The Annals of Statistics
1981, Vol. 9, No. 3, 531–543

ESTIMATION OF THE PARAMETERS OF STOCHASTIC DIFFERENCE EQUATIONS[1]

By Wayne A. Fuller, David P. Hasza and J. Jeffery Goebel

Iowa State University, Kansas State University and Iowa State University

Let Y_t satisfy the stochastic difference equation

$$Y_t = \sum_{i=1}^{q} \psi_{ti}\alpha_i + \sum_{j=1}^{p} \gamma_j Y_{t-j} + e_t,$$

where the $\{\psi_{ti}\}$ are fixed sequences and (or) weakly stationary time series and the e_t are independent random variables, each with mean zero and variance σ^2. The form of the limiting distributions of the least squares estimators of α_i and γ_j depend upon the absolute value of the largest root of the characteristic equation, $m^p - \sum_{j=1}^{p} \gamma_j m^{p-j} = 0$. Limiting distributions of the least squares estimators are established for the situations where the largest root is less than one, equal to one, and greater than one in absolute value. In all three situations the regression t-type statistic is of order one in probability under mild assumptions. Conditions are given under which the limiting distribution of the t-type statistic is standard normal.

1. Introduction.

Let the time series Y_t satisfy

(1) $$Y_t = \sum_{i=1}^{q} \psi_{ti}\alpha_i + \sum_{j=1}^{p} \gamma_j Y_{t-j} + e_t, \qquad t = 1, 2, \cdots,$$

where the $\{\psi_{ti}\}$ are fixed sequences and (or) weakly stationary time series, and the e_t are independent $(0, \sigma^2)$ random variables. If any $\{\psi_{ti}\}$ are stationary time series, then we assume $\{e_t\}$ to be independent of such $\{\psi_{ti}\}$. The model (1) is of order p and the polynomial equation

(2) $$m^p - \sum_{j=1}^{p} \gamma_j m^{p-j} = 0$$

is the characteristic equation of the model. The roots $m_1, m_2, \cdots, m_p$ of (2) are the characteristic roots of the process (1).

Mann and Wald (1943) considered estimation of the parameters of the model with $\{\psi_{ti}\}$ restricted to the constant function and the roots of the characteristic equation less than one in absolute value. White (1958) obtained the limiting joint moment generating function of the numerator and denominator of the least squares estimator of γ_1 for the first order case and no ψ-variables. The moment generating function had three forms, according as the root of the characteristic equation was less than one, equal to one, or greater than one in absolute value.

Anderson (1959), Rao (1961), and Stigum (1974) have studied estimation of the model when at least one of the roots of the characteristic equation is greater than one in absolute value. Venkataraman (1967) and Narasimham (1969) studied the model with $p = 2$ and at least one root greater then one in absolute value. Rao (1978) considered the case $p = 1$ and $\gamma_1 = 1$. None of these studies permitted ψ-variables in the equation. Venkataraman (1968, 1973) studied the second order explosive model (at least one root greater than one) with a

Received January 1978; revised June 1979.

[1] This research was partly supported by Joint Statistical Agreements J.S.A. 77-3 and J.S.A. 78-30 with the U.S. Census Bureau.

AMS 1970 *subject classifications.* Primary 62M10; secondary 62J05.

Key words and phrases. Stochastic difference equation, time series, regression for time series, autoregressive process.

constant term.

If the set $\{\psi_{ti}\}$ contains the constant function and a stationary vector autoregressive time series and if all of the roots of the characteristic equation are less than one in modulus, then the limiting distribution of the estimator is normal; see, for example, Hannan (1970 page 329), and Nicholls (1976).

The limiting behavior of the estimator for a model with fixed ψ-variables and roots of the characteristic equation less than one in absolute value has been investigated by several authors. Among the first to consider the statistical properties of this model were workers at the Cowles Commission; see Anderson and Rubin (1950), Koopmans, Rubin, and Leipnik (1950), and Rubin (1950). Hannan (1965), Amemiya and Fuller (1967), Hatanaka (1974), and Fuller (1976, page 435) studied the situation in which there are nonlinear restrictions on the parameters arising from the specification of autocorrelated errors and lagged dependent variables in the equation. Hannan and Nicholls (1972), Reinsel (1976), and Fuller (1976) considered estimation of model (1) with the roots of (2) less than one in absolute value.

The assumptions of Hatanaka and Fuller would permit polynomials in time to be among the explanatory variables, but the normalization used in their formal theorems must be modified if there is a nonzero coefficient for the time trend. Hannan and Nicholls (1972) mentioned that their results were applicable to deviations from time trends. Rao (1967) discussed the estimation of equation (1) when the set $\{\psi_{ti}\}$ contains only polynomials in time and the roots of the characteristic equation are less than one in absolute value.

Dickey (1976), Fuller (1976), and Dickey and Fuller (1979) considered the estimation of equation (1) assuming one of the roots of the characteristic equation to be one and permitted the set $\{\psi_{ti}\}$ to include the constant function and time. Hasza (1977) discussed the estimation of equation (1) with one of the roots of the characteristic equation greater than one in absolute value. Hasza permitted a set $\{\psi_{ti}\}$ composed of polynomial functions of time to enter the equation.

We present the limiting distribution of the least squares estimator for equation (1) under three situations. If all of the roots of the characteristic equation are less than one in absolute value, we demonstrate that the limiting distribution of the estimator is normal under mild regularity conditions. If one of the roots of the characteristic equation is one and the others are less than one in absolute value, we demonstrate that the limiting distribution depends upon the nature of the set $\{\psi_{ti}\}$ and upon the parameters in the model. When one of the roots of the characteristic equation is greater than one in absolute value and the remaining roots are less than one in absolute value, the least squares estimators normalized by the square roots of the sums of squares of the explanatory variables are normal if and only if the e_t are normal independent random variables.

2. The model. We consider the least squares estimation of the parameters of model (1) assuming the e_t to be independent $(0, \sigma^2)$ random variables such that $E\{|e_t|^{2+\nu}\} < L$ for some real L and $\nu > 0$. The parameters α_i and γ_j are fixed unknown constants and the ψ_{ti} are fixed functions of time. We assume $Y_0, Y_{-1}, \cdots, Y_{-p+1}$ to be known and fixed.

The difference equation (1) may be solved to obtain

$$Y_t = S_t + u_t, \qquad u_t = \sum_{j=0}^{t-1} v_j e_{t-j}, \tag{3}$$
$$S_t = \sum_{j=0}^{p-1} v_{t+j} Y_{-j} + \sum_{j=0}^{t-1} v_j \sum_{i=1}^{q} \alpha_i \psi_{t-j,i},$$

and the v_j satisfy the homogeneous difference equation with characteristic equation (2) and initial conditions $v_0 = 1$ and $v_j = 0$ for $j < 0$. The sequence S_t defines the fixed or systematic part of Y_t, where we set $S_{-t} = Y_{-t}$ for $t = 0, 1, \cdots, p-1$. The random portion of Y_t is u_t.

