Introduction to Statistical Time Series
WAYNE A. FULLER

CHAPTER 10

Unit Root and Explosive Time Series

In Chapter 9, we studied time series in which the mean was a function of time. A second important class of nonstationary time series are those that satisfy nonstationary stochastic difference equations. We begin our discussion with least squares estimators for the univariate unit root autoregressive process. Alternative estimation procedures for unit processes are discussed in Section 10.1.3. Tests based on the methods of Section 10.1.3 have been demonstrated to be more powerful against the unknown mean stationary alternative than tests based on ordinary least squares regression. The remaining sections are devoted to autoregressive processes with a root greater than one, multivariate unit root autoregressive processes, and time series with a moving average unit root.

10.1. UNIT ROOT AUTOREGRESSIVE TIME SERIES

10.1.1. The Autoregressive Process with a Unit Root

Consider the first order time series

$$\begin{aligned} &Y_t = \rho Y_{t-1} + e_t , \qquad t = 1, 2, \ldots , \\ &Y_0 = 0 , \end{aligned} \tag{10.1.1}$$

where the e_t are normal independent $(0, \sigma^2)$ random variables. In Section 2.1, we showed that the solution to the difference equation is

$$Y_t = \sum_{j=0}^{t-1} \rho^j e_{t-j} . \tag{10.1.2}$$

One case of particular interest is obtained when $\rho = 1$. Then Y_t is the sum of t independent $(0, \sigma^2)$ random variables. The time series (10.1.1) with $\rho = 1$ is sometimes called a *random walk*.

By the results of Section 8.1, the maximum likelihood estimator of ρ conditional on Y_1, for any real valued ρ, is

$$\hat{\rho} = \left(\sum_{t=2}^{n} Y_{t-1}^2\right)^{-1} \sum_{t=2}^{n} Y_t Y_{t-1}, \tag{10.1.3}$$

and

$$\hat{\rho} - \rho = \left(\sum_{t=2}^{n} Y_{t-1}^2\right)^{-1} \sum_{t=2}^{n} Y_{t-1} e_t. \tag{10.1.4}$$

If $\rho = 1$, then $Y_t = \Sigma_{j=1}^{t} e_j$ and the mean and variance of the numerator and the denominator of (10.1.4) can be computed. We have

$$\begin{aligned} E\left\{\sum_{t=2}^{n} Y_{t-1}^2\right\} &= E\left\{\sum_{t=1}^{t-1}\left(\sum_{i=1}^{t} e_i\right)^2\right\} \\ &= \tfrac{1}{2}n(n-1)\sigma^2, \end{aligned} \tag{10.1.5}$$

$$\operatorname{Var}\left\{\sum_{t=2}^{n} Y_{t-1}^2\right\} = \tfrac{1}{3}n(n-1)(n^2-n+1)\sigma^4, \tag{10.1.6}$$

$$\operatorname{Var}\left\{\sum_{t=2}^{n} Y_{t-1}e_t\right\} = \sum_{t=1}^{n-1} t\sigma^4 = \tfrac{1}{2}n(n-1)\sigma^4, \tag{10.1.7}$$

$$\operatorname{Cov}\left\{\sum_{t=2}^{n} Y_{t-1}e_t, \sum_{t=2}^{n} Y_{t-1}^2\right\} = \tfrac{1}{3}n(n-1)(n-2)\sigma^4. \tag{10.1.8}$$

Normality was used in obtaining (10.1.6) and (10.1.8). The other results remain true for e_t that are iid$(0, \sigma^2)$.

These results differ in a notable way from those obtained for $|\rho| < 1$. When $|\rho| < 1$, every entry in the covariance matrix of $(\Sigma_{t=2}^{n} Y_{t-1}e_t, \Sigma_{t=2}^{n} Y_{t-1}^2)$ is of order n, and the limit of the expected value of $n^{-1}\Sigma_{t=2}^{n} Y_{t-1}^2$ is $\sigma^2(1-\rho^2)^{-1}$. If $\rho = 1$, the variance of the numerator of (10.1.4) is order n^2 instead of order n. Also, the mean of the denominator of (10.1.4) is of order n^2, and the variance is order n^4. The denominator divided by $0.5n(n-1)\sigma^2$ gives a random variable with mean one and variance $4(n^2-n+1)[3(n^2-n)]^{-1}$. Because the variance converges to 4/3, the normalized denominator does not converge to a constant in mean square.

The moment results and the fact that ΣY_t^2 is a quadratic form in $(e_1, e_2, \ldots, e_n)$ can be used to show that the error in $\hat{\rho}$, as an estimator of one, is $O_p(n^{-1})$. Thus, for $\rho = 1$, the least squares estimator converges in probability to the true value more rapidly than does the estimator for $|\rho| < 1$.

Lemma 10.1.1. *Let $\hat{\rho}$ be defined by (10.1.3), and assume Y_t satisfies (10.1.2) with $\rho = 1$, where the e_t are normal independent $(0, \sigma^2)$ random variables. Then*

$$\hat{\rho} - 1 = O_p(n^{-1}) .$$

Proof. The denominator of $\hat{\rho}$ can be written as

$$\sum_{t=2}^{n} Y_{t-1}^2 = \mathbf{e}'\mathbf{A}_n\mathbf{e} ,$$

where $\mathbf{e}' = (e_1, e_2, \ldots, e_{n-1})$ and

$$\mathbf{A}_n = \begin{pmatrix} n-1 & n-2 & n-3 & \cdots & 1 \\ n-2 & n-2 & n-3 & \cdots & 1 \\ n-3 & n-3 & n-3 & \cdots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix} . \qquad (10.1.9)$$

The matrix $\mathbf{A}_n$ can be diagonalized by the $(n-1) \times (n-1)$ orthonormal matrix $\mathbf{M}_n$ whose ijth element is

$$m_{nij} = 2(2n-1)^{-1/2} \cos[(4n-2)^{-1}(2j-1)(2i-1)\pi] . \qquad (10.1.10)$$

Then

$$\mathbf{M}_n\mathbf{A}_n\mathbf{M}_n' = \mathbf{\Lambda}_n \equiv \mathrm{diag}(\lambda_{1n}, \lambda_{2n}, \ldots, \lambda_{n-1,n}) ,$$

where

$$\lambda_{in} = \tfrac{1}{4}\sec^2[(n-i)\pi(2n-1)^{-1}] \qquad (10.1.11)$$

for $i = 1, 2, \ldots, n-1$. See Rutherford (1946). Hence, the denominator of $\hat{\rho} - 1$ can be written as

$$\mathbf{e}'\mathbf{A}_n\mathbf{e} = \sum_{i=1}^{n-1} \lambda_{in} U_{in}^2 , \qquad (10.1.12)$$

where

$$U_{in} = \sum_{t=1}^{n-1} m_{nit} e_t , \qquad i = 1, 2, \ldots, n-1 ,$$

and the U_{in} are $\mathrm{NI}(0, \sigma^2)$ random variables. By the moment results (10.1.5) and (10.1.6)

$$\sum_{i=1}^{n-1} \lambda_{in} = \tfrac{1}{2}n(n-1)$$

$$\sum_{i=1}^{n-1} \lambda_{in}^2 = \tfrac{1}{6}n(n-1)(n^2-n+1) .$$

If we fix i and take the limit as $n \to \infty$, we have

$$\lim_{n\to\infty} n^{-2}\lambda_{in} = \gamma_i^2 \equiv 4[(2i-1)\pi]^{-2}$$

for $i = 1, 2, \ldots$. Therefore, the largest root, λ_{1n}, increases at the rate n^2 as n increases. Given $\epsilon > 0$,

$$\begin{aligned} P\left\{\sum_{i=1}^{n-1} \lambda_{in} U_{in}^2 < \tfrac{1}{2}n(n-1)\epsilon\right\} &\leq P\{\lambda_{1n} U_{1n}^2 < 0.5n(n-1)\epsilon\} \\ &\leq P\{U_{1n}^2 < 2\epsilon\}. \end{aligned}$$

Because U_{1n}^2 is a chi-square random variable, given $\Delta > 0$, there exists an $\epsilon > 0$ such that $P\{U_{1n}^2 < 2\epsilon\} < \Delta$. It follows that, given $\Delta > 0$, there exists a $K\epsilon^{-1}$ such that

$$P\{(\mathbf{e}'\mathbf{A}_n\mathbf{e})^{-1} n(n-1) > K\epsilon^{-1}\} < \Delta$$

for all n and, by (10.1.6),

$$n(n-1)\left[\sum_{t=2}^{n} Y_{t-1}^2\right]^{-1} = O_p(1).$$

Because $n^{-2}\sum_{t=2}^{n} Y_{t-1}e_t = O_p(n^{-1})$, it follows that $\hat{\rho} - 1 = O_p(n^{-1})$. ▲

The numerator of $\hat{\rho} - 1$ can be written in an alternative informative manner as

$$\sum_{t=2}^{n} Y_{t-1}e_t = \sum_{t=2}^{n}\sum_{j=1}^{t-1} e_j e_t = \frac{1}{2}\left(Y_n^2 - \sum_{t=1}^{n} e_t^2\right). \tag{10.1.13}$$

Hence, as $n \to \infty$,

$$2(n\sigma^2)^{-1}\sum_{t=2}^{n} Y_{t-1}e_t + 1 \xrightarrow{\mathcal{L}} \chi_1^2,$$

where χ_1^2 is the chi-square distribution with one degree of freedom. The probability that a one-degree-of-freedom chi-square random variable is less than one is 0.6826. Therefore, because the denominator is always positive, the probability that $\hat{\rho} < 1$, given $\rho = 1$, approaches 0.6826 as n gets large. Although a chi-square distribution is skewed to the right, the high correlation between the numerator and denominator of $\hat{\rho} - 1$ strongly dampens the skewness. In fact, the distribution of $\hat{\rho}$ displays skewness to the left.

