stationary distribution, the specifications given in (1) to (3) follow.

The model has particularly interested economists and a number of possibilities for estimating the unknown parameters—γ, ρ, ν—have been noted. (See Theil and Nagar [12], Zellner [13], Durbin [5].) Hildreth and Lu [6] proposed a method for obtaining maximum likelihood estimates and showed that the maximum likelihood estimators are consistent. In what follows it will be shown that the maximum likelihood estimators $(\hat{\gamma}, \hat{\rho}, \hat{\nu})$ of (γ, ρ, ν) have asymptotically a multivariate normal distribution with mean vector equal to the true parameter values and variance matrix indicated in Theorem A.[2]

To state the result more precisely, let $\theta' = (\gamma\ \rho\ \nu)$ and let

(6) $$B = A^{-1} = \begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 \\ -\rho & (1+\rho^2) & -\rho & \cdots & 0 & 0 \\ 0 & -\rho & (1+\rho^2) & \cdots & 0 & 0 \\ & & \vdots & & & \\ 0 & 0 & 0 & \cdots & (1+\rho^2) & -\rho \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{pmatrix} = I + \rho^2 I^* - \rho H$$

where I is a T order identity matrix, I^* differs from I by having zeros instead of ones in the upper left and lower right corners, and H has ones in the $2(T-1)$ positions adjacent to the main diagonal and zeros elsewhere. It is shown in the next section that $|B| = 1 - \rho^2$.

Assume

(7) $|z_{tk}|$ is bounded for $k = 1, 2, \cdots, K$; $t = 1, 2, \cdots$;

(8) For any integers j, k with $1 \leqq j, k \leqq K$ and any non-negative integer τ $\lim_{T\to\infty} T^{-1} \sum_{t=\tau+1}^{T} z_{tj}z_{t-\tau,k}$ exists and is finite

(9) $\lim_{T\to\infty} T^{-1}Z'BZ$ is non-singular.

THEOREM A. *If y has a multivariate normal distribution with the parameters*

[2] For γ and ρ the maximum likelihood estimates considered here are asymptotically equal to the least squares estimates considered by Durbin [5], p. 151. His expression for the variance matrix is equivalent to that of Theorem A, but he did not investigate the form of the distribution.

given by (1), (2), (3) *and if* Z *satisfies* (7), (8), (9) *as* T *increases, then the distribution of* $T^{\frac{1}{2}}(\hat{\theta} - \theta)$ *converges to the multivariate normal with mean vector zero and variance matrix*

$$\begin{pmatrix} \nu(\lim_{T\to\infty} T^{-1}Z'BZ)^{-1} & 0 & 0 \\ 0 & (1-\rho^2) & 0 \\ 0 & 0 & 2\nu^2 \end{pmatrix}.$$

The proof resembles in broad outline Cramér's ([3], pp. 500–4) proof for independently distributed observations. For the model given by (1)–(3) above, the likelihood function is proportional to

$$\varphi(\gamma, \rho, \nu) = \nu^{-T/2}|B|^{\frac{1}{2}}e^{-(1/2\nu)(y-Z\gamma)'B(y-Z\gamma)} \tag{10}$$

with logarithm

$$\Phi(\gamma, \rho, \nu) = -\tfrac{1}{2}T \log \nu + \tfrac{1}{2}\log|B| - \tfrac{1}{2}\nu^{-1}(y - Z\gamma)'B(y - Z\gamma). \tag{11}$$

Let $\partial\Phi/\partial\theta$ be a column vector of partial derivatives of Φ evaluated at the true θ and $\partial\Phi/\partial\hat{\theta}$ be evaluated at a solution to the likelihood equations.[3]
Using similar notation for higher derivatives,

$$\partial\Phi/\partial\hat{\theta} = \partial\Phi/\partial\theta + (\partial^2\Phi/\partial\theta^2)\,(\hat{\theta} - \theta) + R = 0, \tag{12}$$

where R is a vector of remainder terms in Taylor expansions of components of $\partial\Phi/\partial\hat{\theta}$.

Let $\tilde{\theta} = \lambda\theta + (1 - \lambda)\hat{\theta}$ for an appropriate λ, $0 \leqq \lambda \leqq 1$. Let $\partial^3\Phi/\partial\tilde{\theta}_j\partial\tilde{\theta}^2$ be the Hessian of $\partial\Phi/\partial\theta_j$ evaluated at $\tilde{\theta}$. Then

$$R_j = \tfrac{1}{2}(\hat{\theta} - \theta)'(\partial^3\Phi/\partial\tilde{\theta}_j\partial\tilde{\theta}^2)(\hat{\theta} - \theta) \qquad \text{for } j = 1, 2, \cdots, K + 2. \tag{13}$$

Writing

$$G = \frac{1}{2}\begin{pmatrix} (\hat{\theta} - \theta)'\partial^3\Phi/\partial\tilde{\theta}_1\partial\tilde{\theta}^2 \\ \vdots \\ (\hat{\theta} - \theta)'\partial^3\Phi/\partial\tilde{\theta}_{K+2}\partial\tilde{\theta}^2 \end{pmatrix} \tag{14}$$

yields $R = G(\hat{\theta} - \theta)$ and using (12),

$$\partial\Phi/\partial\theta + (\partial^2\Phi/\partial\theta^2)(\hat{\theta} - \theta) + G(\hat{\theta} - \theta) = 0 \qquad \text{or} \tag{15}$$

$$T^{\frac{1}{2}}(\hat{\theta} - \theta) = -(T^{-1}(\partial^2\Phi/\partial\theta^2) + T^{-1}G)^{-1}(T^{-\frac{1}{2}}\partial\Phi/\partial\theta). \tag{16}$$

With respect to the factors on the right side of (16), the following two propositions are verified in Sections 3 and 4 respectively.