We shall consider the limiting distribution of the least squares estimators of the parameters of (1) for sequences $\{\psi_{ti}\}$ in a relatively broad class. In particular we permit $\sum_{t=1}^{n} \psi_{ti}^2$ to increase at a slower rate or at a faster rate than the sample size n. Such possibilities complicate the normalization of the estimators required to obtain nonsingular limiting distributions for situations of practical interest. To illustrate the normalization problem, consider $q = 2$ and $\psi_t = (\psi_{t1}, \psi_{t2}) = (t + \eta_{t1}, t + \eta_{t2})$, where $\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} \eta_{ti}\eta_{tj} = \delta_{ij}$ and δ_{ij} is Kronecker's delta. Define $\mathbf{D}_n = \text{diag}\{(\sum_{t=1}^{n} \psi_{ti}^2)^{1/2}\}$, then $\mathbf{H}_n = \mathbf{D}_n^{-1}[\sum_{t=1}^{n} \psi_t'\psi_t]\mathbf{D}_n^{-1}$ approaches singularity as n increases. On the other hand, if we define $x_{t1} = 2t + \eta_{t1} + \eta_{t2}$ and $x_{t2} = \eta_{t1} - \eta_{t2}$ with mild assumptions on e_t and η_{ti}, then the vector $\mathbf{D}_n^{-1} \sum_{t=1}^{n} \mathbf{x}_t' e_t$ converges in distribution to a nonsingular bivariate normal random variable. Regression variables with behavior similar to that of (ψ_{t1}, ψ_{t2}) seem common in economics.

A similar problem arises with the matrix of sums of squares and products of ψ_{ti} and Y_{t-j}. For example, let the model be

$$Y_t = (t, Y_{t-1})(\alpha_1, \gamma_1)' + e_t,$$

where $Y_0 = 0$, $\alpha_1 \neq 0$, and $|\gamma_1| < 1$. In this case S_t approaches $\alpha_1(1 - \gamma_1)^{-1}t - \alpha_1\gamma_1(1 - \gamma_1)^{-2}$ as t increases, and the matrix

$$\mathbf{G}_n = \mathbf{D}_n^{-1}[\textstyle\sum_{t=1}^{n} (t, Y_{t-1})'(t, Y_{t-1})]\mathbf{D}_n^{-1}$$

converges to a singular matrix as $n \to \infty$, where $\mathbf{D}_n = \text{diag}\{(\sum_{t=1}^{n} t^2)^{1/2}, (\sum_{t=1}^{n} Y_{t-1}^2)^{1/2}\}$. Let $W_{t1} = Y_{t-1} - \alpha_1(1 - \gamma_1)^{-1}t$, $\beta_1 = \alpha_1[1 + (1 - \gamma_1)^{-1}]$, and consider the reparameterized model

$$Y_t = (t, W_{t1})(\beta_1, \gamma_1)' + e_t.$$

Let $(\hat{\beta}_1, \hat{\gamma}_1)'$ denote the least squares estimator of $(\beta_1, \gamma_1)'$. Then the limiting distribution of

$$[(\textstyle\sum_{t=1}^{n} t^2)^{1/2}(\hat{\beta}_1 - \beta_1), (\sum_{t=1}^{n} W_{t1}^2)^{1/2}(\hat{\gamma}_1 - \gamma_1)]'$$

is bivariate normal.

A third item to consider in the parameterization of the model is m_1, the root of (2) with largest absolute value. For $p > 1$ and $|m_1| \geq 1$, a degenerate asymptotic distribution for the estimator of m_1 can be avoided by considering

$$Y_t = \textstyle\sum_{i=1}^{q} \psi_{ti}\alpha_i + \sum_{j=1}^{p-1} \beta_{q+j}(Y_{t-j} - m_1 Y_{t-j-1}) + \beta_{q+p} Y_{t-1} + e_t, \tag{4}$$

where $\beta_{q+p} = m_1$ and the roots of

$$m^{p-1} - \textstyle\sum_{j=1}^{p-1} \beta_{q+j} m^{p-1-j} = 0 \tag{5}$$

are $m_2, m_3, \cdots, m_p$.

We use the Gram-Schmidt orthogonalization procedure to reparameterize model (1) and the equivalent model (4). The reparameterization is not introduced for computational purposes but to facilitate the proofs of this paper. Applications of the theorems to the original parameters are noted in the text. Given n observations ($n > q + p$), let

$$\begin{aligned}
x_{t1n} &= \psi_{t1} & & \\
x_{tin} &= \psi_{ti} - \textstyle\sum_{j=1}^{i-1} c_{ijn}x_{tjn} & & i = 2, 3, \cdots, q, \\
x_{tin} &= rS_{t+q-i} - S_{t+q-i-1} - \textstyle\sum_{j=1}^{i-1} c_{ijn}x_{tjn}, & & i = q + 1, q + 2, \cdots, q + p - 1 \\
x_{t,p+q,n} &= S_{t-1} - \textstyle\sum_{j=1}^{p+q-1} c_{p+q,jn}x_{tjn}, & &
\end{aligned} \tag{6}$$

where $r = 0$ if $|m_1| < 1$ and $r = m_1^{-1}$ if $|m_1| \geq 1$, and the c_{ijn} are the multiple regression coefficients obtained by the least squares regression of ψ_{ti} and $rS_{t+q-i} - S_{t+q-i-1}$ on x_{tjn}, $j = 1, 2, \cdots, i - 1$ and $t = 1, 2, \cdots, n$. The $c_{p+q,jn}$ are obtained by the least squares regression of S_{t-1} on $x_{tjn}, j = 1, 2, \cdots, p + q - 1$. It is understood that $c_{ijn} = 0$ if $\sum_{t=1}^{n} x_{tjn}^2 = 0$. Define for $p > 1$

$$W_{t1n} = rY_{t-1} - Y_{t-2} - \textstyle\sum_{j=1}^{q} c_{q+1,jn}x_{tjn},$$

(7) $$W_{tin} = rY_{t-i} - Y_{t-i-1} - \sum_{j=1}^{q} c_{q+i,jn} x_{tjn} - \sum_{j=1}^{i-1} c_{q+i,q+j,n} W_{tjn}, \qquad i = 2, 3, \cdots, p-1$$

$$W_{tpn} = Y_{t-1} - \sum_{j=1}^{q} c_{q+p,jn} x_{tjn} - \sum_{j=1}^{p-1} c_{q+p,q+j,n} W_{tjn}.$$

Let $\mathbf{A}_n$ be the nonsingular transformation matrix defined by (6) and (7) so that

$$\mathbf{X}'_{tn} = (x_{t1n}, x_{t2n}, \cdots, x_{tqn}, W_{t1n}, \cdots, W_{tpn})'$$
$$= \mathbf{A}_n(\psi_{t1}, \psi_{t2}, \cdots, \psi_{tq}, Y_{t-1}, \cdots, Y_{t-p})'.$$

Then

(8) $$W_{tin} = x_{t,q+i,n} + \sum_{j=1}^{p} a_{q+i,q+j,n} u_{t-j},$$

where a_{ijn} is the (ij)th element of $\mathbf{A}_n$, and model (1) can be written as

(9) $$Y_t = \mathbf{X}_{tn}\boldsymbol{\theta}_n + e_t,$$

where $\boldsymbol{\theta}'_n = (\alpha_1, \alpha_2, \cdots, \alpha_q, \gamma_1, \cdots, \gamma_p)\mathbf{A}_n^{-1}$.

3. The stationary case. In this section we assume that the roots $m_1, m_2, \cdots, m_p$ of (2) are less than one in absolute value. We apply a version of the central limit theorem for martingale differences given by Scott (1973) to obtain limiting properties of the least squares estimator and associated test statistics. See also Brown (1971) and Dvoretzky (1972).