To establish the limiting distribution of the least squares estimator of ρ, given that $\rho = 1$, we will use the following lemma.

Lemma 10.1.2. Let $\{U_{tn}: 1 \leq t \leq n, n \geq 1\}$ denote a triangular array of random variables defined on the probability space $(\Omega, \mathcal{A}, P)$. Assume

$$E[(U_{tn}, U_{tn}^2, U_{tn}U_{sn})] = (0, \sigma^2, 0)$$

for $1 \le t \ne s \le n$. Let $\{w_i: i = 1, 2, \ldots\}$ be a sequence of real numbers, and let $\{w_{in}: i = 1, 2, \ldots, n;\ n = 1, 2, \ldots\}$ be a triangular array of real numbers. If

$$\lim_{n\to\infty} \sum_{i=1}^{n} w_{in}^2 = \sum_{i=1}^{\infty} w_i^2 < \infty$$

and

$$\lim_{n\to\infty} w_{in} = w_i\,, \qquad i = 1, 2, \ldots,$$

then

$$\sum_{i=1}^{n} w_{in}U_{in} = \sum_{i=1}^{n} w_i U_{in} + o_p(1)\,.$$

Proof. Let $\epsilon > 0$ be given. Then we can choose an M such that

$$\sigma^2 \sum_{i=M+1}^{\infty} w_i^2 < \frac{\epsilon}{9}$$

and

$$\sigma^2 \left| \sum_{i=1}^{n} w_{in}^2 - \sum_{i=1}^{\infty} w_i^2 \right| < \frac{\epsilon}{9}$$

for all $n > M$. Furthermore, given M, we can choose $N_0 > M$ such that $n > N_0$ implies

$$\sigma^2 \sum_{i=1}^{M} (w_{in} - w_i)^2 < \frac{\epsilon}{9}$$

and

$$\sigma^2 \sum_{i=M+1}^{n} w_{in}^2 < \frac{3\epsilon}{9}\,.$$

Hence, for all $n > N_0$,

$$V\left\{\sum_{i=1}^{n} w_{in}U_{in} - \sum_{i=1}^{n} w_i U_{in}\right\} < \epsilon\,. \qquad \blacktriangle$$

We now give the primary result for the distribution of the least squares estimator of the autoregressive parameter when the true process is a random walk.

Theorem 10.1.1. Let

$$Y_t = Y_{t-1} + e_t\,, \qquad t = 1, 2, \ldots,$$

where $Y_0 = 0$. Assume $\{e_t\}_{t=1}^{n}$ satisfies

$$E\{(e_t, e_t^2) \mid \mathscr{A}_{t-1}\} = (0, \sigma^2) \qquad \text{a.s.},$$
$$E\{|e_t|^{2+\delta} \mid \mathscr{A}_{t-1}\} < M < \infty \qquad \text{a.s.}$$

for some $\delta > 0$, where $\mathscr{A}_t$ is the sigma-field generated by $\{e_1, e_2, \ldots, e_t\}$. Then

$$n(\hat{\rho} - 1) \xrightarrow{\mathscr{L}} (2G)^{-1}(T^2 - 1),$$

where $\hat{\rho}$ is defined in (10.1.3),

$$(G, T) = \left(\sum_{i=1}^{\infty} \gamma_i^2 U_i^2, \sum_{i=1}^{\infty} 2^{1/2} \gamma_i U_i \right),$$
$$\gamma_i = (-1)^{i+1} 2[(2i-1)\pi]^{-1},$$

$\{U_i\}_{i=1}^{\infty}$ is a sequence of NI(0, 1) random variables, and (G, T) is defined as a limit in mean square.

Proof. The estimator $\hat{\rho}$ is scale invariant. Hence, there is no loss of generality in assuming $\sigma^2 = 1$. Let

$$(G_n, T_n) = \left(n^{-2} \sum_{t=2}^{n} Y_{t-1}^2, n^{-1/2} \sum_{t=1}^{n-1} e_t \right).$$

From the definition of $\hat{\rho}$ and from (10.1.13), we have

$$n(\hat{\rho} - 1) = \left(n^{-2} \sum_{t=2}^{n} Y_{t-1}^2 \right)^{-1} n^{-1} \sum_{t=2}^{n} Y_{t-1} e_t$$
$$= (2G_n)^{-1}(T_n^2 - 1) + o_p(1),$$

where $n^{-1} \sum_{t=1}^{n} e_t^2 = \sigma^2 + o_p(1)$ by Corollary 5.3.5.

To establish the limit distribution result, we show that (G_n, T_n) converges in distribution to (G, T). Let

$$\mathbf{U}_n = (U_{1n}, U_{2n}, \ldots, U_{n-1,n})' = \mathbf{M}_n \mathbf{e}_n,$$

where $\mathbf{e}_n = (e_1, e_2, \ldots, e_{n-1})'$ and $\mathbf{M}_n$ is the matrix with (i, j)th element given by (10.1.10). Then

$$T_n = n^{-1/2} \mathbf{J}_n' \mathbf{e}_n = n^{-1/2} \mathbf{J}_n' \mathbf{M}_n^{-1} \mathbf{U}_n,$$

where $\mathbf{J}_n = (1, 1, \ldots, 1)'$. Let k_{in} be the ith element of $n^{-1/2}\mathbf{J}_n'\mathbf{M}_n^{-1}$. The element k_{in} is the covariance between T_n and U_{in}. Using this fact and the results of Jolley (1961, p. 78), it can be shown that, for fixed i,

$$\lim_{n \to \infty} k_{in} = 2^{1/2} \gamma_i .$$

See Dickey (1976, p. 29). Also, $2\sum_{i=1}^{\infty}\gamma_i^2 = 1$; see Jolley (1961, p. 56). Let

$$(G_n, T_n) = \left(\sum_{i=1}^{n-1} n^{-2}\lambda_{in}U_{in}^2, \sum_{i=1}^{n-1} k_{in}U_{in}\right),$$

$$(\overset{*}{G}_n, \overset{*}{T}_n) = \left(\sum_{i=1}^{n-1} \gamma_i^2 U_{in}^2, \sum_{i=1}^{n-1} 2^{1/2}\gamma_i U_{in}\right),$$

where λ_{in} is defined in (10.1.11). Then $T_n - \overset{*}{T}_n$ converges in probability to zero by Lemma 10.1.2. Also,

$$G_n - \overset{*}{G}_n = \sum_{i=1}^{n-1} (n^{-2}\lambda_{in} - \gamma_i^2)U_{in}^2 = o_p(1)$$

because

$$\lim_{n\to\infty} \sum_{i=1}^{n-1} |n^{-2}\lambda_{in} - \gamma_i^2| = 0 .$$

The entries in $n^{1/2}\mathbf{M}_n$ are bounded, and the sum of squares of each row of $\mathbf{M}_n$ is one. Therefore, by Corollary 5.3.4, for fixed l,

$$(U_{1n}, U_{2n}, \ldots, U_{ln})' \xrightarrow{\mathscr{L}} (U_1, U_2, \ldots, U_l)' ,$$

where $(U_1, U_2, \ldots, U_l) \sim N(\mathbf{0}, \mathbf{I})$. Note that

$$(\overset{*}{G}_n, \overset{*}{T}_n) = \left(\sum_{i=1}^{l} \gamma_i^2 U_{in}^2, \sum_{i=1}^{l} 2^{1/2}\gamma_i U_{in}\right) + \left(\sum_{i=l+1}^{n-1} \gamma_i^2 U_{in}^2, \sum_{i=l+1}^{n-1} 2^{1/2}\gamma_i U_{in}\right).$$

Because the sequence $\{\gamma_i^2\}_{i=1}^{\infty}$ is summable, the vector

$$\left(\sum_{i=l+1}^{n-1} \gamma_i^2 U_{in}^2, \sum_{i=l+1}^{n-1} 2^{1/2}\gamma_i U_{in}\right)$$

converges to zero in probability as l tends to infinity, uniformly in n. Therefore, by Lemma 6.3.1, (G_n, T_n) converges in distribution to (G, T) and the result is established. ▲

We gave the proof of Theorem 10.1.1 for e_t that are martingale differences. The result also holds for e_t that are iid$(0, \sigma^2)$ random variables.

The limiting distribution of Theorem 10.1.1 can be obtained by using Theorems 5.3.5 and 5.3.6. By Corollary 5.3.6, with $\sigma^2 = 1$,

$$(G_n, T_n) \xrightarrow{\mathscr{L}} (G, T) ,$$

where

$$(G, T) = \left[\int_0^1 W^2(t)\, dt, W(1) \right] \tag{10.1.14}$$

and $W(t)$ is the standard Wiener process defined in Section 5.3.

In Theorem 10.1.1, we assumed $Y_0 = 0$ to simplify the presentation. Because

$$p\lim \left[\sum_{t=1}^{n} \left(\sum_{j=1}^{t} e_j \right)^2 \right]^{-1} \sum_{t=1}^{n} \left(Y_0 + \sum_{j=1}^{t} e_j \right)^2 = 1$$

for any finite Y_0, the conclusion of Theorem 10.1.1 is true for any finite Y_0.

The first part of Table 10.A.1 contains percentiles for the empirical distribution of $\hat{\rho}$ given that $\rho = 1$. Normally distributed errors were used to construct the table for $n = 25, 50, 100, 250$, and 500. The total number of observations available is n, while $n - 1$ sums of squares and products are used in calculating the regression coefficient. The entries for $n = \infty$ are the percentiles for the limiting distribution of Theorem 10.1.1.

The table may also be used for the distribution of $\hat{\rho}$ for $\rho = -1$, since, for symmetrically distributed e_t, the distribution for $\rho = -1$ is the mirror image of the distribution for $\rho = 1$. If the e_t are not symmetrically distributed, the limiting distribution for $\rho = -1$ remains the mirror image of the limiting distribution for $\rho = 1$.