---

[3] It is hoped that this unconventional notation will be forgiven in view of the writing saved in equations that follow. Conditions for the existence of a solution to the likelihood equations have not been thoroughly investigated. The likelihood equations can readily be solved (uniquely if Z is of rank K) for $\hat{\gamma}$, $\hat{\nu}$ as functions of $\hat{\rho}$ If these expressions in $\hat{\rho}$ are substituted for $\hat{\gamma}$, $\hat{\nu}$ one is left with a polynomial in $\hat{\rho}$ of degree $4K + 1$. In empirical cases that have been investigated the existence of a solution is clear, but it has not been shown that a solution must always exist. See Hildreth and Lu [6], pp. 11–13 and 20–39.

PROPOSITION 1. *$T^{-\frac{1}{2}}\partial\Phi/\partial\theta$ has a limiting multivariate normal distribution with mean* 0 *and variance matrix*

$$\begin{pmatrix} \nu^{-1}(\lim_{T\to\infty} T^{-1}Z'BZ) & 0 & 0 \\ 0 & (1-\rho^2)^{-1} & 0 \\ 0 & 0 & \frac{1}{2}\nu^{-2} \end{pmatrix} = C.$$

PROPOSITION 2. $\text{plim}\ (T^{-1}\partial^2\Phi/\partial\theta^2 + T^{-1}G) = \text{plim}\ \partial^2\Phi/\partial\theta^2 = -C.$

Theorem A follows from these propositions and two results of Mann and Wald [11] which have been restated by Chernoff ([2], pp. 5, 6). These, using Chernoff's numbering, are:

THEOREM 2. *If $L(X_n) \to L(X)$ (the distribution law of X_n converges to the distribution law of X) and g is a function that is continuous except possibly on a set of probability* 0 *with respect to $L(X)$*, then $L(g(X_n)) \to L(g(X))$.

THEOREM 3. *If $L(X_n) \to L(X)$, then $L(X_n + o_p(1)) \to L(X)$. X_n, X may be random variables, random vectors or other mappings to metric spaces.*

An immediate consequence of these theorems is

CONSEQUENCE. Let $W_T = M_TX_T$ where $\{W_T\}$, $\{X_T\}$ are sequences of random vectors and $\{M_T\}$ is a sequence of random matrices. If M is the probability limit of M_T and the distribution of X_T converges to the distribution of a random vector X, then the distribution of W_T converges to the distribution of MX.

To justify the consequence, let $\bar{W}_T = MX_T = g(X_T)$. Then the distribution of $\bar{W}_T$ converges to that of MX by the Mann-Wald result restated as Theorem 2 above. Now write $W_T = MX_T + (M_T - M)X_T$. Since $\text{plim}\ M_T = M$, the last term is $o_p(1)$, i.e., converges to 0 in probability, and the distribution of W_T converges to the distribution of MX by Theorem 3.

Let $X_T = T^{-\frac{1}{2}}\partial\Phi/\partial\theta$. Applying the consequence and the two propositions to equation (16), it is seen that the distribution of $T^{\frac{1}{2}}(\hat{\theta} - \theta)$ converges to the limit of the distribution of $-C^{-1}X_T$. X_T is asymptotically $n(0, C)$, multivariate normal with mean 0 and variance C, by Proposition 1 and therefore $-C^{-1}X_T$ is asymptotically distributed according to $n(0, C^{-1})$ as stated in Theorem A. Since $C = E(\partial\Phi/\partial\theta)(\partial\Phi/\partial\theta)'$, the maximum likelihood estimates are asymptotically efficient (Cramér [3], p. 489).

If ρ were known, the best unbiased estimator of γ would be

$$\bar{\gamma} = (Z'BZ)^{-1}Z'By \tag{17}$$

which would be $n(\gamma, \nu(Z'BZ)^{-1})$ for any sample size. Referring to Theorem A, it is seen that $T^{\frac{1}{2}}\hat{\gamma}$ has the same asymptotic distribution as $T^{\frac{1}{2}}\bar{\gamma}$.

One purpose of studying asymptotic distributions is to aid in the development of distributions that may serve as useful approximations for moderate sized samples. This question is tentatively considered in Section 5 and some suggestions for further study are offered.