THEOREM 1. *Let model* (1) *hold with the roots of the characteristic equation* (2) *less than one in absolute value. Let* $\{e_t\}$ *be a sequence of independent* $(0, \sigma^2)$ *random variables with* $E\{|e_t|^{2+\nu}\} < L$ *for some real* L *and* ν *greater than zero. Considering the parameterization in* (9), *define*

$$\hat{\boldsymbol{\theta}}_n = (\textstyle\sum_{t=1}^{n} \mathbf{X}'_{tn}\mathbf{X}_{tn})^{-1} \sum_{t=1}^{n} \mathbf{X}'_{tn} Y_t.$$

Should the matrix be singular, the inverse is replaced by the Moore-Penrose generalized inverse. Assume

(10) $$\lim_{n\to\infty} \sup_{1\le t\le n} (\textstyle\sum_{s=1}^{n} x_{sin}^2)^{-1} x_{tin}^2 = 0, \qquad i = 1, 2, \cdots, q,$$

and

(11) $$\lim_{n\to\infty} \sup_{1\le t\le n} (n + \textstyle\sum_{s=1}^{n} x_{sin}^2)^{-1} x_{tin}^2 = 0, \qquad i = q+1, q+2, \cdots, q+p.$$

Let $\mathbf{D}_n$ *be the diagonal matrix whose elements are the square roots of the diagonal elements of* $\sum_{t=1}^{n} \mathbf{X}'_{tn}\mathbf{X}_{tn}$ *and define*

$$\mathbf{G}_n = \mathbf{D}_n^{-1} \textstyle\sum_{t=1}^{n} \mathbf{X}'_{tn}\mathbf{X}_{tn}\mathbf{D}_n^{-1}.$$

Let $\mathbf{G}_n^{1/2}$ *be the symmetric positive definite square root of* $\mathbf{G}_n$ (*see* Bellman, 1960, *page* 92). *Then*

$$\sigma^{-1}\mathbf{G}_n^{1/2}\mathbf{D}_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n) \to_{\mathscr{L}} N(\mathbf{O}, \mathbf{I}) \qquad \textit{as} \quad n \to \infty.$$

PROOF. The probability that $|\mathbf{G}_n| \neq 0$ converges to one as n increases. We have

$$\mathbf{D}_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n) = [\mathbf{D}_n^{-1} \textstyle\sum_{t=1}^{n} \mathbf{X}'_{tn}\mathbf{X}_{tn}\mathbf{D}_n^{-1}]^{-1}\mathbf{D}_n^{-1} \sum_{t=1}^{n} \mathbf{X}'_{tn} e_t.$$

Consider

$$\textstyle\sum_{t=1}^{n} W_{tin}^2 = \sum_{t=1}^{n} (x_{t,q+i,n} + \sum_{j=1}^{p} a_{q+i,q+j,n} u_{t-j})^2.$$

By the definition of u_t,

(12) $$n^{-1} \textstyle\sum_{t=1}^{n} u_t u_{t-j} \to_P \gamma_u(j),$$

where $u_t = 0$ for $t \le 0$ and $\gamma_u(j)$ is the covariance function of a stationary autoregressive process with characteristic equation (2). It follows that $[n^{-1}\sum_{t=1}^n W_{tjn}^2]^{-1}$ is $O_p(1)$. Now

(13) $$(n + \sum_{t=1}^n x_{tjn}^2)^{-2}\operatorname{Var}\{\sum_{t=1}^n x_{tjn}u_{t-j}\} = (n + \sum_{t=1}^n x_{tjn}^2)^{-2}\sum_{t=1}^n\sum_{s=1}^n x_{tjn}x_{sjn}E\{u_{t-j}u_{s-j}\}$$

and the right side of this equation converges to zero because $|E\{u_t u_j\}|$ is bounded by a multiple of $\lambda^{|t-j|}$ for some $|\lambda| < 1$. Therefore, for $j = 1, 2, \cdots, p$

$$[\sum_{t=1}^n E\{W_{tjn}^2\}]^{-1/2}[\sum_{t=1}^n W_{tjn}^2]^{1/2} \to_P 1.$$

Let

$$\mathbf{H}_n = [E\{\mathbf{D}_n^2\}]^{-1/2}[\sum_{t=1}^n \mathbf{x}_{tn}'\mathbf{x}_{tn} + n\mathbf{F}_n][E\{\mathbf{D}_n^2\}]^{-1/2},$$

$$\mathbf{x}_{tn} = (x_{t1n}, x_{t2n}, \cdots, x_{t,q+p,n}),$$

$$\mathbf{F}_n = \mathbf{A}_n\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{\Gamma}_n\end{pmatrix}\mathbf{A}_n',$$

$$\mathbf{\Gamma}_n = E\{n^{-1}\sum_{t=1}^n (u_{t-1}, u_{t-2}, \cdots, u_{t-p})'(u_{t-1}, u_{t-2}, \cdots, u_{t-p})\}.$$

Note that $\mathbf{H}_n^{-1}$ is well defined because $\mathbf{\Gamma}_n$ is positive definite and that

$$p\lim_{n\to\infty}(\mathbf{G}_n^{-1/2} - \mathbf{H}_n^{-1/2}) = \mathbf{0}.$$

Consider the linear combination

$$\boldsymbol{\eta}'\mathbf{H}_n^{-1/2}\mathbf{D}_n^{-1}\sum_{t=1}^n \mathbf{X}_{tn}'\, e_t$$

where $\boldsymbol{\eta}$ is a vector of arbitrary real numbers such that $\boldsymbol{\eta}'\boldsymbol{\eta} \ne 0$. Because

$$E\{\sum_{t=1}^n W_{tjn}e_t\sum_{t=1}^n W_{tin}e_t\} = \sigma^2\sum_{t=1}^n E\{W_{tin}W_{tjn}\},$$

we can write

$$\boldsymbol{\eta}'\mathbf{H}_n^{-1/2}\mathbf{D}_n^{-1}\sum_{t=1}^n \mathbf{X}_{tn}'e_t = \sum_{t=1}^n Z_{tn} + o_p(1) = S_{nn} + o_p(1),$$

where $S_{nn} = \sum_{t=1}^n Z_{tn}$,

(14)
$$Z_{tn} = (g_{tn} + v_{tn})e_t,$$
$$g_{tn} = \sum_{i=1}^q \eta_i(\sum_{s=1}^n x_{sin}^2)^{-1/2}x_{tin} + \sum_{j=1}^p \eta_{q+j,n}(\sum_{s=1}^n E\{W_{sjn}^2\})^{-1/2}x_{t,q+j,n},$$
$$v_{tn} = \sum_{i=1}^p \eta_{q+i,n}(\sum_{s=1}^n E\{W_{sin}^2\})^{-1/2}(\sum_{j=1}^p a_{q+i,q+j,n}u_{t-j}),$$
$$\eta_{q+i,n} = \sum_{j=1}^p h_{q+j,q+i,n}^{(-1/2)}\eta_{q+j},$$

and $h_{jin}^{(-1/2)}$ is the (ji)th element of $\mathbf{H}_n^{-1/2}$. Observe that $\{g_{tn}: t = 1, 2, \cdots, n\}$ is fixed and that the v_{tn} are fixed linear combinations of $\{u_{t-j}: j = 1, 2, \cdots, p\}$ for a particular n.