Corollary 10.1.1.1. Let the model of Theorem 8.5.1 hold. Then

$$\lim_{n \to \infty} (P\{\hat{\rho} - \rho > a \mid \rho = 1\} - P\{\hat{\rho} - \rho < -a \mid \rho = -1\}) = 0$$

for all real a. If, in addition, the e_t are independently and identically distributed with distribution that is symmetric about zero, then

$$P\{\hat{\rho} - \rho > a \mid \rho = 1\} = P\{\hat{\rho} - \rho < -a \mid \rho = -1\}$$

for all real a and all $n \geq 2$.

Proof. Let $X_t = \Sigma_{j=0}^{t-1} (-1)^j e_{t-j}$. In a manner similar to that used to obtain (10.1.13), we have

$$\sum_{t=2}^{n} X_{t-1} e_t = \sum_{t=2}^{n} e_t \sum_{j=1}^{t-1} (-1)^{j-1} e_{t-j} = 0.5 \left(\sum_{t=1}^{n} e_t^2 - X_n^2 \right).$$

Therefore, for the estimator of ρ when $\rho = -1$, $P\{\hat{\rho} + 1 < a\}$ is

$$P\left\{ \sum_{t=2}^{n} X_{t-1} e_t + a \sum_{t=2}^{n} X_{t-1}^2 < 0 \right\}$$
$$= P\left\{ 0.5 \left(X_n^2 - \sum_{t=1}^{n} e_t^2 \right) - a \sum_{t=1}^{n-1} \left[\sum_{j=0}^{t-1} (-1)^j e_{t-j} \right]^2 > 0 \right\}.$$

Although the sign of e_i in the weighted sum $\Sigma_{j=0}^{t-1}(-1)^j e_{t-j}$ is not the same for all $t > i$, the sign is always opposite of that for e_{i-1} and e_{i+1}, and it follows that

$$\sum_{t=1}^{n-1} X_t^2 = \sum_{t=1}^{n-1} \left[\sum_{j=1}^{t} (-1)^j e_j\right]^2 .$$

If the e_t are iid$(0, \sigma^2)$ and if the distribution of e_t, $t = 1, 2, \ldots$, is symmetric, then the distributional properties of the sequence $(-e_1, e_2, -e_3, \ldots)$ are the same as the distributional properties of the sequence $(e_1, e_2, e_3, \ldots)$, and we conclude that, for symmetric e_t and for any a,

$$P\left\{\frac{1}{2}\left[\left(\sum_{t=1}^{n} e_t\right)^2 - \sum_{t=1}^{n} e_t^2\right] - a\sum_{t=1}^{n-1}\left(\sum_{j=1}^{t} e_j\right)^2 > 0\right\}$$
$$= P\left\{\frac{1}{2}\left[\left(\sum_{t=1}^{n} (-1)^t e_t\right)^2 - \sum_{t=1}^{n} e_t^2\right] - a\sum_{t=1}^{n-1}\left(\sum_{j=1}^{t} (-1)^j e_j\right)^2 > 0\right\}.$$

Also, the large sample distribution of $\hat{\rho}$ given $\rho = -1$ is the mirror image of the distribution for $\rho = 1$ for e_t that are not symmetric, because the U_i have a limiting normal distribution for the sequence $(-e_1, e_2, -e_3, \ldots)$ as well as for the sequence $(e_1, e_2, e_3, \ldots)$. ▲

A natural statistic to use in testing the hypothesis that $\rho = 1$ is the test statistic one would calculate in ordinary linear regression:

$$\hat{\tau} = \left[s^{-2}\sum_{t=2}^{n} Y_{t-1}^2\right]^{1/2} (\hat{\rho} - 1), \qquad (10.1.15)$$

where

$$s^2 = (n-2)^{-1}\sum_{t=2}^{n} \hat{e}_t^2 = (n-2)^{-1}\sum_{t=2}^{n} (Y_t - \hat{\rho}Y_{t-1})^2 .$$

Corollary 10.1.1.2. Let the model of Theorem 10.1.1 hold. Then

$$\hat{\tau} \xrightarrow{\mathscr{L}} (4G)^{-1/2}(T^2 - 1),$$

where $\hat{\tau}$ is defined in (10.1.15).

Proof. We have

$$s^2 = (n-2)^{-1}\left[\sum_{t=2}^{n} e_t^2 - (\hat{\rho} - \rho)\sum_{t=2}^{n} Y_{t-1}e_t\right]$$
$$= (n-2)^{-1}\sum_{t=2}^{n} e_t^2 + O_p(n^{-1}) .$$

Therefore, $s^2 \xrightarrow{P} \sigma^2$ by Corollary 5.3.8. Because

$$\begin{aligned}\hat{\tau} &= 2^{-1}(s^2 G_n)^{-1/2}\left(T_n^2 - n^{-1}\sum_{t=1}^{n} e_t^2\right)\\ &= (4G_n)^{-1/2}(T_n^2 - 1) + o_p(1),\end{aligned}$$

we have the result. ▲

Percentage points for the empirical distribution of $\hat{\tau}$ are given in Table 10.A.2. The percentiles for $n = \infty$ are the percentiles of the distribution of Corollary 10.1.1.2.

To extend the results for the first order process with $\rho = 1$ to the pth order autoregressive process, we consider the time series

$$Y_t = \sum_{j=1}^{t} Z_j, \qquad t = 1, 2, \ldots, \tag{10.1.16}$$

where $\{Z_t: t \in (0, \pm 1, \pm 2, \ldots)\}$ is a $(p-1)$st order stationary autoregressive time series with the representation

$$Z_t = \sum_{i=2}^{p} \theta_i Z_{t-i+1} + e_t, \tag{10.1.17}$$

the e_t are martingale difference random variables with the properties defined in Theorem 10.1.1, and the absolute value of the largest root of

$$m^{p-1} - \sum_{i=2}^{p} \theta_i m^{p-i} = 0 \tag{10.1.18}$$

is less than some $\lambda < 1$. By (10.1.16) and (10.1.17), we may write

$$Y_t + \sum_{j=1}^{p} \alpha_j Y_{t-j} = e_t, \qquad t = p+1, p+2, \ldots, \tag{10.1.19}$$

where $p - 1$ of the roots of

$$m^p + \sum_{j=1}^{p} \alpha_j m^{p-j} = 0$$

are the $p - 1$ roots of (10.1.18), and the remaining root is one. In a problem where a root of one is suspected, it is operationally desirable and theoretically convenient to write the autoregressive equation so that the unit root is isolated as a coefficient. To this end, we write

$$Y_t = \theta_1 Y_{t-1} + \sum_{j=2}^{p} \theta_j (Y_{t-j+1} - Y_{t-j}) + e_t \tag{10.1.20}$$

for $t = p+1, p+2, \ldots,$ where $p \geq 2$, $\theta_i = \Sigma_{j=i}^{p} \alpha_j$, $i = 2, 3, \ldots, p$, and $\theta_1 = -\Sigma_{j=1}^{p} \alpha_j$. If there is a unit root, $\theta_1 = 1$.

If one knew that $\theta_1 = 1$, one would regress $Y_t - Y_{t-1} = Z_t$ on $Z_{t-1}, Z_{t-2}, \ldots, Z_{t-p+1}$ and obtain an estimator of $(\theta_2, \theta_3, \ldots, \theta_p)$, say $(\tilde{\theta}_2, \tilde{\theta}_3, \ldots, \tilde{\theta}_p)$, with the properties described in Section 8.2. It is interesting that, given $\theta_1 = 1$, the regression of Y_t on $Y_{t-1}, Z_{t-1}, \ldots, Z_{t-p+1}$ yields an estimator of θ_1, say $\tilde{\theta}_1$, such that the large sample distribution of $nc(\tilde{\theta}_1 - 1)$ is the same as that of $n(\hat{\rho} - 1)$, where $\hat{\rho}$ is the estimator (10.1.3) and c is a constant defined in Theorem 10.1.2. The large sample distribution of $(\tilde{\theta}_2, \tilde{\theta}_3, \ldots, \tilde{\theta}_p)$ is the same as that of the regression coefficients in the regression of Z_t on $Z_{t-1}, Z_{t-2}, \ldots, Z_{t-p+1}$.

Theorem 10.1.2. Let Y_t be defined by (10.1.16). Suppose the e_t satisfy the assumptions of Theorem 10.1.1. Let $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_p)'$ be the vector of regression coefficients obtained by regressing Y_t on $Y_{t-1}, Z_{t-1}, \ldots, Z_{t-p+1}$, $t = p+1, p+2, \ldots, n$, and define $\mathbf{D}_n$ to be the diagonal matrix

$$\mathbf{D}_n = \operatorname{diag}\{n, n^{1/2}, \ldots, n^{1/2}\}.$$

Define $c = \Sigma_{i=0}^{\infty} w_i = (1 - \Sigma_{i=2}^{p} \theta_i)^{-1}$, where

$$Z_t = \sum_{i=0}^{\infty} w_i e_{t-i},$$

and the w_i satisfy the homogeneous difference equation with characteristic equation (10.1.18) and initial conditions $w_0 = 1$ and $w_i = 0$ for $i < 0$. Let

$$\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta} = \left[\left(c \sum_{t=2}^{n} W_{t-1}^2\right)^{-1} \sum_{t=2}^{n} W_{t-1} e_t,\ \tilde{\theta}_2 - \theta_2,\ \tilde{\theta}_3 - \theta_3, \ldots, \tilde{\theta}_p - \theta_p\right]',$$

where $W_t = \Sigma_{j=1}^{t} e_j$, and $(\tilde{\theta}_2, \tilde{\theta}_3, \ldots, \tilde{\theta}_p)'$ is the vector of regression coefficients obtained by regressing Z_t on $Z_{t-1}, Z_{t-2}, \ldots, Z_{t-p+1}$. Then

$$\operatorname*{plim}_{n\to\infty} \mathbf{D}_n(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}) = \mathbf{0}.$$

Furthermore, $n(\hat{\theta}_1 - 1)$ is independent of $n^{1/2}(\hat{\theta}_2 - \theta_2, \ldots, \hat{\theta}_p - \theta_p)$ in the limit.