3. Proof of Proposition 1. Differentiating (11) and letting

$u' = (u_1, u_2, \cdots, u_T)$, yields

$$\partial\Phi/\partial\theta = \begin{pmatrix} \partial\Phi/\partial\gamma \\ \partial\Phi/\partial\rho \\ \partial\Phi/\partial\nu \end{pmatrix} = \begin{pmatrix} \nu^{-1}Z'Bu \\ -\rho(1-\rho^2)^{-1} - \tfrac{1}{2}\nu^{-1}u'B^*u \\ -\tfrac{1}{2}T\nu^{-1} + \tfrac{1}{2}\nu^{-2}u'Bu \end{pmatrix} \tag{18}$$
$$= \nu^{-1}\begin{pmatrix} Z'Bu \\ -\tfrac{1}{2}u'B^*u \\ \tfrac{1}{2}\nu^{-1}(u'Bu - \nu T) \end{pmatrix} + O(1)$$

where $B^* = \partial B/\partial\rho = 2\rho I^* - H$.

Define

$$v = Mu \tag{19}$$

where

$$\begin{aligned} m_{11} &= (1-\rho^2)^{\frac{1}{2}}, \\ m_{tt} &= 1 \quad \text{for} \quad t = 2, 3, \cdots, T, \\ m_{t,t-1} &= -\rho \quad \text{for} \quad t = 2, 3, \cdots, T, \text{ and} \\ m_{st} &= 0 \quad \text{otherwise.} \end{aligned} \tag{20}$$

The last $T - 1$ components of v then agree with those defined by equations (5). Also

$$|M| = (1-\rho^2)^{\frac{1}{2}}, \qquad M'M = B, \qquad |B| = (1-\rho^2), \tag{21}$$
$$M^{-1}M'^{-1} = B^{-1} = A, \qquad v =_d n(0, \nu I).$$

Note that

$$-\tfrac{1}{2}u'B^*u = \textstyle\sum_{t=1}^{T-1} u_t u_{t+1} - \rho \sum_{t=2}^{T-1} u_t^2 = \sum_{t=1}^{T-1} u_t v_{t+1} + \rho u_1^2. \tag{22}$$

If we let $\bar{B}$ be equal to B^* in every element except the upper left and set $\bar{b}_{11} = 2\rho$, then

$$-\tfrac{1}{2}u'\bar{B}u = \textstyle\sum_{t=1}^{T-1} u_t v_{t+1} \qquad \text{and} \tag{23}$$

$$Eu'\bar{B}u = 0, \qquad E(u'\bar{B}u)^2 = \textstyle\sum_{t=1}^{T-1} Eu_t^2 v_{t+1}^2 = (T-1)\nu^2(1-\rho^2)^{-1}. \tag{24}$$

Define

$$\xi_T = \nu^{-1}T^{-\frac{1}{2}}\begin{pmatrix} Z'Bu \\ -\tfrac{1}{2}u'\bar{B}u \\ \tfrac{1}{2}\nu^{-1}(u'Bu - \nu T) \end{pmatrix} \tag{25}$$
$$= \nu^{-1}T^{-\frac{1}{2}}\begin{pmatrix} Z'M'v \\ -\tfrac{1}{2}v'M'^{-1}\bar{B}M^{-1}v \\ \tfrac{1}{2}\nu^{-1}(v'v - \nu T) \end{pmatrix} = T^{-\frac{1}{2}}\partial\Phi/\partial\theta + O_p(T^{-\frac{1}{2}}).$$

Clearly $E\xi_T = 0$,

$$(26) \qquad \operatorname{Var} \xi_T = E\xi_T\xi_T' = \begin{bmatrix} (\nu T)^{-1}Z'BZ & 0 & 0 \\ 0 & (T-1)/T(1-\rho^2) & 0 \\ 0 & 0 & \frac{1}{2}\nu^{-2} \end{bmatrix}$$

and

$$(27) \qquad \lim_{T\to\infty} \operatorname{Var} \xi_T = C$$

as defined in the statement of Proposition 1. Since the sequences of second moments are uniformly integrable[4] (Loève [10], pp. 182–3), C is also the matrix of second moments of the limiting distribution. To prove Proposition 1 it remains to show that ξ_T, and therefore $T^{-\frac{1}{2}}\partial\Phi/\partial\theta$, is asymptotically multivariate normal.

This is done by showing that the distribution of an arbitrary linear combination of the components of ξ_T is normal in the limit (see Anderson [1], p. 37). Let the arbitrary linear combination be

$$(28) \qquad \zeta_T = T^{-\frac{1}{2}}(\alpha'Z'M'v + \beta v'M'^{-1}\bar{B}M^{-1}v + v'v - \nu T)$$

where α is any vector of order K and β is any scalar.[5]

Suppose Q is an orthogonal matrix which diagonalizes $(\beta M'^{-1}\bar{B}M^{-1} + I)$. Let

$$(29) \qquad Q'(\beta M'^{-1}\bar{B}M^{-1} + I)Q = \Lambda,$$

$$(30) \qquad w = Q'v.$$

w is $n(0, \nu I)$ and

$$(31) \qquad \zeta_T = T^{-\frac{1}{2}}(\alpha'Z'M'Qw + w'\Lambda w - \nu T).$$

Let λ_{tT} for $t = 1, 2, \cdots, T$ be the diagonal elements of the matrix Λ corresponding to sample size T.[6] From (31) and the fact that ζ_T has mean zero

$$(32) \qquad \textstyle\sum_{t=1}^{T} \lambda_{tT} = T.$$

Hence

$$(33) \quad E(w'\Lambda w)^2 = 2\nu^2 \textstyle\sum_{t=1}^{T} \lambda_{tT}^2 + \nu^2 \sum_{t=1}^{T}\sum_{q=1}^{T} \lambda_{tT}\lambda_{qT} = 2\nu^2 \sum_{t=1}^{T} \lambda_{tT}^2 + \nu^2 T^2.$$