Because $Y_0, Y_{-1}, \cdots, Y_{-p+1}$ are fixed, the sigma field $\mathscr{B}_{kn}$ generated by $(Z_{1n}, Z_{2n}, \cdots, Z_{kn})$ is the sigma field generated by $(e_1, e_2, \cdots, e_k)$. Therefore,

$$E\{Z_{tn}^2 \mid \mathscr{B}_{t-1,n}\} = \delta_{tn}^2 = (g_{tn} + v_{tn})^2\sigma^2.$$

By the definition of the v_{tn} and by (12) and (13), $V_{nn}^2 s_{nn}^{-2}$ converges in probability to one as $n \to \infty$, where

$$V_{nn}^2 = \sum_{t=1}^n (g_{tn} + v_{tn})^2\sigma^2, \qquad \text{and} \quad s_{nn}^2 = \sum_{t=1}^n (g_{tn}^2 + E\{v_{tn}^2\})\sigma^2.$$

To apply the results of Scott (1973) it is sufficient to show that $\{Z_{tn}\}$ satisfies the Lindeberg type condition

$$s_{nn}^{-2}\sum_{t=1}^n E\{Z_{tn}^2 I(|Z_{tn}| \ge \epsilon s_{nn})\} \to 0,$$

where $I(A)$ is the indicator function for the set A. We have

$$
\begin{aligned}
&s_{nn}^{-2} \sum_{t=1}^{n} E\{Z_{tn}^2 I(|Z_{tn}| \geq \epsilon s_{nn})\} \\
&\quad = s_{nn}^{-2} \sum_{t=1}^{n} E\{(g_{tn} + v_{tn})^2 e_t^2 I(|(g_{tn} + v_{tn})e_t| \geq \epsilon s_{nn})\} \qquad (15)\\
&\quad \leq s_{nn}^{-2-\nu} \epsilon^{-\nu} L 2^{2+\nu} \left(\sum_{t=1}^{n} |g_{tn}|^{2+\nu} + \sum_{t=1}^{n} E\{|v_{tn}|^{2+\nu}\}\right).
\end{aligned}
$$

By (14), we can write $v_{tn} = \sum_{j=1}^{p} \xi_{tjn} u_{t-j}$, where $\xi_{tjn} = O(n^{-1/2})$ and it follows that $E\{|v_{tn}|^{2+\nu}\} = O(n^{-1-\nu/2})$. By the definitions of $\mathbf{H}_n$ and $\mathbf{H}_n^{-1/2}$, $s_{nn}^2 = \boldsymbol{\eta}'\boldsymbol{\eta}\sigma^2$ and

$$s_{nn}^{-2-\nu} \sum_{t=1}^{n} |g_{tn}|^{2+\nu} \leq \sigma^{-2}(\sigma^2 \boldsymbol{\eta}'\boldsymbol{\eta})^{-\nu} \sup_{1\leq t\leq n} |g_{tn}|^{\nu}.$$

Now

$$
\begin{aligned}
\sup_{1\leq t\leq n} g_{tn}^2 \leq \sup_{1\leq t\leq n} (q+p)^2 \Big[&\sum_{i=1}^{q} \left(\sum_{s=1}^{n} x_{sin}^2\right)^{-1} x_{tin}^2 \eta_i^2 \\
&+ \sum_{j=1}^{p} \left(\sum_{s=1}^{n} E\{W_{sjn}^2\}\right)^{-1} x_{t,q+j,n}^2 \eta_{q+j,n}^2 \Big],
\end{aligned}
$$

which converges to zero by (10) and (11). It follows that $s_{nn}^{-1} S_{nn}$ converges in distribution to a $N(0, 1)$ random variable. The conclusion follows because $\boldsymbol{\eta}$ was arbitrary. □

For a particular n, the elements of $\boldsymbol{\theta}_n$ are fixed linear combinations of the parameters of the original problem. Therefore, for large samples, Theorem 1 justifies the use of the ordinary regression statistics in making inferential statements concerning the parameters of regression model (1).

In our proof we assumed the e_t to be independent with bounded $(2 + \nu)$th moment. The result can also be obtained under the assumption that the e_t are independent and identically distributed; see Lemma 2 of Brown (1971).

In our derivation we treated ψ_{ti} as fixed. Condition (10) holds almost surely for stationary processes satisfying mild conditions; see Hannan and Heyde (1972). Therefore, Theorem 1 holds if such stationary processes are included in the set $\{\psi_{ti}\}$ and if $\{e_t\}$ is independent of such $\{\psi_{ti}\}$. The theorem also holds for processes that are the sum of a fixed sequence satisfying condition (10) and a zero mean weakly stationary process.

4. The unit root case. In this section we assume that model (1) holds and that one of the roots of the characteristic equation (2) is of unit absolute value with the remaining roots less than one in absolute value. In this case the nature of the limiting distribution of the estimators and test statistics depends upon the parameters of the model and upon the nature of the ψ_{ti} included in model (1). Only in special circumstances will the limiting distribution of the least squares regression coefficients be normal.

We first consider two cases of practical interest in which the standard hypothesis testing procedures are not applicable in large samples. In case (a) we consider the situation in which the φ_{ti} of model (1) satisfy $\varphi_{t1} \equiv 1$ while condition (17) of Theorem 2 holds for $i = 2, 3, \cdots, q$. In case (b) we consider the situation in which $\varphi_{t1} \equiv 1$, $\varphi_{t2} = t$, and condition (17) holds for $i = 3, 4, \cdots, q$.

We introduce an additional modification of the parameterization of (9), letting

$$W_{tpn}^{\dagger} = W_{tpn} - n^{-1} \sum_{s=1}^{n} W_{spn} = W_{tpn} - \bar{W}_{\cdot pn}$$

for case (a) and

$$W_{tpn}^{\dagger} = W_{tpn} - \bar{W}_{\cdot pn} - b_{wn}[t - \tfrac{1}{2}(n + 1)]$$

for case (b), where $\bar{W}_{\cdot pn}$ is the sample mean of W_{tpn} and b_{wn} is the least squares coefficient obtained by regressing W_{tpn} on $t - \frac{1}{2}(n + 1)$. This transformation differs from that used in Section 2 because the coefficients for x_{t1n} and x_{t2n} defining $W_{tpn}^{\dagger}$ are functions of the random variables $\{u_t\}_{t=1}^{n}$.

Let $\mathbf{A}_{(u)n}$ be the matrix whose first $q + p - 1$ rows are the first $q + p - 1$ rows of $\mathbf{A}_n$ and whose last row $\mathbf{A}_{(u)p+q,\cdot,n}$ is given by the above transformation so that

$$W_{tpn}^{\dagger} = \mathbf{A}_{(u)p+q,\cdot,n}(\psi_{t1}, \psi_{t2}, \cdots, \psi_{tq}, Y_{t-1}, \cdots, Y_{t-p})'.$$

Assuming $m_1 = 1$, we write our transformed regression equation as

$$Y_t = \mathbf{X}_{(u)tn}\boldsymbol{\theta}_{(u)tn} + e_t, \tag{16}$$

where

$$\mathbf{X}'_{(u)tn} = \mathbf{A}_{(u)n}(\psi_{t1}, \psi_{t2}, \cdots, \psi_{tq}, Y_{t-1}, \cdots, Y_{t-p})'$$
$$\boldsymbol{\theta}'_{(u)tn} = (\alpha_1, \alpha_2, \cdots, \alpha_q, \gamma_1, \cdots, \gamma_p)\mathbf{A}_{(u)n}^{-1}.$$

The asymptotic distributions obtained in Theorem 2 involve the distributions of two statistics, say $\hat{\tau}_\mu$ and $\hat{\tau}_\tau$, that were characterized by Dickey and Fuller (1979). See also Fuller (1976, Section 8.5).