Proof. Using the definition of Z_t as the weighted sum of past e_t, we can write

$$\begin{aligned} Y_t &= \sum_{j=1}^{t} \left(\sum_{i=0}^{\infty} w_i\right) e_j - \sum_{j=1}^{t} \sum_{i=j}^{\infty} w_i e_{t-j+1} + \sum_{j=0}^{\infty} e_{-j} \sum_{i=j+1}^{j+t} w_i \\ &= cW_t + Q_t + R_t, \end{aligned}$$

where $Q_t = \Sigma_{j=1}^{t} g_j e_{t-j+1}$, $g_j = -\Sigma_{i=j}^{\infty} w_i$, and $R_t = \Sigma_{j=0}^{\infty} e_{-j} \Sigma_{i=j+1}^{j+t} w_i$. Since the absolute value of the largest root of (10.1.18) is less than $\lambda < 1$, by Exercise 2.24, $|w_i|$ is less than $M_1 \lambda^i$ and $|g_j|$ is less than $M_2 \lambda^j$ for some finite M_1 and M_2. It follows that $E\{Q_t Q_{t+h}\}$ and $E\{Z_t Z_{t+h}\}$ are bounded by a multiple of $\lambda^{|h|}$. Also, the expectations of $W_t Q_j$, $Y_t Z_j$, $n^{-1} Y_t Y_j$, and R_t^2 are bounded. Using these results, it can be shown that

$$n^{-2}\left[\sum_{t=p+1}^{n} Y_{t-1}^2 - c^2 \sum_{t=p+1}^{n} W_{t-1}^2\right] = o_p(1), \qquad (10.1.21)$$

and

$$n^{-3/2}\left[\sum_{t=p+1}^{n} Y_{t-1} Z_{t-j}\right] = o_p(1), \qquad j = 1, 2, \ldots, p-1.$$

Let $\mathbf{A}_n$ be the matrix of sums of squares and cross products of Y_{t-1}, Z_{t-1}, $Z_{t-2}, \ldots, Z_{t-p+1}$, and let $\mathbf{B}_n$ be the matrix

$$\mathbf{B}_n = \begin{pmatrix} c^2 \sum_{t=p+1}^{n} W_{t-1}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_n \end{pmatrix},$$

where $\mathbf{C}_n$ is the $(p-1) \times (p-1)$ matrix of sums of squares and cross products of $Z_{t-1}, Z_{t-2}, \ldots, Z_{t-p+1}$. The lower right $(p-1) \times (p-1)$ submatrix of $\mathbf{A}_n$ is $\mathbf{C}_n$. Furthermore, the ijth element of $n^{-1}\mathbf{C}_n$ is converging in probability to $\gamma_Z(i-j)$. By the convergence results for $n^{-2} \sum W_{t-1}^2$, we have $\mathbf{D}_n^{-1}\mathbf{B}_n\mathbf{D}_n^{-1} = O_p(1)$ and $\mathbf{D}_n\mathbf{B}_n^{-1}\mathbf{D}_n = O_p(1)$. Therefore, using equation (10.1.21),

$$\mathbf{D}_n(\mathbf{A}_n^{-1} - \mathbf{B}_n^{-1})\mathbf{D}_n = o_p(1).$$

Similarly,

$$n^{-1}\left(\sum_{t=p+1}^{n} Y_{t-1} e_t - c \sum_{t=p+1}^{n} W_{t-1} e_t\right) = o_p(1),$$

$$\frac{n^{-1} \sum_{t=p+1}^{n} Y_{t-1} e_t}{n^{-2} \sum_{t=p+1}^{n} Y_{t-1}^2} = \frac{nc \sum_{t=p+1}^{n} W_{t-1} e_t}{c^2 \sum_{t=p+1}^{n} W_{t-1}^2} + o_p(1),$$

and the probability limit result is established.

From (10.1.13),

$$n^{-1} \sum_{t=p+1}^{n} W_{t-1} e_t = 0.5(n^{1/2}\bar{e})^2 - 0.5\sigma^2 + O_p(n^{-1/2}).$$

Let $\mathbf{U}_{.n,l} = (U_{1n}, U_{2n}, \ldots, U_{ln})$, $\mathbf{X}_t = (Z_{t-1}, Z_{t-2}, \ldots, Z_{t-p+1})$, and $\mathbf{A} = E\{\mathbf{X}_t'\mathbf{X}_t\}$, where U_{in} is defined in (10.1.12). Now

$$n^{-1/2}\left(\sum_{i=1}^{l} b_{1i}m_{nit} + b_2 + \sum_{j=1}^{p-1} b_{3j}Z_{t-j}\right)e_t\,,$$

where m_{nit} is defined in (10.1.10) and $b_{11},\ldots,b_{1l}, b_2, b_{31},\ldots,b_{3,p-1}$ are fixed constants, satisfies the assumptions of Theorem 5.3.4. Therefore,

$$\left[\mathbf{U}_{.n,l}, n^{-1/2}\sum_{t=1}^{n} e_t, \left(n^{-1/2}\mathbf{A}^{-1/2}\sum_{t=1}^{n}\mathbf{X}_t'e_t\right)'\right]' \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}\sigma^2)$$

for every fixed l. It follows, by Lemma 6.3.1, that

$$\left[G_n, T_n, \left(n^{-1/2}\mathbf{A}^{-1/2}\sum_{t=1}^{n}\mathbf{X}_t'e_t\right)'\right]$$

converges in distribution to $(G, T, \boldsymbol{\xi}')$, where (G_n, T_n) is defined in the proof of Theorem 10.1.1, $\boldsymbol{\xi}$ is a $N(\mathbf{0}, \mathbf{I}\sigma^2)$ random variable, and (G, T) is independent of $\boldsymbol{\xi}$. ▲

Corollary 10.1.2.1. Let the assumptions of Theorem 10.1.2 hold. Let

$$\mathbf{K}_n^2 = \operatorname{diag}\left\{\sum_{t=p+1}^{n} Y_{t-1}^2, \sum_{t=p+1}^{n} Z_{t-1}^2, \sum_{t=p+1}^{n} Z_{t-2}^2, \ldots, \sum_{t=p+1}^{n} Z_{t-p+1}^2\right\},$$

$$\tilde{\mathbf{K}}_n^2 = \operatorname{diag}\left\{c^2\sum_{t=p+1}^{n} W_{t-1}^2, \sum_{t=p+1}^{n} Z_{t-1}^2, \sum_{t=p+1}^{n} Z_{t-2}^2, \ldots, \sum_{t=p+1}^{n} Z_{t-p+1}^2\right\}.$$

Then

$$\operatorname*{plim}_{n\to\infty}\,[\mathbf{K}_n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) - \tilde{\mathbf{K}}_n(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})] = \mathbf{0}\,,$$

where W_t, c, $\hat{\boldsymbol{\theta}}$, and $\tilde{\boldsymbol{\theta}}$ are defined in Theorem 10.1.1.

Proof. By the proof of Theorem 10.1.2,

$$n^{-1}\left(\sum_{t=p+1}^{n} Y_{t-1}^2\right)^{1/2} = n^{-1}\left(c^2\sum_{t=p+1}^{n} W_{t-1}^2\right)^{1/2} + o_p(1)\,,$$

and $n^{-1}(\Sigma_{t=p+1}^{n} Y_{t-1}^2)^{1/2} = O_p(1)$. Because

$$\operatorname*{plim}_{n\to\infty} n^{-1}\sum_{t=p+1}^{n} Z_{t-1}^2 = \gamma_Z(0)$$

by Theorem 6.3.5, the result follows. ▲

On the basis of Corollary 10.1.2.1, Table 10.A.2 can be used to investigate the hypothesis that one of the roots in a pth order autoregressive process is unity. That is, the pivotal statistic for θ_1 computed by an ordinary least squares program,

$$\hat{\tau} = [\hat{V}\{\hat{\theta}_1\}]^{-1/2}(\hat{\theta}_1 - 1), \tag{10.1.22}$$

where $\hat{V}\{\hat{\theta}_1\}$ is the ordinary least squares estimator of the variance of $\hat{\theta}_1$, has the limiting distribution given in Corollary 10.1.1.2. Also, the regression pivotals can be used to test hypotheses about the coefficients of the stationary process. For example, the hypothesis that the original process is of order $p-1$ is equivalent to the hypothesis that $\alpha_p = 0$, which is equivalent to the hypothesis that $\theta_p = 0$. A test statistic for this hypothesis is the ordinary regression pivotal statistic

$$t_p = [\hat{V}\{\hat{\theta}_p\}]^{-1/2}\hat{\theta}_p ,$$

where $\hat{V}\{\hat{\theta}_p\}$ is the ordinary least squares estimator of the variance of $\hat{\theta}_p$. Under the assumptions of Theorem 10.1.2, t_p is asymptotically normal when $\theta_p = 0$.

We note that the coefficient for Y_{t-1} in the regression of Y_t on $Y_{t-1}, Z_{t-1}, \ldots, Z_{t-p+1}$ is not the largest root of the fitted autoregressive equation. To see this, consider the second order process. The fitted equation is

$$\hat{Y}_t = (\hat{m}_1 + \hat{m}_2)Y_{t-1} - \hat{m}_1\hat{m}_2 Y_{t-2} ,$$

where $(\hat{m}_1, \hat{m}_2)$ are the two roots of the estimated characteristic equation. It follows that the estimated equation in Y_{t-1} and Z_{t-1} is

$$\hat{Y}_t = [\hat{m}_1 + \hat{m}_2(1 - \hat{m}_1)]Y_{t-1} + \hat{m}_1\hat{m}_2(Y_{t-1} - Y_{t-2}) .$$

Also, the error in the coefficient of Y_{t-1} as an estimator of one is

$$\hat{\theta}_1 - 1 = (\hat{m}_1 - 1)(1 - \hat{m}_2) = (\hat{m}_1 - 1)(1 - m_2) + O_p(n^{-3/2}) .$$

Because $c = (1 - m_2)^{-1}$, the limiting distribution of $n(\hat{m}_1 - 1)$ is the same as the limiting distribution of $n(\hat{\rho} - 1)$ given in Table 10.A.1. We state the generalization of this property as a corollary.