From the definitions of Λ and w we have

$$(34) \qquad w'\Lambda w = -2\beta \textstyle\sum_{t=1}^{T-1} u_t v_{t+1} + \sum_{t=1}^{T} v_t^2$$

---

[4] This may be confirmed by noting in (24) that $Z'M'$ is bounded and the latent roots of $M'^{-1}BM^{-1}$ are bounded. The latter is proved below.

[5] Since multiplication by a scalar does not change the form of a normally distributed random variable, the only loss of generality in this expression is that the coefficient of the last component of ξ_T has been tacitly assumed nonzero. If it is zero, the final two terms of (28) vanish and the argument which follows applies with the deletion of appropriate terms.

[6] If the notation were to be made complete, T subscripts would have to be added to Z, M, Q, Λ, w. These have been omitted to make expressions in which they enter less cumbersome.

so

$$E(w'\Lambda w)^2 = 4\beta^2\nu^2(T-1)/(1-\rho^2) + \nu^2 T(T+2) \tag{35}$$

and, using (34) and (35),

$$\textstyle\sum_{t=1}^{T} \lambda_{tT}^2 = 2\beta^2(T-1)/(1-\rho^2) + T. \tag{36}$$

It is also useful in what follows to know that the λ_{tT} are bounded.

From (29), it is sufficient to show that the latent roots of $M'^{-1}\bar{B}M^{-1}$ are bounded. These are the same as the roots of $\bar{B}M^{-1}M'^{-1} = \bar{B}A$ where $a_{st} = (1-\rho^2)^{-1}\rho^{|t-s|}$. A root of $\bar{B}A$, say λ_{tT}^*, satisfies

$$\bar{B}Ax = \lambda_{tT}^* x \tag{37}$$

where x is the latent vector corresponding to λ_{tT}^*. For given T let $\bar{x}$ be a maximal element of $\{|x_t| : t = 1, 2, \cdots, T\}$ and let $\bar{e}$ be the maximal absolute value of elements of $\bar{B}Ax$. Clearly

$$|\lambda_{tT}^*| \leqq \bar{e}/\bar{x}. \tag{38}$$

Consider a typical component of Ax, say $A_{(t)}x$,

$$\begin{aligned} A_{(t)}x &= (1-\rho^2)^{-1}\left(\textstyle\sum_{s=1}^{t} \rho^{t-s}x_s + \sum_{s=t+1} \rho^{s-t}x_s\right), \\ |A_{(t)}x| &\leqq 2(1-\rho^2)^{-1}(1-|\rho|)^{-1}\bar{x}. \end{aligned} \tag{39}$$

Since no element of $\bar{B}$ is greater than two in absolute value and no row contains more than three nonzero elements, it follows that

$$\bar{e} \leqq 12(1-\rho^2)^{-1}(1-|\rho|)^{-1}\bar{x} \quad \text{and} \tag{40}$$

$$|\lambda_{tT}^*| \leqq 12(1-\rho^2)^{-1}(1-|\rho|)^{-1}. \tag{41}$$

Since the bound does not depend on T the conclusion follows.

Next consider the moment generating function of ζ_T.

$$\begin{aligned} \psi_T(s) &= Ee^{s\zeta_T} \\ &= (2\pi)^{-T/2}\int \exp\{sT^{-\frac{1}{2}}(\beta'Z'M'Qw + w'\Lambda w - \nu T) - \tfrac{1}{2}\nu^{-1}w'w\}\,dw \\ &= e^{-s\nu T^{\frac{1}{2}}}(2\pi)^{-T/2} \\ &\quad \cdot \int \exp\{-\tfrac{1}{2}\nu^{-1}[w'(I - 2\nu sT^{-\frac{1}{2}}\Lambda)w - 2\nu sT^{-\frac{1}{2}}\alpha'Z'M'Qw]\}\,dw. \end{aligned} \tag{42}$$

Let

$$J = (I - 2\nu sT^{-\frac{1}{2}}\Lambda), \qquad d = Q'MZ\alpha \tag{43}$$

and let L be a bound for $|\lambda_{tT}|$. Then for $s \,\varepsilon\, (-1/4\nu L, 1/4\nu L)$, J is positive definite for all T and "completing the square" of the exponent of e under the integral in (41) yields

$$\psi_T(s) = \exp\{-s\nu T^{\frac{1}{2}} + \tfrac{1}{2}\nu s^2 T^{-1} d'J^{-1}d\}|J|^{-\frac{1}{2}}. \tag{44}$$