THEOREM 2. *Let model* (1) *hold with $m_1 = 1$ and $m_2, m_3, \cdots, m_p$ less than one in absolute value, where the m_i are the roots of* (2). *Let e_t be independent* $(0, \sigma^2)$ *random variables such that $E\{|e_t|^{2+\nu}\} < L$ for some real L and $\nu > 0$. Let*

$$\hat{\boldsymbol{\theta}}_{(u)n} = \left(\sum_{t=1}^n \mathbf{X}'_{(u)tn}\mathbf{X}_{(u)tn}\right)^{-1} \sum_{t=1}^n \mathbf{X}'_{(u)tn} Y_t,$$

where $\mathbf{X}_{(u)tn}$ is defined in (16). *Let $\mathbf{D}_{(u)n}$ be the diagonal matrix whose elements are the square roots of the diagonal elements of $\sum_{t=1}^n \mathbf{X}'_{(u)tn}\mathbf{X}_{(u)tn}$. Let*

$$\mathbf{G}_{(u)n} = \mathbf{D}_{(u)n}^{-1} \sum_{t=1}^n \mathbf{X}'_{(u)tn}\mathbf{X}_{(u)tn}\mathbf{D}_{(u)n}^{-1},$$
$$\hat{\boldsymbol{\delta}}_n = \sigma^{-1}\mathbf{G}_{(u)n}^{1/2}\mathbf{D}_{(u)n}(\hat{\boldsymbol{\theta}}_{(u)n} - \boldsymbol{\theta}_{(u)n}),$$

where $\mathbf{G}_{(u)n}^{1/2}$ is the positive definite square root of $\mathbf{G}_{(u)n}$. Assume that (10) *and* (11) *are satisfied and that*

$$\lim_{n\to\infty}\left(n^2 \sum_{t=1}^n x_{tin}^2\right)^{-1} \sum_{t=1}^n \sum_{h=0}^{n-t} t x_{tin} x_{t+h,in} = 0 \tag{17}$$

for $i = 2, 3, \cdots, q$ with case (a) *and for $i = 3, 4, \cdots, q$ with case* (b). *Assume that for $i = q+1, q+2, \cdots, q+p$,*

$$\lim_{n\to\infty} n^{-2} \sum_{t=1}^n x_{tin}^2 = 0, \tag{18}$$

$$\lim_{n\to\infty} n^{-2}\left(n + \sum_{t=1}^n x_{tin}^2\right)^{-1} \sum_{t=1}^n \sum_{h=0}^{n-t} t x_{tin} x_{t+h,in} = 0. \tag{19}$$

Then the last element of $\hat{\boldsymbol{\delta}}_n$ converges in distribution to the statistic $\hat{\tau}_\mu$ for case (a) *and to $\hat{\tau}_\tau$ for case* (b), *where $\hat{\tau}_\mu$ and $\hat{\tau}_\tau$ are characterized in Dickey and Fuller* (1979). *The limiting distribution of the vector of the remaining $q + p - 1$ elements of $\hat{\boldsymbol{\delta}}_n$ is normal with zero mean and identity covariance matrix for both cases.*

PROOF. We have

$$u_t = \sum_{i=1}^t \sum_{j=0}^\infty v_j^* e_i - \sum_{i=1}^t \sum_{j=t-i+1}^\infty v_j^* e_i$$
$$= A_t + B_t,$$

where the v_i^* satisfy the homogeneous difference equation with characteristic equation (5). It follows that $\sum_{j=0}^\infty |v_j^*| < \infty$ and $\sum_{j=t-i+1}^\infty |v_j^*| < M\lambda^{t-i+1}$ for some $M < \infty$ and some $0 < \lambda < 1$.

Let $A_{tn}^\dagger$ and $B_{tn}^\dagger$ denote the portions of A_t and B_t that are orthogonal to ψ_{t1} under(a) and orthogonal to ψ_{t1} and ψ_{t2} under (b). Then $\sum_{t=1}^n (B_{tn}^\dagger)^2 = O_p(n)$ because $E\{|B_{tn}^\dagger|^2\}$ is bounded. Also

$$\sum_{t=1}^n x_{t,p+q,n} B_{tn}^\dagger = o_p(n^{3/2})$$

because

$$\operatorname{Var}\{\sum_{t=1}^{n} x_{t,p+q,n} B_{tn}^{\dagger}\} \le n \sum_{t=1}^{n} x_{t,p+q,n}^{2} \operatorname{Var}\{B_{tn}^{\dagger}\} = o(n^{3}).$$

Therefore,

$$\sum_{t=1}^{n} (W_{tpn}^{\dagger})^{2} = \sum_{t=1}^{n} [x_{t,p+q,n} + A_{tn}^{\dagger} + B_{tn}^{\dagger}]^{2}$$
$$= \sum_{t=1}^{n} [x_{t,p+q,n} + A_{tn}^{\dagger}]^{2} + o_{p}(n^{3/2}).$$

By construction, for $i = 2, 3, \cdots, q$ under (a) and for $i = 3, 4, \cdots, q$ under (b),

$$\sum_{t=1}^{n} W_{tpn}^{\dagger} x_{tin} = \sum_{t=1}^{n} W_{tpn} x_{tin}$$

and, therefore,

$$E\{\sum_{t=1}^{n} W_{tpn}^{\dagger} x_{tin}\} = E\{\sum_{t=1}^{n} \sum_{j=1}^{p} a_{q+p,q+j,n} u_{t-j} x_{tin}\} = 0.$$

Also

(20) $$\operatorname{Var}\{\sum_{t=1}^{n} u_{t-m} x_{tin}\} \le 2\,\sigma^{2} (\sum_{j=0}^{\infty} |v_{j}^{*}|)^{2} \sum_{t=1}^{n} \sum_{h=0}^{t-1} t x_{tin} x_{t+h,in}$$

for $m = 1, 2, \cdots, p$. By (19), the first q elements of the last row of $\mathbf{G}_{(u)n}$ converge in probability to zero because $[\sum_{t=1}^{n} (A_{tn}^{\dagger})^{2}]^{-1} = O_{p}(n^{-2})$. For $j = 1, 2, \cdots, p-1$

$$W_{tjn} = x_{t,q+j,n} + \sum_{i=1}^{j} a_{q+j,q+i,n}(u_{t-i} - u_{t-i-1})$$
$$= x_{t,q+j,n} + \sum_{i=1}^{j} a_{q+j,q+i,n} \sum_{j=0}^{t-i} v_{j}^{*} e_{t-j},$$

where $\sum_{j=0}^{t-i} v_{j}^{*} e_{t-j}$ is converging to a stationary autoregressive process with characteristic equation (5). Therefore,

$$p\lim[\sum_{t=1}^{n} (W_{tpn}^{\dagger})^{2} \sum_{t=1}^{n} W_{tjn}^{2}]^{-1/2} \sum_{t=1}^{n} W_{tpn}^{\dagger} W_{tjn} = 0$$

for $j = 1, 2, \cdots, p-1$, and the first $q + p - 1$ elements of the last row of $\mathbf{G}_{(u)n}$ converge in probability to zero.

The first q elements of

(21) $$\mathbf{D}_{(u)n}^{-1} \sum_{t=1}^{n} \mathbf{X}'_{(u)tn} e_{t}$$

are linear combinations of the e_t where the coefficients are fixed. The next $(p-1)$ elements are of the form

$$[\sum_{t=1}^{n} W_{tjn}^{2}]^{-1/2} \sum_{t=1}^{n} W_{tjn} e_{t},$$

where the W_{tjn} are linear combinations of stationary processes. Because the x_{tin} satisfy assumptions (10) and (11), the limiting distribution of the vector composed of the first $q + p - 1$ elements of (21) is multivariate normal by Theorem 1.