Corollary 10.1.2.2. Assume that the model (10.1.16)–(10.1.17) holds, and express Y_t in the form (10.1.19), where one of the roots of

$$m^p + \sum_{i=1}^{p} \alpha_i m^{p-i} = 0$$

is equal to one and the other roots are less than one in absolute value. Let $\tilde{\boldsymbol{\alpha}}' = (\tilde{\alpha}_1, \tilde{\alpha}_2, \ldots, \tilde{\alpha}_p)$ be the least squares estimator of $\boldsymbol{\alpha}$ estimated subject to the restriction that $\tilde{m}_1$ is real, where $(\tilde{m}_1, \tilde{m}_2, \ldots, \tilde{m}_p)$ are the roots of the estimated characteristic equation with $\tilde{m}_1 \geqslant |\tilde{m}_2| \geqslant \cdots \geqslant |\tilde{m}_p|$. Then

$$n(\tilde{m}_1 - 1) \xrightarrow{\mathscr{L}} (2G)^{-1}(T^2 - 1),$$

where (G, T) is defined in Theorem 10.1.1.

Proof. By (10.1.16) and (10.1.17), we have

$$m^p + \sum_{i=1}^{p} \alpha_i m^{p-i} = (m-1)\left(m^{p-1} - \sum_{i=1}^{p-1} \theta_{i+1} m^{p-i-1}\right),$$

$\hat{\theta}_1 - 1 = O_p(n^{-1})$, and $\hat{\theta}_i - \theta_i = O_p(n^{-1/2})$ for $i = 2, \ldots, p$, where the $\hat{\theta}_i$, $i = 1, 2, \ldots, p$, are the regression coefficients of Theorem 10.1.2. Therefore, by Corollary 5.8.1, the roots of

$$m^p + \sum_{i=1}^{p} \hat{\alpha}_i m^{p-i} = 0,$$

where the relation between $\hat{\theta}_i$ and $\hat{\alpha}_i$ is defined by (10.1.20), converge in probability to the roots of the equation defined with $(\alpha_1, \ldots, \alpha_p)$. The estimated polynomial for the ordinary least squares estimator evaluated at $m = 1$ is

$$1 + \sum_{i=1}^{p} \hat{\alpha}_i = \prod_{i=1}^{p} (1 - \hat{m}_i) = (1 - \hat{m}_1) \prod_{i=2}^{p} (1 - \hat{m}_i).$$

Using the fact that $\hat{\theta}_1 = -\sum_{i=1}^{p} \hat{\alpha}_i$, and the limiting result for the roots,

$$\begin{aligned} \hat{\theta}_1 - 1 &= (\hat{m}_1 - 1)\left(\prod_{i=2}^{p} (1 - m_i) + o_p(1)\right) \\ &= (\hat{m}_1 - 1)\left(1 - \sum_{i=2}^{p} \theta_i + o_p(1)\right). \end{aligned}$$

This equality also demonstrates that given $\epsilon > 0$, there is some N_ϵ such that for $n > N_\epsilon$, the probability that $\hat{m}_1$, computed with the ordinary least squares estimator of $\hat{\alpha}$, is real is greater than $1 - \epsilon$. The conclusion follows because we demonstrated in the proof of Theorem 10.1.2 that

$$n(\hat{\theta}_1 - 1) \xrightarrow{\mathscr{L}} (2G)^{-1}\left(1 - \sum_{i=2}^{p} \theta_i\right)(T^2 - 1). \qquad \blacktriangle$$

For the model (10.1.1) with $|\rho| < 1$, the limiting behavior of the estimator of ρ is the same whether the mean is known or estimated. The result is no longer true when $\rho = 1$. Consider the estimator

$$\hat{\rho}_\mu = \frac{\sum_{t=2}^{n} (Y_t - \bar{y}_{(0)})(Y_{t-1} - \bar{y}_{(-1)})}{\sum_{t=2}^{n} (Y_{t-1} - \bar{y}_{(-1)})^2}, \tag{10.1.23}$$

where

$$[\bar{y}_{(0)}, \bar{y}_{(-1)}] = (n-1)^{-1} \sum_{t=2}^{n} (Y_t, Y_{t-1}) .$$

When $\rho = 1$, we have $(3^{-1}n)^{-1/2}\bar{y} \xrightarrow{\mathscr{L}} N(0, 1)$ and this random variable makes a contribution to the limiting distribution of $n(\hat{\rho}_\mu - 1)$. Also see the moment results of Exercise 10.4. The limit random variable associated with $\bar{y}$ is denoted by H in Theorem 10.1.3.

Theorem 10.1.3. Let the assumptions of Theorem 10.1.1 hold. Let $\hat{\rho}_\mu$ be defined by (10.1.23), and let

$$\hat{\tau}_\mu = [\hat{V}\{\hat{\rho}_\mu\}]^{-1/2}(\hat{\rho}_\mu - 1) , \qquad (10.1.24)$$

where

$$\hat{V}\{\hat{\rho}_\mu\} = \left[\sum_{t=2}^{n} (Y_t - \bar{y}_{(-1)})^2\right]^{-1} s^2$$

and

$$s^2 = (n-3)^{-1} \sum_{t=2}^{n} [Y_t - \bar{y}_{(0)} - \hat{\rho}_\mu(Y_{t-1} - \bar{y}_{(-1)})]^2 .$$

Then

$$n(\hat{\rho}_\mu - 1) \xrightarrow{\mathscr{L}} (G - H^2)^{-1}[0.5(T^2 - 1) - TH]$$

and

$$\hat{\tau}_\mu \xrightarrow{\mathscr{L}} (G - H^2)^{-1/2}[0.5(T^2 - 1) - TH] ,$$

where

$$H = \sum_{i=1}^{\infty} 2^{1/2}\gamma_i^2 U_i = \int_0^1 W(t)\, dt ,$$

$U_i \sim \mathrm{NI}(0, 1)$, $W(t)$ is the Wiener process, and G, T, and γ_i are defined in Theorem 10.1.1.

Proof. The method of proof is the same as that of Theorem 10.1.1. Let

$$H_n = n^{-3/2} \sum_{t=2}^{n} Y_{t-1} ,$$

and define $\mathbf{U}_n$ and $\mathbf{M}_n$ as in the proof of Theorem 10.1.1. Then

$$H_n = \mathbf{L}_n'\mathbf{e}_n = \mathbf{L}_n'\mathbf{M}_n^{-1}\mathbf{U}_n ,$$

where $\mathbf{L}_n' = n^{-3/2}(n-1, n-2, \ldots, 1)$. Let q_{in} be the ith element of $\mathbf{L}_n'\mathbf{M}_n^{-1}$. This element is the covariance between H_n and $\mathbf{U}_n$, and it can be shown that for fixed i

$$\lim_{n\to\infty} q_{in} = 2^{1/2}\gamma_i^2 .$$

The remainder of the proof follows that of Theorem 10.1.1. ▲

The second part of Table 10.A.1 contains empirical percentiles for the distribution of $n(\hat{\rho}_\mu - 1)$, and the second part of Table 10.A.2 contains the empirical percentiles of the corresponding studentized statistic given that $\rho = 1$. The values for $n = \infty$ are the percentiles of the limiting distribution of Theorem 10.1.3.

It can be demonstrated that $p\lim n(\hat{\rho}_\mu - \hat{\rho}) = 0$ when $\rho = -1$. That is, estimating the mean does not alter the limiting distribution when $\rho = -1$. Therefore, the first part of Tables 10.A.1 and 10.A.2 can be used to approximate the distributions of $\hat{\rho}_\mu$ and $\hat{\tau}_\mu$ when $\rho = -1$.

The distributions of Theorem 10.1.3 are obtained under the model (10.1.1) with $\rho = 1$. This point can be emphasized by writing an extended model as

$$Y_t = (1-\rho)\mu + \rho Y_{t-1} + e_t ,$$

where ρ and μ are unknown parameters. The null model for the test $\hat{\tau}_\mu$ *is* $\rho = 1$. The extended model then reduces to (10.1.1). The alternative model with $\rho \neq 1$ permits a nonzero value for $\theta_0 = (1-\rho)\mu$. Thus, the test based on $\hat{\tau}_\mu$ is invariant to the mean μ of the alternative model.

The limiting distribution of $n^{3/2}(\hat{\theta}_1 - 1)$, where $\hat{\theta}_1$ is the least squares estimator and $\theta_1 = 1$, for the model

$$Y_t = \theta_0 + \theta_1 Y_{t-1} + e_t$$

with $\theta_0 \neq 0$ is normal, and is discussed in Section 10.1.2.