Now

(45) $|J|^{-\frac{1}{2}} = \prod_{t=1}^{T} (1 - 2\nu s T^{-\frac{1}{2}} \lambda_{tT})^{-\frac{1}{2}} = \exp \{-\frac{1}{2} \sum_{t=1}^{T} \log (1 - 2\nu s T^{-\frac{1}{2}} \lambda_{tT})\}$

and

(46) $$\log (1 - 2\nu s T^{-\frac{1}{2}} \lambda_{tT}) = -2\nu s \lambda_{tT} T^{-\frac{1}{2}} - \tfrac{1}{2} \cdot 4\nu^2 s^2 \lambda_{tT}^2 T^{-1} - \tfrac{1}{3} \cdot 8\nu^3 s^3 \lambda_{tT}^3 T^{-3/2} - \cdots$$
$$= -2\nu s \lambda_{tT} T^{-\frac{1}{2}} - 2\nu^2 s^2 \lambda_{tT}^2 T^{-1} - k_{tT}$$

where, still taking $s \varepsilon (-1/4\nu L, 1/4\nu L)$,

$$|k_{tT}| \leqq \tfrac{1}{3} \cdot 8\nu^3 |s|^3 |\lambda_{tT}|^3 T^{-3/2} (1 + 2\nu |s| \, |\lambda_{tT}| T^{-\frac{1}{2}} + 4\nu^2 s^2 \lambda_{tT}^2 T^{-3/2} + \cdots)$$

(47) $$= \tfrac{1}{3} \cdot 8\nu^3 |s|^3 |\lambda_{tT}|^3 T^{-3/2} (1 - 2\nu |s| \, |\lambda_{tT}| T^{-\frac{1}{2}})^{-1}$$
$$\leqq \tfrac{1}{3} \cdot 16\nu^3 |s^3| \, |\lambda_{tT}|^3 T^{-3/2} = O(T^{-3/2}).$$

Thus

(48) $$\sum_{t=1}^{T} \log (1 - 2\nu s T^{-\frac{1}{2}} \lambda_{tT})$$
$$= -2\nu s T^{-\frac{1}{2}} \sum_{t=1}^{T} \lambda_{tT} - 2\nu^2 s^2 T^{-1} \sum_{t=1}^{T} \lambda_{tT}^2 + O(T^{-\frac{1}{2}})$$

and together with (32), (36), (44) and (45) this implies

(49) $$\psi_T(s) = \exp \{\tfrac{1}{2} \nu s^2 T^{-1} d' J^{-1} d + 4\beta^2 (T - 1) \nu^2 s^2 (T(1 - s^2))^{-1} + \nu^2 s^2\}$$
$$+ O(T^{-\frac{1}{2}}).$$

It is easily verified that

(50) $$J^{-1} = (I - 2\nu s T^{-\frac{1}{2}} \Lambda)^{-1} = (I + 2\nu s T^{-\frac{1}{2}} \Lambda J^{-1})$$

and therefore

(51) $$d' J^{-1} d = d' d + 2\nu s T^{-\frac{1}{2}} d' \Lambda J^{-1} d.$$

For $s \varepsilon (-1/4\nu L, 1/4\nu L)$ the absolute value of a typical element of ΛJ^{-1} satisfies

(52) $$\left| \frac{\lambda_{tT}}{1 - \dfrac{2\nu s}{\sqrt{T}} \lambda_{tT}} \right| < 2L$$

Therefore

(53) $$2\nu s T^{-\frac{1}{2}} d' \Lambda J^{-1} d \leqq 4\nu s L T^{-\frac{1}{2}} d' d = O(T^{\frac{1}{2}})$$

since assumption (9), p. 4 implies that $d' d = \alpha' Z' B Z \alpha = O(T)$.

Combining (49), (51), (53)

(54) $$\psi_T(s) = \exp \left\{ \frac{\nu s^2}{2T} d' d - \frac{4\beta^2 (T - 1) \nu^2 s^2}{T(1 - e^2)} - 2\nu^2 s^2 + O(T^{-\frac{1}{2}}) \right\}$$

and

(55) $$\lim_{T \to \infty} \psi_T(s) = \exp \left\{ \frac{s^2}{2} \left(\frac{\nu}{T} \alpha' Z' B Z \alpha + \frac{4\beta^2 \nu^2}{(1 - \rho^2)} + 2\nu^2 \right) \right\}$$

which is the moment generating function of a normally distributed random variable with mean zero and variance indicated by the coefficient of $s^2/2$ in (55). By Curtiss' [4] extension of a Theorem of Lévy and Cramér it follows that the limiting distribution of ζ_T is

$$n(0, (\nu/T)\alpha' Z' BZ\alpha + 4\beta^2\nu^2/(1 - \rho^2) + 2\nu^2).$$

4. Proof of Proposition 2. This consists of two parts: (2a) plim $T^{-1}(\partial^2\Phi/\partial\theta^2) = C$ and (2b) plim $T^{-1}G = 0$. Concerning (2a),

$$(56) \quad (\partial^2\Phi/\partial\theta^2) = \begin{pmatrix} \partial^2\Phi/\partial\gamma^2 & \partial^2\Phi/\partial\gamma\partial\rho & \partial^2\Phi/\partial\gamma\partial\nu \\ \cdots & \partial^2\Phi/\partial\rho^2 & \partial^2\Phi/\partial\rho\partial\nu \\ \cdots & \cdots & \partial^2\Phi/\partial\nu^2 \end{pmatrix}$$

$$= -\nu^{-1}\begin{pmatrix} Z'BZ & -Z'B^*u & \nu^{-1}Z'Bu \\ \cdots & \left(\dfrac{\nu(1+\rho^2)}{(1-\rho^2)^2} + u'I^*u\right) & -\tfrac{1}{2}\nu^{-1}u'B^*u \\ \cdots & \cdots & (\nu^{-2}u'Bu - \tfrac{1}{2}T\nu^{-1}) \end{pmatrix}.$$