The last element of $\mathbf{D}_{(u)n}^{-1} \sum_{t=1}^{n} \mathbf{X}'_{(u)tn} e_{t}$ is

$$\frac{\sum_{t=1}^{n} W_{tpn}^{\dagger} e_{t}}{[\sum_{t=1}^{n} (W_{tpn}^{\dagger})^{2}]^{1/2}} = \frac{\sum_{t=1}^{n} (x_{t,q+p,n} + u_{t-1}^{\dagger}) e_{t}}{[\sum_{t=1}^{n} (x_{t,q+p,n} + u_{t-1}^{\dagger})^{2}]^{1/2}} = \frac{n^{-1} \sum_{t=1}^{n} u_{t-1}^{\dagger} e_{t}}{[n^{-2} \sum_{t=1}^{n} (u_{t-1}^{\dagger})^{2}]^{1/2}} + o_{p}(1),$$

where $u_{t-1}^{\dagger} = A_{t}^{\dagger} + B_{t}^{\dagger}$. The limiting distribution of this element is that of the $\hat{\tau}_{\mu}$-statistic for case (a) and that of $\hat{\tau}_{\tau}$ for case (b). See Dickey (1976), Hasza (1977), and Dickey and Fuller (1979). □

We have presented the theorem for the positive unit root. A similar theorem holds for a root of negative one. The kinds of fixed sequences that alter the distribution for the negative unit root differ from those that alter the distribution for the positive unit root. For example, the presence of the function cos π t will produce a limiting distribution of the t-type statistic for the negative unit root case that is the limiting distribution of $-\hat{\tau}_{\mu}$, and the presence of cos π t and t cos π t will produce a limiting distribution for the t-type statistic that is the limiting distribution of $-\hat{\tau}_{\tau}$.

Theorem 2 gives the limiting distribution of coefficients that are random linear combinations of the original coefficients of model (3). Under the assumptions of Theorem 2, the estimated intercept and regression coefficient of t, the $\hat{\alpha}_1$ and $\hat{\alpha}_2$ of model (1), are *not* normally distributed in the limit. Dickey (1977) has tabulated the distribution of the t-type statistic associated with the estimated intercept $\hat{\alpha}_1$. The limiting distribution is bimodal and symmetric with 5 percent points well beyond 2.

We now consider a situation where the limiting distribution of the least squares estimator is normal.

THEOREM 3. *Let model* (1) *hold with* $m_1 = 1$ *and* $m_2, m_3, \cdots, m_p$ *less than one in absolute value. Let* e_t *be independent* $(0, \sigma^2)$ *random variables such that* $E\{|e_t|^{2+\nu}\} < L$ *for some real* L *and* $\nu > 0$. *Assume that* (10) *and* (11) *are satisfied. Assume*

$$\lim_{n\to\infty} n^{-2} \sum_{t=1}^{n} x_{t,q+p,n}^2 = \infty$$

and

$$\lim_{n\to\infty}\left[\sum_{t=1}^{n} x_{tin}^2 \sum_{t=1}^{n} x_{t,q+p,n}^2\right]^{-1} \sum_{t=1}^{n}\sum_{h=0}^{t-1} t\, x_{tin}x_{t+h,in} = 0 \qquad (22)$$

for $i = 2, \cdots, q+p-1$. *If* $\mathbf{D}_n, \hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_n$, *and* $\mathbf{G}_n^{1/2}$ *are as in Theorem* 1, *then*

$$\sigma^{-1}\mathbf{G}_n^{1/2}\mathbf{D}_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n) \to_{\mathscr{L}} N(\mathbf{O}, \mathbf{I}) \qquad \text{as} \quad n \to \infty.$$

PROOF. We have

$$\frac{\sum_{t=1}^{n} W_{tpn}e_t}{\left[\sum_{t=1}^{n} W_{tpn}^2\right]^{1/2}} = \frac{\sum_{t=1}^{n} x_{t,p+q,n}e_t}{\left[\sum_{t=1}^{n} x_{t,p+q,n}^2\right]^{1/2}} + o_p(1)$$

because $\sum_{t=1}^{n} u_{t-i}e_t = O_p(n^{1/2})$ for $i = 1, 2, \cdots, p$. Also

$$\frac{\sum_{t=1}^{n} W_{tpn}W_{tjn}}{\left[\sum_{t=1}^{n} W_{tpn}^2 \sum_{t=1}^{n} W_{tjn}^2\right]^{1/2}} = \frac{\sum_{t=1}^{n}\left(\sum_{i=1}^{p} a_{q+i,q+p,n}u_{t-i}\right)\left(\sum_{i=1}^{p} a_{q+i,q+j,n}u_{t-i}\right)}{\left[\sum_{t=1}^{n} W_{tpn}^2 \sum_{t=1}^{n} W_{tjn}^2\right]^{1/2}} = o_p(1)$$

for $j = 1, 2, \cdots, p-1$ because $\sum_{t=1}^{n} u_{t-i}(u_{t-j} - u_{t-j-1}) = O_p(n)$ for $i, j = 1, 2, \cdots, p$. Similarly

$$\frac{\sum_{t=1}^{n} W_{tpn}x_{tin}}{\left[\sum_{t=1}^{n} W_{tpn}^2 \sum_{t=1}^{n} x_{tin}^2\right]^{1/2}} = \frac{\sum_{t=1}^{n}\sum_{j=1}^{p} a_{q+p,q+j,n}u_{t-j}x_{tin}}{\left[\sum_{t=1}^{n} x_{t,q+p,n}^2 \sum_{t=1}^{n} x_{tin}^2\right]^{1/2}} = o_p(1)$$

for $i = 1, 2, \cdots, q$ by (20) and assumption (22). Therefore, $\mathbf{D}_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n) = \mathbf{G}_n^{-1}\,\mathbf{d} + o_p(1)$, where

$$\mathbf{d}' = \left[\frac{\sum_{t=1}^{n} x_{t1n}e_t}{\left[\sum_{t=1}^{n} x_{t1n}^2\right]^{1/2}}, \cdots, \frac{\sum_{t=1}^{n} W_{t,p-1,n}e_t}{\left[\sum_{t=1}^{n} W_{t,p-1,n}^2\right]^{1/2}}, \frac{\sum_{t=1}^{n} x_{t,q+p,n}e_t}{\left[\sum_{t=1}^{n} x_{t,q+p,n}^2\right]^{1/2}}\right].$$

The result follows by an argument similar to that of Theorem 1. □

Under the conditions of Theorem 3 the fixed portion of Y_t dominates the behavior of Y_t and we obtain the limiting normal distribution. Under the conditions of Theorem 2 the random portion of Y_t dominates the behavior of Y_t and we do not obtain the normal distribution in the limit.

Neither theorem covers the situation in which the fixed and random portions of Y_t grow at the same rate, for example, where

$$\lim_{n\to\infty} n^{-2} \sum_{t=1}^{n} x_{t,p+q,n}^2 = K,$$

and $0 < K < \infty$. Because $x_{t,p+q,n}$ depends upon the parameters of the model there seems no simple general characterization of the limiting distribution when the fixed and random portions grow at the same rate. It does seem that the percentiles of the limiting distribution of the regression t-type statistics would fall between those of Theorem 2 and those of Theorem 3.

5. The explosive case. We now consider equation (1) assuming the roots m_1, m_2, $\cdots$, m_p of (2) satisfy $|m_1| > 1$ and $|m_j| < 1$ for $j = 2, 3, \cdots, p$. In this section we assume the e_t to be normal independent $(0, \sigma^2)$ random variables.