It follows from Theorem 10.1.3 that the distribution of the ordinary least squares estimator of θ_0 for the model (10.1.1) with $\rho = 1$ is not normal. Let $\Delta Y_t = Y_t - Y_{t-1}$, and consider the equation

$$\Delta Y_t = \theta_0 + (\theta_1 - 1)Y_{t-1} + e_t , \qquad (10.1.25)$$

with the unknown θ_0 equal to zero. The estimated regression equation can be written

$$\Delta \hat{Y}_t = \Delta \bar{Y} + (\hat{\theta}_1 - 1)(Y_{t-1} - \bar{y}_{(-1)}) , \qquad (10.1.26)$$

where $\bar{y}_{(-1)}$ is defined in (10.1.23), $\Delta \bar{Y} = (n-1)^{-1} \sum_{t=2}^{n} \Delta Y_t$, and

$$\hat{\theta}_1 - 1 = \left(\sum_{t=2}^{n} (Y_{t-1} - \bar{y}_{(-1)})^2\right)^{-1} \sum_{t=2}^{n} (Y_{t-1} - \bar{y}_{(-1)})(\Delta Y_t - \Delta\bar{Y}) .$$

Under the null model, $\Delta Y_t = e_t$ and $\Delta\bar{Y} = \bar{e}$. It follows that

$$\hat{\theta}_0 = \bar{e} - (\hat{\theta}_1 - 1)\bar{y}_{(-1)}$$

and

$$n^{1/2}\hat{\theta}_0 \xrightarrow{\mathcal{L}} T - (G - H^2)^{-1}[0.5(T^2 - 1) - TH]H , \qquad (10.1.27)$$

where G, T, and H are defined in Theorem 10.1.1 and Theorem 10.1.3. Dickey (1976) and Dickey and Fuller (1981) have discussed the distribution of $n^{1/2}\hat{\theta}_0$. Under the normal distribution null model, $\Delta\bar{Y}$ of (10.1.26) is distributed as a normal random variable. Also, because $\hat{\theta}_1$ is a quadratic function of $(Y_1, \ldots, Y_n)$, $\Delta\bar{Y}$ is uncorrelated with $\hat{\theta}_1$ for normal e_t. Therefore, estimating the equation in the form (10.1.26) yields two uncorrelated estimators whose null distributions are easily interpreted.

The distributional results for the estimator of the parameters of the first order autoregressive model with intercept extend to the pth order process in much the same manner as the results for the model with no intercept. It follows from Theorem 10.1.4 that the regression pivotal calculated to test $\theta_1 = 1$ has the tabled distribution $\hat{\tau}_\mu$ when the fitted model has an intercept.

Theorem 10.1.4. Let Y_t be defined by (10.1.16). Suppose the assumptions of Theorem 10.1.2 hold. Let $\hat{\boldsymbol{\theta}} = (\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_p)'$ be the vector of regression coefficients obtained by regressing Y_t on $(1, Y_{t-1}, Z_{t-1}, \ldots, Z_{t-p+1})$, $t = p+1, p+2, \ldots, n$. Define $\mathbf{D}_n$ to be the diagonal matrix

$$\mathbf{D}_n = \text{diag}(n^{1/2}, n, n^{1/2}, n^{1/2}, \ldots, n^{1/2}) ,$$

and let $\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}$ be the vector such that $\tilde{\theta}_0 = \hat{\theta}_0$;

$$\tilde{\theta}_1 - \theta_1 = \left[c \sum_{t=2}^{n} (W_{t-1} - \bar{w}_{(-1)})^2\right]^{-1} \sum_{t=2}^{n} (W_{t-1} - \bar{w}_{(-1)})e_t ,$$

where W_t and c are defined in Theorem 10.1.2, and $\bar{w}_{(-1)} = (n-1)^{-1} \sum_{t=1}^{n-1} W_t$, and $(\tilde{\theta}_2, \tilde{\theta}_3, \ldots, \tilde{\theta}_p)$ is the vector of regression coefficients obtained by regressing Z_t on $(Z_{t-1}, Z_{t-2}, \ldots, Z_{t-p+1})$. Let τ_μ^* be the ordinary regression pivotal computed by dividing $\hat{\theta}_1 - 1$ by the regression estimated standard error of $\hat{\theta}_1$. Then

$$\operatorname*{plim}_{n\to\infty} \mathbf{D}_n(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}) = \mathbf{0} ,$$

$n(\hat{\theta}_1 - \theta_1)$ is independent of $n^{1/2}[\hat{\theta}_2 - \theta_2, \ldots, \hat{\theta}_p - \theta_p]$ in the limit, and the limiting distribution of τ_μ^* is the limiting distribution of $\hat{\tau}_\mu$ given in Theorem 10.1.3.

Proof. Omitted. ▲

There has been extensive research on models with unit roots. We give only a few references. Hasza and Fuller (1982) studied models with two unit roots. Dickey, Hasza, and Fuller (1984) gave results for seasonal models, and Tiao and Tsay (1983) gave results for models with complex roots. Said and Dickey (1984) show that fitting a high order autoregressive model is an appropriate way to test for a unit root in a model of unknown form. Chan and Wei (1988) and Pantula (1989) give results that cover models with a number of unit roots. Phillips (1986, 1987a–c, 1988) and his co-workers have given many results on both univariate and vector processes. Shin (1990) and Yap and Reinsel (1995a,b) give tests for a unit root in the autoregressive portion of an autoregressive moving average. Ahtola and Tiao (1984), Chan (1988), Chan and Wei (1987), and Phillips (1987b) have studied the behavior of estimators when ρ is close to one but not equal to one.

Example 10.1.1. In this example we analyze the one-year treasury bill interest rate for the period January 1960 through August 1979. This is one of three interest rate series studied by Stock and Watson (1988). The data are given in Table 10.B.1 of Appendix 10.B. We postulate the time series to be an autoregressive process, and we are interested in testing the hypothesis that the process has a unit root. In this example, we use the ordinary least squares method of fitting. If we fit a sixth order autoregressive process, the fitted equation is

$$\begin{aligned}\hat{Y}_t = \underset{(0.064)}{0.104} &+ \underset{(0.067)}{1.332}\, Y_{t-1} - \underset{(0.111)}{0.453}\, Y_{t-2} + \underset{(0.115)}{0.127}\, Y_{t-3} \\ &+ \underset{(0.115)}{0.043}\, Y_{t-4} - \underset{(0.111)}{0.079}\, Y_{t-5} + \underset{(0.067)}{0.014}\, Y_{t-6}\end{aligned}$$

and the regression residual mean square is 0.0833. There are 230 observations in the regression, and the residual mean square has 223 degrees of freedom. The coefficients for Y_{t-3}, Y_{t-4}, Y_{t-5}, and Y_{t-6} are small relative to the least squares standard errors. If we drop Y_{t-6} from the regression, the t-statistic for Y_{t-5} is -1.036. If we drop Y_{t-5} from the regression, the t-statistic for Y_{t-4} is -0.884, and if we drop Y_{t-4} from the regression, the t-statistic for Y_{t-3} is 1.427. By Theorem 8.2.1, the t-statistics are approximately normally distributed for a stationary process. By Theorem 10.1.4, the t-statistics are approximately normally distributed if the process contains a single unit root. Thus, the testing procedure is appropriate in either case. The F-test for the coefficients of Y_{t-4}, Y_{t-5}, and Y_{t-6}, computed as the sum of the squared t-statistics divided by 3, is 0.63. We compute the F-statistic in this way because different numbers of observations were used in the different regressions. The value of F is such that one easily accepts the hypothesis that the three coefficients are zero. The t-statistic for the coefficient of Y_{t-3} is such that one could either drop the coefficient or retain it in the analysis. We choose to retain the coefficient, and we proceed with the analysis using the third order model.

We are interested in testing for a unit root against the alternative of a stationary process, but we are willing to entertain the possibility of two unit roots. Dickey and Pantula (1987) suggest a sequential testing procedure in such situations. One begins with the greatest number of unit roots one is willing to consider. In our case, we begin by testing for a second unit root under the maintained hypothesis of a single unit root. We are willing to assume there is no intercept in the model for the differences. The estimated regression is

$$\Delta^2 \hat{Y}_t = \underset{(0.076)}{-\ 0.763}\, \Delta Y_{t-1} + \underset{(0.065)}{0.103}\, \Delta^2 Y_{t-1}\,,$$

where the numbers in parentheses are the regression estimated standard errors. Because the τ-statistic $\hat{\tau} = (0.076)^{-1}(-0.763) = -10.09$ is less than the 1% tabular value of -2.58 in the first part of Table 10.A.2, we reject the hypothesis of a second unit root.

The regression equation for the test of a unit root in the original time series is

$$\Delta \hat{Y}_t = \underset{(0.062)}{0.082} - \underset{(0.011)}{0.012}\, Y_{t-1} + \underset{(0.065)}{0.343}\, \Delta Y_{t-1} - \underset{(0.066)}{0.094}\, \Delta Y_{t-2}\,.$$

Because the $\hat{\tau}_\mu$ statistic of -1.09 for the coefficient of Y_{t-1} is greater than the 10% tabular value of -2.57 from the second part of Table 10.A.2, we would accept the hypothesis of a unit root at that level. The analysis of these data is extended in Example 10.1.2. ▲▲

10.1.2. Random Walk with Drift

In this section, we obtain the limiting distribution of the least squares estimator for (θ_0, θ_1) of the model

$$Y_t = \begin{cases} \theta_0 + \theta_1 Y_{t-1} + e_t\,, & t = 1, 2, \ldots, \\ Y_0\,, & t = 0\,, \end{cases} \tag{10.1.28}$$

under the assumption that $\theta_0 \neq 0$ and $\theta_1 = 1$. The model (10.1.28) with $\theta_0 \neq 0$ is sometimes called a random walk with *drift*, where θ_0 is the drift parameter. Under (10.1.28)

$$Y_t = Y_0 + \theta_0 t + W_t\,, \tag{10.1.29}$$

where $W_t = \sum_{j=1}^{t} e_j$. When $\theta_0 \neq 0$, the time trend will dominate the long-run behavior of Y_t of (10.1.29). This is because $W_t = O_p(t^{1/2})$ and hence W_t is small in probability relative to t. The least squares estimator of $\boldsymbol{\theta} = (\theta_0, \theta_1)'$ is $\hat{\boldsymbol{\theta}} = (\hat{\theta}_0, \hat{\theta}_1)'$, where

$$\begin{aligned}
\hat{\theta}_1 &= \left[\sum_{t=2}^{n} (Y_{t-1} - \bar{y}_{(-1)})^2\right]^{-1} \sum_{t=2}^{n} (Y_{t-1} - \bar{y}_{(-1)})Y_t\,, \\
\hat{\theta}_0 &= \bar{y}_{(0)} - \hat{\theta}_1 \bar{y}_{(-1)}\,, \qquad\qquad (10.1.30)\\
\bar{y}_{(i)} &= (n-1)^{-1} \sum_{t=2}^{n} Y_{t+i}\,, \qquad i = 0, -1\,.
\end{aligned}$$

We give the limiting distribution of the least squares estimator under the assumption that $\theta_0 \neq 0$ and $\theta_1 = 1$.