Thus plim $T^{-1}(\partial^2\Phi/\partial\gamma^2) = -\nu^{-1}$ lim $T^{-1}Z'BZ$. $T^{-1}(\partial^2\Phi/\partial\gamma\partial\rho)$ is normal with mean zero and variance $T^{-2}\nu^{-1}Z'B^*AB^*Z$ which converges to zero under assumptions (7) and (8). $T^{-1}(\partial^2\Phi/\partial\gamma\partial\nu)$ is normal with mean zero and variance $T^{-2}\nu^{-3}Z'BZ$ which also converges to zero so

$$(57) \qquad \text{plim } (\partial^2\Phi/\partial\gamma\partial\rho) = \text{plim } (\partial^2\Phi/\partial\gamma\partial\nu) = 0.$$

$$(58) \qquad \text{plim } T^{-1}(\partial^2\Phi/\partial\rho^2) = -\nu^{-1} \text{ plim } T^{-1}\textstyle\sum_{t=2}^{T-1}u_t^2 = -(1-\rho^2)^{-1}$$

since $T^{-1}\sum_{t=2}^{T-1}u_t^2$ has mean $(T-2)\nu/T(1-\rho^2)$ and variance $2(T-2)\nu^2/T^2(1-\rho^2)^2 + 4(T-3)\nu^2\rho^2/T^2(1-\rho^2)^3 + O_p(T^{-1})$

$$(59) \quad \text{plim } T^{-1}(\partial^2\Phi/\partial\rho\partial\nu) = -\nu^{-2}(\text{plim } T^{-1}\textstyle\sum_{t=1}^{T-1}u_t v_{t+1} + \text{plim } \rho u_1^2 T^{-1}) = 0$$

since $T^{-1}\sum_{t=1}^{T-1}u_t v_{t+1}$ has mean 0 and variance $(T-1)\nu^2/T^2(1-\rho^2) = O(T^{-1})$.

$$(60) \quad \text{plim } T^{-1}(\partial^2\Phi/\partial\nu^2) = -\nu^{-3} \text{ plim } T^{-1}\textstyle\sum_{t=1}^{T}v_t^2 + \tfrac{1}{2}\nu^{-2} = -\tfrac{1}{2}\nu^{-2}$$

since $T^{-1}\sum_{t=1}^{T}v_t^2$ has mean ν and variance $2T\nu^2T^{-2} = O(T^{-1})$.

(2b) can conveniently be verified using the knowledge that the maximum likelihood estimators are consistent.[7] Referring to (14), a typical row of G may be written

$$(61) \qquad G_{(k)} = (\theta - \theta)'(\partial^3\Phi/\partial\tilde{\theta}_k\partial\tilde{\theta}^2) \quad \text{for} \quad k = 1, 2, \cdots, K+2$$

and where $\partial^3\Phi/\partial\tilde{\theta}_k\partial\tilde{\theta}^2$ is a matrix of third derivatives involving θ_k and evaluated at $\tilde{\theta}$ which lies between $\hat{\theta}$ and the true θ. With consistency of $\hat{\theta}$ the problem thus reduces to checking that third derivatives of $T^{-1}\Phi$ are $O_p(1)$. Inspection of the

[7] Consistency of $\hat{\gamma}$, $\hat{\rho}$ was shown in Hildreth and Lu [6], pp. 52–4. Since

$$\hat{\nu} = T^{-1}\hat{u}'\hat{B}\hat{u} = T^{-1}u'Bu + o_p(1) = T^{-1}\textstyle\sum_{t=1}^{T}v_t^2 + o_p(1),$$

it is clear that $\hat{\nu}$ is also consistent.

matrix of second derivatives in (56) indicates that third derivatives either vanish or are proportional to expressions evaluated in (2a) and found to be $O_p(1)$or smaller.

5. Conjectured approximations. An alternative expression for the relevant terms of the log likelihood function is

$$\begin{aligned}\Phi(\gamma, \rho, \nu) = & -\tfrac{1}{2}T \log \nu + \tfrac{1}{2} \log (1 - \rho^2) \\ & - \tfrac{1}{2}\nu^{-1}[(\dot\gamma - \gamma)'Z'Z(\dot\gamma - \gamma) + \dot s + \rho^2(\ddot\gamma - \gamma)'Z'I^*Z(\ddot\gamma - \gamma) \\ & + \rho^2\ddot s - \rho(\dddot\gamma - \gamma)'Z'HZ(\dddot\gamma - \gamma) - \rho\dddot s] \end{aligned} \tag{62}$$

where

$$\begin{aligned} \dot\gamma &= (Z'Z)^{-1}Z'y, & \dot s &= (y - Z\dot\gamma)'(y - Z\dot\gamma), \\ \ddot\gamma &= (Z'I^*Z)^{-1}Z'I^*y, & \ddot s &= (y - Z\ddot\gamma)'I^*(y - Z\ddot\gamma), \\ \dddot\gamma &= (Z'HZ)^{-1}Z'Hy, & \dddot s &= (y - Z\dddot\gamma)H(y - Z\dddot\gamma). \end{aligned} \tag{63}$$

From (63) and (1) it is seen that $\dot\gamma$, $\ddot\gamma$, $\dddot\gamma$ are unbiased, normally distributed estimators of γ. Inspection of (62) reveals that the likelihood function depends on y only through $(\dot\gamma, \ddot\gamma, \dddot\gamma, \dot s, \ddot s, \dddot s,)$ and the latter is therefore a sufficient statistic.