Using the difference equation (4) we obtain

$$Y_t - m_1 Y_{t-1} = \sum_{j=0}^{p-2} v_{t+j}^*(Y_{-j} - m_1 Y_{-j-1}) + \sum_{i=1}^{q} \alpha_i \psi_{ti}^* + u_t^*, \tag{23}$$

where the v_j^* are weights satisfying the homogeneous difference equation with characteristic equation (5) and initial conditions $v_0^* = 1$ and $v_j^* = 0$ for $j < 0$,

$$\psi_{ts}^* = \sum_{j=0}^{t-1} v_j^* \psi_{t-j,s} \qquad \text{and} \qquad u_t^* = \sum_{j=0}^{t-1} v_j^* e_{t-j}.$$

The solution of difference equation (23) is

$$Y_t = m_1^t Y_0 + m_1^t \sum_{j=1}^{t} \sum_{i=0}^{p-2} m_1^{-j} v_{j+i}^*(Y_{-i} - m_1 Y_{-i-1})$$
$$+ m_1^t \sum_{i=1}^{q} \alpha_i \sum_{j=1}^{t} m_1^{-j} \psi_{ji}^* + m_1^t \sum_{j=1}^{t} m_1^{-j} u_j^*.$$

Lemma 1 demonstrates that, given assumption (10), the quantity $\sum_{j=1}^{t} m_1^{-j} \psi_{ji}^*$ converges as t increases.

LEMMA 1. *Let $\{x_t\}_{t=1}^{\infty}$ be a sequence of real numbers. Let $|m| > 1$ and suppose*

$$\lim_{n\to\infty} \sup_{1\le t\le n} \left(\sum_{j=1}^{n} x_j^2\right)^{-1} x_t^2 = 0. \tag{24}$$

Then $\sum_{t=1}^{n} m^{-t} x_t$ converges and $\sum_{t=n}^{\infty} m^{-t} x_t = O(\lambda^n)$ for some $0 < \lambda < 1$.

PROOF. Let $S_n = \sum_{t=1}^{n} x_t^2$. Condition (24) implies that given $\epsilon > 0$, there exists an N such that for $n > N$, $x_n^2 < \epsilon S_{n-1}$. Therefore, $S_n < (1 + \epsilon) S_{n-1}$ for all $n > N$. Choose $\epsilon > 0$ and $N_1 > N$ so that $1 + \epsilon < \rho < |m|$ and $S_n < \rho^n$ for all $n > N_1$. Then $|x_n| < \rho^n$ for all $n > N_1$. This implies that $\sum_{t=1}^{n} m^{-t} x_t$ converges and that $\sum_{t=n+1}^{\infty} m^{-t} x_t < m^{-n} \rho^n (1 - m^{-1}\rho)$ for $n > N_1$. □

On the basis of Lemma 1 we define the quantities

$$\bar{u} = \operatorname{plim}_{t\to\infty} \bar{u}_t \qquad \text{and} \qquad \bar{\psi}_j = \lim_{t\to\infty} \bar{\psi}_{tj},$$

where

$$\bar{u}_t = \sum_{i=1}^{t} m_1^{-i} u_i^* \qquad \text{and} \qquad \bar{\psi}_{tj} = \sum_{i=1}^{t} m_1^{-i} \psi_{ij}^*.$$

Then we may write Y_t as

$$Y_t = m_1^t(\bar{A}_t + \bar{u}_t) = m_1^t H_t, \tag{25}$$

where $\bar{A}_t = Y_0 + \sum_{i=0}^{p-2} \sum_{j=1}^{t} m_1^{-j} v_{j+i}^*(Y_{-i} - m_1 Y_{-i-1}) + \sum_{i=1}^{q} \alpha_i \bar{\psi}_{ti}$.
We now obtain the probability limit of the properly normalized sum of squares of Y_{t-1}.

LEMMA 2. *Let the model of this section hold. Then*

$$\operatorname{plim}_{n\to\infty} m_1^{-2n} \sum_{t=1}^{n} Y_{t-1}^2 = (m_1^2 - 1)^{-1} H^2,$$

where $H = \operatorname{plim} H_t$ and H_t is defined in (25).

PROOF. The sum of squares

$$\sum_{t=1}^{n} Y_{t-1}^2 = \sum_{t=1}^{n} m_1^{2t-2} H^2 + 2 \sum_{t=1}^{n} m_1^{2t-2} H(H_{t-1} - H)$$
$$+ \sum_{t=1}^{n} m_1^{2t-2} (H_{t-1} - H)^2.$$

We have

$$E\{(H_{t-1}-H)^2\} = E\{(\textstyle\sum_{j=t}^{\infty} m_1^{-j}u_j^* + \sum_{i=0}^{p-1}\sum_{j=t}^{\infty} m_1^{-j}v_{j+i}^*(Y_{-i}-m_1Y_{-i-1}) + \sum_{i=1}^{q}\sum_{j=t}^{\infty}\alpha_i m_1^{-j}\psi_{ji}^*)^2\} = O(\lambda^{2t})$$

for some $0<\lambda<1$. It follows that

$$\textstyle\sum_{t=1}^{n} m_1^{2t-2-2n}(H_t-H)^2 = O_p(\sum_{t=1}^{n} m_1^{2t-2n}\lambda^{2t}) = O_p(\lambda^{2n}).$$

Therefore, $m_1^{-2n}\sum_{t=1}^{n} Y_{t-1}^2 = (m_1^2-1)^{-1}H^2 + O_p(\lambda^{2n})$. □

We next establish the asymptotic behavior of the normalized sum of the cross products of Y_{t-1} with x_{tin}.

LEMMA 3. *Let the model of this section hold. Then, for $i = 1, 2, \cdots, q$,*

$$[m_1^{2n}\textstyle\sum_{t=1}^{n} x_{tin}^2]^{-1/2}\sum_{t=1}^{n} x_{tin}Y_{t-1}\to_P 0.$$

PROOF. We have

$$\textstyle\sum_{t=1}^{n} x_{tin}Y_{t-1} = \sum_{t=1}^{n} x_{tin}m_1^{t-1}H + \sum_{t=1}^{n} x_{tin}(H_{t-1}-H)m_1^{t-1}.$$

By the proof of Lemma 2, $E\{(H_{t-1}-H)^2\} = O(\lambda^{2t})$, where $|\lambda|<1$, and

$$\textstyle\sum_{t=1}^{n} x_{tin}(H_{t-1}-H)m_1^{t-1} = O_p(\sum_{t=1}^{n} x_{tin}\lambda^t m_1^{t-1}).$$

By Lemma 1

$$(\textstyle\sum_{t=1}^{n} x_{tin}^2)^{-1/2}\sum_{t=1}^{n} x_{tin}\lambda^t m_1^{-n+t-1} \le [(\sum_{t=1}^{n} x_{tin}^2)^{-1/2}\sup_{1\le t\le n}|x_{tin}|](1-|\lambda|)^{-1}$$

which converges to zero as $n\to\infty$. □

THEOREM 4. *Let model* (1) *hold with $|m_1|>1$ and $m_2, m_3, \cdots, m_p$ less than one in absolute value. Let the e_t be normal independent* $(0, \sigma^2)$ *random variables. Let*

$$\lim_{n\to\infty}\sup_{1\le t\le n}[\textstyle\sum_{s=1}^{n} x_{sin}^2]^{-1}x_{tin}^2 = 0$$

for $i = 1, 2, \cdots, q$ and

$$\lim_{n\to\infty}\sup_{1\le t\le n}[n+\textstyle\sum_{s=1}^{n} x_{sin}^2]^{-1}x_{tin}^2 = 0$$

for $i = q+1, q+2, \cdots, q+p-1$. Let $\boldsymbol\theta_n, \hat{\boldsymbol\theta}_n, \mathbf{X}_{tn}, \mathbf{G}_n, \mathbf{D}_n$, and $\mathbf{G}_n^{1/2}$ be as defined in Theorem 1. Then

$$\sigma^{-1}\mathbf{G}_n^{1/2}\mathbf{D}_n(\hat{\boldsymbol\theta}_n-\boldsymbol\theta_n)\to_{\mathscr{L}} N(\mathbf{O}, \mathbf{I}).$$