Theorem 10.1.5. Let the process Y_t satisfy (10.1.28) where $\{e_t\}$ satisfies the conditions of Theorem 10.1.1 or is a sequence of independent identically distributed $(0, \sigma^2)$ random variables. Assume $\theta_0 \neq 0$ and $\theta_1 = 1$. Let the least squares estimator be given by (10.1.30). Then

$$\begin{pmatrix} n^{1/2}(\hat{\theta}_0 - \theta_0) \\ n^{3/2}(\hat{\theta}_1 - \theta_1) \end{pmatrix} \xrightarrow{\mathscr{L}} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 2^{-1}\theta_0 \\ 2^{-1}\theta_0 & 3^{-1}\theta_0^2 \end{pmatrix}^{-1} \sigma^2\right].$$

Proof. We have

$$\begin{aligned}
\sum_{t=2}^{n} (Y_{t-1} - \bar{y}_{(-1)})^2 = \theta_0^2 \sum_{t=2}^{n} (t - \bar{t})^2 + 2\theta_0 \sum_{t=2}^{n} (t - \bar{t})(W_{t-1} - \bar{w}_{(-1)}) \\
+ \sum_{t=2}^{n} (W_{t-1} - \bar{w}_{(-1)})^2
\end{aligned}$$

and

$$\sum_{t=2}^{n} (Y_{t-1} - \bar{y}_{(-1)})e_t = \theta_0 \sum_{t=2}^{n} (t - \bar{t})e_t + \sum_{t=2}^{n} (W_{t-1} - \bar{w}_{(-1)})e_t\,,$$

where $\bar{t} = 0.5(n+1)$ and $\bar{w}_{(-1)} = (n-1)^{-1} \Sigma_{t=2}^{n} W_{t-1}$. By our previous results,

$$\sum_{t=2}^{n} (W_{t-1} - \bar{w}_{(-1)})[n^{-2}(W_{t-1} - \bar{w}_{(-1)}), n^{-5/2}(t - \bar{t}), n^{-1}e_t] = O_p(1)\,.$$

Therefore, using $\Sigma_{t=2}^{n}(t - \bar{t})^2 = (12)^{-1}n(n-1)(n-2)$,

$$n^{3/2}(\hat{\theta}_1 - 1) = 12n^{-3/2}\theta_0^{-1} \sum_{t=2}^{n} (t - \bar{t})e_t + O_p(n^{-1/2})\,.$$

Similarly,

$$\hat{\theta}_0 - \theta_0 = n^{-1} \sum_{t=1}^{n} e_t - 6n^{-2} \sum_{t=1}^{n} (t - \bar{t})e_t + O_p(n^{-1})$$

and

$$n^{1/2}(\hat{\theta}_0 - \theta_0) \xrightarrow{\mathscr{L}} N(0, 4\sigma^2) .$$

The joint limiting distribution follows by Theorem 6.3.4. ▲

The regression estimated equation for the model (10.1.28) can be written in the form

$$\Delta Y_t = \Delta \bar{Y} + (\hat{\theta}_1 - 1)(Y_{t-1} - \bar{y}_{(-1)}) , \tag{10.1.31}$$

introduced in (10.1.26). This form has the advantage that $\Delta \bar{Y}$ is asymptotically normally distributed with mean θ_0 for all θ_0 when $\theta_1 = 1$.

A time series of the form (10.1.28) with $\theta_0 \neq 0$ will tend, over time, to increase if $\theta_0 > 0$ and to decrease if $\theta_0 < 0$. A natural alternative model for a time series displaying such behavior is as the sum of a stationary time series and a time trend. Let

$$Y_t = \phi_0 + \phi_1 t + X_t , \tag{10.1.32}$$

where X_t is an autoregressive time series satisfying

$$X_t = \theta_1 X_{t-1} + \sum_{j=2}^{p} \theta_j (X_{t-j+1} - X_{t-j}) + e_t . \tag{10.1.33}$$

By substituting (10.1.33) into (10.1.32), we obtain

$$Y_t = \psi_0 + \psi_1 t + \theta_1 Y_{t-1} + \sum_{j=2}^{p} \theta_j (Y_{t-j+1} - Y_{t-j}) + e_t , \tag{10.1.34}$$

where $\psi_0 = \phi_0(1 - \theta_1) + \theta_1 \phi_1 - \phi_1 \sum_{j=2}^{p} \theta_j$ and $\psi_1 = \phi_1(1 - \theta_1)$. Thus, one way to investigate the hypothesis that Y_t contains a unit root is to fit the regression model (10.1.34) by ordinary least squares and test the hypothesis that $\theta_1 = 1$. The distribution of the test statistic differs from that obtained when time is not included in the regression.

Theorem 10.1.6. Let Y_t satisfy (10.1.34) where $\{e_t\}_{t=1}^{n}$ is such that

$$E\{(e_t, e_t^2) \mid \mathscr{A}_{t-1}\} = (0, \sigma^2) \qquad \text{a.s.,}$$
$$E\{|e_t|^{2+\delta} \mid \mathscr{A}_{t-1}\} < M < \infty \qquad \text{a.s.}$$

for some $\delta > 0$ and $t = 1, 2, \ldots$, where $\mathscr{A}_t$ is the sigma-field generated by $(e_1, e_2, \ldots, e_t)$. Assume $\theta_1^0 = 1$.

Let $(\hat{\psi}_0, \hat{\psi}_1, \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_p)$ be the ordinary least squares regression coefficients, and let

$$\hat{\tau}_\tau = [\hat{V}\{\hat{\theta}_1\}]^{-1/2}(\hat{\theta}_1 - 1)$$

be the ordinary regression pivotal for θ_1. Then

$$\hat{\tau}_\tau \xrightarrow{\mathscr{L}} 0.5(G - H^2 - 3K^2)^{-1/2}[(T - 2H)(T - 6K) - 1],$$

where

$$K = \sum_{i=1}^{\infty} 2^{1/2}(2\gamma_i^3 - \gamma_i^2)U_i = 2\int_0^1 tW(t)\,dt - H,$$

$W(t)$ is the Wiener process, and T, G, H, γ_i, and U_i are defined in Theorem 10.1.1 and Theorem 10.1.3. Also, the limiting distribution of

$$n^{1/2}[(\hat{\theta}_2 - \theta_2^0), (\hat{\theta}_3 - \theta_3^0), \ldots, (\hat{\theta}_p - \theta_p^0)]$$

is the same as the limiting distribution of the least squares estimator obtained by regressing ΔY_t on $(\Delta Y_{t-1}, \ldots, \Delta Y_{t-p+1})$.

Proof. Omitted. See Dickey and Fuller (1979). ▲

In (10.1.34), we expressed Y_t as a function of the vector $(1, t, Y_{t-1}, \Delta Y_{t-1}, \ldots, \Delta Y_{t-p+1})$ and computed the estimates by ordinary least squares. A second estimation procedure that has been heavily used in practice is to adjust the observations by subtracting the mean functions as estimated by ordinary least squares. The conclusions of Theorem 10.1.4 and of Theorem 10.1.6 also hold for the coefficients obtained in the regression of y_t on $(y_{t-1}, \Delta y_{t-1}, \ldots, \Delta y_{t-p})$, where y_t is the deviation of Y_t from the estimated mean function.

10.1.3. Alternative Estimators

In the stationary case, we presented a number of alternative estimators of the parameters of the autoregressive process. The maximum likelihood estimator associated with (8.1.9), the least squares estimator (8.2.6), and the weighted symmetric estimator associated with (8.2.14) all have the same limiting distribution when Y_t is a stationary autoregressive process. However, the limiting distributions differ if the process has a unit root.

We begin our discussion with the first order model and let

$$Y_t = \begin{cases} \rho Y_{t-1} + e_t, & t = 2, 3, \ldots, \\ 0, & t = 0, \end{cases} \tag{10.1.35}$$

where the e_t are independent $(0, \sigma^2)$ random variables. The simple symmetric estimator for ρ minimizes (8.2.14) with $w_t \equiv 0.5$ and is

$$\tilde{\rho}_s = \left[\tfrac{1}{2}(Y_1^2 + Y_n^2) + \sum_{t=2}^{n-1} Y_t^2\right]^{-1} \sum_{t=2}^{n} Y_{t-1}Y_t. \tag{10.1.36}$$

The error in the simple symmetric estimator is

$$\tilde{\rho}_s - \rho = \left[(Y_n^2 - Y_1^2) + 2\sum_{t=2}^{n} Y_{t-1}^2\right]^{-1}\left[2\sum_{t=2}^{n} Y_{t-1}e_t - \rho(Y_n^2 - Y_1^2)\right], \tag{10.1.37}$$

and if $\rho = 1$, the error reduces to

$$\tilde{\rho}_s - 1 = -\left[(Y_n^2 - Y_1^2) + 2\sum_{t=2}^{n} Y_{t-1}^2\right]^{-1}\sum_{t=2}^{n} e_t^2, \tag{10.1.38}$$

where we have used (10.1.13).