Furthermore, by the Lehmann-Scheffé [8] ratio test[8] this array is minimal sufficient. Equating derivatives of the log likelihood to zero and rearranging terms yields

$$\hat\gamma = [Z'Z + \hat\rho^2Z'I^*Z - \hat\rho Z'HZ]^{-1}[Z'Z\gamma + \hat\rho^2Z'I^*Z\ddot\gamma - \hat\rho Z'HZ\dddot\gamma] \tag{64}$$

$$\begin{aligned}\hat\rho = & \tfrac{1}{2}(y - Z\hat\gamma)'H(y - \hat Z\gamma)[(y - Z\hat\gamma)I^*(y - Z\hat\gamma)]^{-1} \\ & - \hat\rho\hat\nu[(1 - \hat\rho^2)(y - Z\hat\gamma)'(y - Z\hat\gamma)]^{-1}, \end{aligned} \tag{65}$$

$$\hat\nu = T^{-1}(y - Z\hat\gamma)'(I + \hat\rho^2I^* - \hat\rho H)(y - Z\hat\gamma). \tag{66}$$

(64) shows $\hat\gamma$ as a random mixture (generalized weighted average) of three unbiased, normally distributed estimators of γ. If $\hat\rho$ were a constant, $\hat\gamma$ would be normally distributed with mean γ and variance $\nu T^{-1}Z'\hat BZ$ (where $\hat B = I + \hat\rho^2I^* - \hat\rho H$) for any sample size. It therefore seems reasonable to conjecture that $\hat\gamma$ is approximately normal for medium sized samples and that the asymptotic parameters may also be pretty good approximations. Of course, such an offhand

---

[8] See also Lindgren [9], pp. 191–200. If y and w are two sample points

$$\phi(y)/\phi(w) = \exp \{-\tfrac{1}{2}\nu^{-1} [(y - Z\gamma)'B(y - Z\gamma) - (w - Z\gamma)'B(w - Z\gamma)]\}$$

which is independent of ν, ρ, γ if and only if $y'y = w'w$, $y'Z = w'Z$, $y'I^*Z = w'I^*Z$, $y'HZ = w'HZ$. Thus these $3K + 1$ quantites constitute a minimal sufficient statistic. Since the statistic defined in (50) is a function of these quantities, it is also minimal sufficient. Note that since $y'Z = y'I^*Z + y_1Z_{(1)} + y_TZ_{(T)}$ where $Z_{(t)}$ is the tth row of Z, the minimal sufficient statistic is generally of dimension $2K + 3$.

conjecture should be checked analytically and perhaps by Monte Carlo studies as soon as possible.

From (65) and (66) one might conjecture that much larger sample sizes would be needed in order that the finite sample distributions of $\hat{\rho}$ and $\hat{\nu}$ be usefully approximated by their asymptotic distributions. Hence there seems a larger incentive to examine (65) and (66) more closely to see if their forms suggest alternative approximate distributions that might prove useful in moderate sized samples.

Let $\hat{u} = y - Z\hat{\gamma}$ and rewrite

$$\hat{\nu} = T^{-1}\hat{u}'\hat{B}\hat{u}. \tag{66$'$}$$

If ρ were known the best unbiased estimator of γ would be $\tilde{\gamma} = (Z'BZ)^{-1}Z'B$ and the corresponding estimator of ν would be $\tilde{\nu} = T^{-1}\tilde{u}'B\tilde{u}$ where $\tilde{u} = y - Z\tilde{\gamma}$. In this latter case it is known that $\tilde{u}'B\tilde{u}$ is distributed as χ^2_{T-K}. The present case differs in that one additional parameter, ρ, has been fitted to the data. This should make the corresponding sum of squares of residuals $(\hat{u}'\hat{B}\hat{u} = (\hat{u}\hat{B}^{\frac{1}{2}})'(\hat{B}^{\frac{1}{2}}\hat{u}))$ smaller than in the case where ρ is known. One might conjecture that the present sum of squares is still approximately χ^2 but with a smaller number of degrees of freedom. $T - K - 1$ would be appropriate if ρ entered linearly in determining y and might be considered a possibility even in the absence of linearity.