PROOF. We have

$$\mathbf{D}_n(\hat{\boldsymbol\theta}_n-\boldsymbol\theta_n) = \mathbf{G}_n^{-1}\mathbf{D}_n^{-1}\textstyle\sum_{t=1}^{n}\mathbf{X}_{tn}'e_t.$$

By Lemmas 2 and 3, the last row of $\mathbf{G}_n$ is converging to the vector $(0, 0, \cdots, 0, 1)$. Now the last element of $\mathbf{D}_n^{-1}\sum_{t=1}^{n}\mathbf{X}_{tn}'e_t$ is

$$[\textstyle\sum_{t=1}^{n} Y_{t-1}^2]^{-1/2}\sum_{t=1}^{n} Y_{t-1}e_t + o_p(1).$$

By Lemma 2 and by arguments similar to those used in the proof of Lemma 2,

$$m_1^n[\textstyle\sum_{t=1}^{n} Y_{t-1}^2]^{-1/2} = (m_1^2-1)^{1/2}H^{-1} + O_p(\lambda^{2n})$$

for some $0<\lambda<1$ and

(26) $$m_1^{-n} \sum_{t=1}^n Y_{t-1}e_t = H \sum_{j=1}^n m_1^{-j} e_{n-j+1} + o_p(1).$$

It follows that

$$[\sum_{t=1}^n Y_{t-1}^2]^{-1/2} \sum_{t=1}^n Y_{t-1}e_t \to_{\mathscr{L}} N(0, 1).$$

By the arguments of Theorem 1, the vector of the first $q + p - 1$ elements of $\mathbf{D}_n^{-1} \sum_{t=1}^n \mathbf{X}'_{tn} e_t$ converges to a multivariate normal random variable. From (26) it is clear that the last element is independent of the first $q + p - 1$ elements in the limit. □

It follows from Theorem 4 that the usual regression test statistics associated with model (1) are applicable in large samples if the e_t are normally distributed. The assumption of normal e_t was necessary to obtain the normal limiting distribution for $(\sum_{t=1}^n Y_{t-1}^2)^{-1/2} \sum_{t=1}^n Y_{t-1}e_t$. This quantity is $O_p(1)$ under the milder assumptions on the e_t used in Section 3.

Acknowledgments. This research was partly supported by Joint Statistical Agreements J.S.A. 77-3 and J.S.A. 78-30 with the U.S. Census Bureau. We express our appreciation to the referee who identified a flaw in the original manuscript and made useful suggestions for improvement.

REFERENCES

AMEMIYA, T. and FULLER, W. A. (1967). A comparative study of alternative estimators in a distributed lag model. *Econometrica* **35** 509–529.

ANDERSON, T. W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *Ann. Math. Statist.* **30** 676–687.

ANDERSON, T. W. and RUBIN, H. (1950). The asymptotic properties of estimates of the parameters in a single equation in a complete system of stochastic equations. *Ann. Math. Statist.* **21** 570–582.

BELLMAN, R. (1960). *Introduction to Matrix Analysis.* McGraw-Hill, New York.

BROWN, B. M. (1971). Martingale central limit theorems. *Ann. Math. Statist.* **42** 59–66.

DICKEY, D. A. (1976). Estimation and hypothesis testing in nonstationary time series. Unpublished Ph.D. thesis, Iowa State Univ., Ames, Iowa.

DICKEY, D. A. (1977). Distributions associated with the nonstationary autoregressive process. Paper presented at the eastern regional meeting of the Institute of Mathematical Statistics, Chapel Hill, N.C. (April, 1977).

DICKEY, D. A. and FULLER, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Amer. Statist. Assoc.* **74** 427–431.

DVORETZKY, A. (1972). Asymptotic normality for sums of dependent random variables. In *Sixth Berkeley Symp. Math. Statist. Prob.* **2** 513–535.

FULLER, W. A. (1976). *Introduction to Statistical Time Series.* Wiley, New York.

GRENANDER, U. (1954). On the estimation of regression coefficients in the case of an autocorrelated disturbance. *Ann. Math. Statist.* **25** 252–272.

HANNAN, E. J. (1965). The estimation of relationships involving distributed lags. *Econometrica* **33** 206–224.

HANNAN, E. J. (1970). *Multiple Time Series.* Wiley, New York.

HANNAN, E. J. and HEYDE, C. C. (1972). On limit theorems for quadratic functions of discrete time series. *Ann. Math. Statist.* **43** 2058–2066.

HANNAN, E. J. and NICHOLLS, D. F. (1972). The estimation of mixed regression, autoregression, moving average and distributed lag models. *Econometrica* **40** 529–548.

HATANAKA, M. (1974). An efficient two-step estimator for the dynamic adjustment model with autoregressive errors. *J. Econometrics* **2** 199–220.

HASZA, D. P. (1977). Estimation in nonstationary time series. Unpublished Ph.D. thesis. Iowa State Univ., Ames, Iowa.

KOOPMANS, T. C., RUBIN, H. and LEIPNIK, R. B. (1950). Measuring the equation systems of dynamic economics. In *Statistical Inference in Dynamic Economic Models.* (T. C. Koopmans, ed.) Wiley, New York.

MANN, H. B. and WALD, A. (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica* **11** 173–220.

NARASIMHAM, G. V. L. (1969). Some properties of estimators occurring in the theory of linear stochastic processes. In *Lecture Notes in Operations Research and Mathematical Economics.* (M. Beckman and H. P. Künzi, eds.) Springer, Berlin.

NICHOLLS, D. F. (1976). The efficient estimation of vector linear time series models. *Biometrika* **63** 381–390.

RAO, C. R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. In *Fifth Berkeley Symp. Math. Statist. Prob.* **1** 355–372.

RAO, M. M. (1961). Consistency and limit distributions of estimators of parameters in explosive stochastic difference equations. *Ann. Math. Statist.* **32** 195–218.

RAO, M. M. (1978). Asymptotic distribution of an estimator of the boundary parameter of an unstable process. *Ann. Statist.* **6** 185–190.

REINSEL, G. (1976). Maximum likelihood estimation of stochastic linear difference equations with autoregressive moving average errors. Technical Report 112, Dept. Statist., Carnegie-Mellon Univ.

RUBIN, H. (1950). Consistency of maximum-likelihood estimates in the explosive case. In *Statistical Inference in Dynamic Economic Models.* (T. C. Koopmans, ed.) Wiley, New York.

SCOTT, D. J. (1973). Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach. *Advances Appl. Probability* **5** 119–137.

STIGUM, B. P. (1974). Asymptotic properties of dynamic stochastic parameter estimates (III). *J. Multivariate Anal.* **4** 351–381.

VENKATARAMAN, K. N. (1967). A note on the least square estimators of the parameters of a second order linear stochastic difference equation. *Calcutta Statist. Assoc. Bull.* **16** 15–28.

VENKATARAMAN, K. N. (1968). Some limit theorems on a linear stochastic difference equation with a constant term, and their statistical applications. *Sankhyā Series A* **30** 51–74.

VENKATARAMAN, K. N. (1973). Some convergence theorems on a second order linear explosive stochastic difference equation with a constant term. *J. Indian Statist. Assoc.* **11** 47–69.

WHITE, J. S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *Ann. Math. Statist.* **29** 1188–1197.

WAYNE A. FULLER
DEPARTMENT OF STATISTICS
SNEDECOR HALL
IOWA STATE UNIVERSITY
AMES, IOWA 50011

DAVID P. HASZA
DEPARTMENT OF STATISTICS
CALVIN HALL 101B
KANSAS STATE UNIVERSITY
MANHATTAN, KANSAS 66506

J. JEFFERY GOEBEL
STATISTICS LABORATORY
IOWA STATE UNIVERSITY
AMES, IOWA 50011