Similarly, rewrite

$$\hat{\rho} = \tfrac{1}{2}\hat{u}'H\hat{u}(\hat{u}'I^*\hat{u})^{-1} - \hat{\rho}\hat{\nu}((1 - \hat{\rho}^2)\hat{u}'I^*\hat{u})^{-1}. \tag{65$'$}$$

The second term on the right is clearly small when T is sufficiently large, and unless $|\hat{\rho}|$ is very close to 1, the term becomes fairly small for moderate T. (Note that $|\hat{\rho}/(1 - \hat{\rho}^2)| < 2.25$ when $|\hat{\rho}| \leqq .8$ and that $\hat{\nu}/\hat{u}'I^*\hat{u} \leqq (3 + \hat{\rho}^2)T^{-1} + 3(\hat{u}_1^2 + \hat{u}_T^2)(T\hat{u}'I^*\hat{u})^{-1}$.) Assuming that the first term on the right of (65$'$) dominates the form of the distribution, the problem of finding a useful approximation still seems formidable. However, some of the same considerations that led Thiel and Nagar [12] to consider a transformed β-distribution as an approximation to the Von-Neumann ratio would also apply here.

Define

$$\hat{d} = \textstyle\sum_{t=2}^{T} (\hat{u}_t - \hat{u}_{t-1})^2 / \sum_{t=1}^{T} \hat{u}_t^2 \tag{67}$$

as the Von-Neumann ratio for maximum likelihood residuals and let

(68) $\bar{\rho} = \frac{1}{2}u'Hu(\hat{u}'\hat{u})^{-1} = \sum_{t=2}^{T} \hat{u}_t\hat{u}_{t-1}/\sum_{t=1}^{T} \hat{u}_t^2$ be an approximation to $\hat{\rho}$.

(69) $\hat{d} = 2 - 2\bar{\rho} - (\hat{u}_1^2 + \hat{u}_T^2)/\sum_{t=1}^{T} \hat{u}_t^2$.

If $\hat{d}$ is approximately distributed as a linear function of a random variable having a β-distribution then so is $\bar{\rho}$ whenever T is large enough that the last term of (69) is small in probability.

If one pursues this, admittedly rather tenuous, hint that the distribution of $\hat{\rho}$ may be approximated by a transformed β distribution, the natural range is $(-1, 1)$ rather than $(0, 1)$, the range of the standard β distribution.

If a random variable X has the usual β density

$$(70)\quad f(x) = (B(p, q))^{-1}x^{p-1}(1 - x)^{q-1} \quad \text{for } 0 \leqq x \leqq 1 \quad \text{and} \quad p, q > 0$$

then the random variable $W = 2X - 1$ has density

$$(71)\quad q(w) = 2^{-(p+q-1)}(B(p, q))^{-1}(w + 1)^{p-1}(1 - w)^{q-1} \quad \text{for } -1 \leqq w \leqq 1 \quad \text{and} \quad p, q > 0$$

and

$$(72)\qquad EW = (p - q)(p + q)^{-1},$$

$$(73)\qquad E(W - EW)^2 = 4pq(p + q)^{-2}(p + q + 1)^{-1}.$$

p and q should be determined by studying the mean and variance of $\hat{\rho}$ to obtain tolerable approximations for moderate sized samples. If one were to take the asymptotic mean and variance as first approximations the following relations would result

$$(74)\quad (p - q)(p + q)^{-1} = \rho, \qquad 4pq(p + q)^{-2}(p + q + 1)^{-1} = (1 - \rho^2)T^{-1}$$

or

$$(75)\qquad p = \tfrac{1}{2}(T - 1)(1 + \rho), \qquad q = \tfrac{1}{2}(T - 1)(1 - \rho).$$

REFERENCES

[1] ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.

[2] CHERNOFF, HERMAN (1956). Large sample theory: parametric case. *Ann. Math. Statist.* **27** 1–22.

[3] CRAMÉR, HAROLD (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.

[4] CURTISS, J. H. (1942). A note on the theory of moment generating functions. *Ann. Math. Statist.* **13** 430.

[5] DURBIN, JAMES (1960). Estimation of parameters in time series regression models. *J. Roy. Statist. Soc. Ser. B* **22** 139.

[6] HILDRETH, CLIFFORD and LU, J. Y. (1960). *Demand relations with autocorrelated disturbances*. Research Bulletin 276, Agricultural Experiment Station, Michigan State Univ.

[7] HILDRETH, CLIFFORD (1966). Asymptotic distribution of maximum likelihood estimators in linear models with autoregressive disturbances. RAND Memorandum RM 5059 PR, Santa Monica.

[8] LEHMANN, ERIC and SCHEFFÉ, HENRY (1950). Completeness, similar regions and unbiased estimation. *Sankhyā* **10** 327.

[9] LINDGREN, B. W. (1960). *Statistical Theory*. Macmillan, New York.

[10] LOÈVE, MICHEL (1963). *Probability Theory* (3d ed.) Van Nostrand, Princeton.

[11] MANN, H. B. and WALD, A. (1943). On stochastic limit and order relationships. *Ann. Math. Statist.* **14** 217–226.

[12] THEIL, H. and NAGAR, A. L. (1961). Testing the independence of regression disturbances. *J. Amer. Statist. Assoc.* **56** 793–806.

[13] ZELLNER, ARNOLD and TIAO, G. C. (1964). Bayesian analysis of the regression model with autocorrelated errors. *J. Amer. Statist. Assoc.* **59** 763–778.