An estimator of σ^2 associated with (10.1.36) is $\tilde{\sigma}_s^2 = (n-2)^{-1}Q_s(\tilde{\rho}_s)$, where Q_s is (8.2.14) with $w_t \equiv 0.5$. Then

$$\tilde{\sigma}_s^2 = (n-2)^{-1}\left[\sum_{t=2}^{n}(Y_t - \tilde{\rho}_s Y_{t-1})^2 + 0.5(1-\tilde{\rho}_s^2)(Y_1^2 - Y_n^2)\right]. \tag{10.1.39}$$

If this estimator of variance is used to construct a pivotal statistic, the pivotal reduces to a function of $\tilde{\rho}_s$. Letting

$$\hat{\tau}_s = \left\{\tilde{\sigma}_s^2[0.5(Y_1^2 + Y_n^2) + \sum_{t=2}^{n-1} Y_t^2]^{-1}\right\}^{-1/2}(\tilde{\rho}_s - 1), \tag{10.1.40}$$

we have

$$\hat{\tau}_s = -(n-2)^{1/2}(1+\tilde{\rho}_s)^{-1/2}(1-\tilde{\rho}_s)^{1/2}. \tag{10.1.41}$$

The estimator constructed with the weights of (8.2.15) is called the weighted symmetric estimator and can be written as

$$\tilde{\rho}_w = \left(\sum_{t=2}^{n-1} Y_t^2 + n^{-1}\sum_{t=1}^{n} Y_t^2\right)^{-1}\sum_{t=2}^{n} Y_{t-1}Y_t. \tag{10.1.42}$$

The error in the weighted symmetric estimator when $\rho = 1$ is

$$\tilde{\rho}_w - 1 = \left(\sum_{t=2}^{n-1} Y_t^2 + n^{-1}\sum_{t=1}^{n} Y_t^2\right)^{-1}\left(\sum_{t=2}^{n} Y_{t-1}e_t + Y_1^2 - n^{-1}\sum_{t=1}^{n} Y_t^2\right). \tag{10.1.43}$$

The pivotal statistic for the weighted symmetric estimator is

$$\hat{\tau}_w = \left[\tilde{\sigma}_w^2\left(\sum_{t=2}^{n-1} Y_t^2 + n^{-1}\sum_{t=1}^{n} Y_t^2\right)^{-1}\right]^{-1/2}(\tilde{\rho}_w - 1), \tag{10.1.44}$$

where $\tilde{\sigma}_w^2$ is the estimator of σ^2 constructed by dividing $Q_w(\tilde{\rho}_w)$ by $n-2$. The limiting distributions of the symmetric estimators follow from (10.1.38) and (10.1.43).

Theorem 10.1.7. Let the assumptions of Theorem 10.1.1 hold, and let $\tilde{\rho}_s$ be defined by (10.1.36), $\hat{\tau}_s$ by (10.1.40), $\tilde{\rho}_w$ by (10.1.42), and $\hat{\tau}_w$ by (10.1.44). Then

$$
\begin{aligned}
n(\tilde{\rho}_s - 1) &\xrightarrow{\mathscr{L}} -(2G)^{-1}, \\
\hat{\tau}_s &\xrightarrow{\mathscr{L}} -0.5G^{-1/2}, \\
n(\tilde{\rho}_w - 1) &\xrightarrow{\mathscr{L}} 0.5G^{-1}(T^2 - 1) - 1, \\
\hat{\tau}_w &\xrightarrow{\mathscr{L}} 0.5G^{-1/2}(T^2 - 1) - G^{1/2},
\end{aligned}
$$

where G and T are defined in Theorem 10.1.1.

Proof. From (10.1.38) we have

$$
\begin{aligned}
n(\tilde{\rho}_s - 1) &= -0.5\left[(2n^2)^{-1}(Y_n^2 + Y_1^2 + 2\sum_{t=2}^{n-1} Y_t^2)\right]^{-1}\sigma^2 + o_p(1) \\
&= -0.5\left[n^{-2}\sum_{t=1}^{n} Y_t^2\right]^{-1}\sigma^2 + o_p(1),
\end{aligned}
$$

and the first result follows from Theorem 10.1.1. The limiting distribution for $\hat{\tau}_s$ is a consequence of (10.1.41).

Using (10.1.43), and the limiting distributions of $n^{-2}\Sigma_{t=1}^{n} Y_t^2$ and of $n^{-1}\Sigma_{t=2}^{n} Y_{t-1}e_t$ derived in Theorem 10.1.1, we obtain the results for $n(\tilde{\rho}_w - 1)$ and $\hat{\tau}_w$. ▲

The first part of Table 10.A.3 contains percentiles of the distribution of $\hat{\tau}_s$, and the first part of Table 10.A.4 contains the percentiles of the distribution of $\hat{\tau}_w$. The percentiles were generated using the Monte Carlo method. The percentiles for the limiting distributions were constructed by approximating the infinite sums that define (G, T, H, K) with finite sums. The raw Monte Carlo percentiles were smoothed using a function of n^{-1}. The standard errors of the estimates of Table 10.A.3 are generally less than 0.01, and those of Table 10.A.4 are generally less than 0.007.

The difference $\tilde{\rho}_s - 1$ is always negative, and this is reflected in the percentiles of Table 10.A.3. The percentiles of $n(\tilde{\rho}_s - 1)$ are defined in terms of those of Table 10.A.3 by equation (10.1.41). The distribution of $\tilde{\rho}_s$ has somewhat smaller tails than the distribution of the ordinary least squares estimator. For example, the distance between the 0.025 and 0.975 percentiles of the limiting distribution is 11.03 for $\tilde{\rho}_s$ and is 12.10 for $\hat{\rho}$. About 10% of the values of $\tilde{\rho}_w$ are greater than one, and the 90th percentile of $\hat{\tau}_w$ is close to zero.

The distributions of the symmetric estimators are a function of other variables included in the regression model, as is the distribution of the ordinary least squares estimator. Let $\hat{f}_{tk}$ be an estimated "mean function" for the time series Y_t. Examples of $\hat{f}_{tk}$ are the sample mean

$$\hat{f}_{yt1} = \bar{y}_n ,$$

the linear regression trend function

$$\hat{f}_{yt2} = b_0 + b_1 t , \tag{10.1.45}$$

where (b_0, b_1) is the vector of regression coefficients obtained in the ordinary least squares regression of Y_t on $(1, t)$, and the quadratic regression trend function

$$\hat{f}_{yt3} = b_0 + b_1 t + b_2 t^2 ,$$

where (b_0, b_1, b_2) is the vector of regression coefficients obtained in the ordinary least squares regression of Y_t on $(1, t, t^2)$.

The limiting distributions of the τ-statistics are given in Theorem 10.1.8. The limiting distributions of $n(\tilde{\rho}_{gj} - 1)$, $g = s$, w and $j = 1, 2$, are obtained by squaring the denominators in the expressions. For example, the limiting distribution of $n(\hat{\rho}_{s1} - 1)$ is the distribution of $-0.5(G - H^2)^{-1}$.

Theorem 10.1.8. Let the assumptions of Theorem 10.1.7 hold. Let $\tilde{\rho}_{sj}$ be the estimator obtained by replacing Y_t with $Y_t - \hat{f}_{ytj}$ in (10.1.36), and let $\hat{\tau}_{sj}$ be the pivotal obtained by replacing Y_t with $Y_t - \hat{f}_{ytj}$ in (10.1.40). Let $\tilde{\rho}_{wj}$ and $\hat{\tau}_{wj}$ be constructed in an analogous manner using (10.1.42) and (10.1.44), respectively. Then

$$\hat{\tau}_{s1} \xrightarrow{\mathcal{L}} -0.5(G - H^2)^{-1/2} ,$$

$$\hat{\tau}_{s2} \xrightarrow{\mathcal{L}} -0.5(G - H^2 - 3K^2)^{-1/2} ,$$

$$\hat{\tau}_{w1} \xrightarrow{\mathcal{L}} (G - H^2)^{-1/2}[0.5(T^2 - 1) - TH - G + 2H^2] ,$$

$$\hat{\tau}_{w2} \xrightarrow{\mathcal{L}} \frac{0.5[(T - 2H)(T - 6K) - 1] + (H - 3K)^2 - (G - H^2 - 3K^2)}{(G - H^2 - 3K^2)^{1/2}} ,$$

where G, T, H, and K are defined in Theorem 10.1.6.

Proof. The limiting distributions of the simple symmetric test statistics are obtained by replacing Y_t with $Y_t - \hat{f}_{ytj}$ in (10.1.40), where $\tilde{\rho}_s - 1$ is given in (10.1.38), and the limiting distributions of the weighted symmetric test statistics are obtained by replacing Y_t with $Y_t - \hat{f}_{ytj}$ in (10.1.44), where $\tilde{\rho}_w - 1$ is given in

(10.1.43). We have

$$n^{-1}[(Y_1-\bar{y}_n)^2,(Y_1-\hat{f}_{yt2})^2]\xrightarrow{\mathscr{L}}[H^2,(H-3K)^2],$$

$$\left[\bar{e}_n\bar{y}_n, n^{-1}\sum_{t=2}^{n}\hat{f}_{y,t-1,2}\hat{f}_{yt2}\right]\xrightarrow{\mathscr{L}}[HT,0.5\{(T-2H)(T-6K)-1\}],$$

$$n^{-2}\left[\sum_{t=2}^{n}(Y_t-\bar{y}_n)^2,\sum_{t=1}^{n}(Y_t-\hat{f}_{yt2})^2\right]\xrightarrow{\mathscr{L}}[G-H^2,G-H^2-3K^2].$$

The limiting distributions of the test statistics follow from the joint limiting distribution of the components. ▲

The distributions of the pivotal statistics for the mean adjusted case are given in the second parts of Tables 10.A.3 and 10.A.4. The distributions for the linear trend adjusted case are given in the third parts of the tables. In addition, the distribution for the simple symmetric pivotal is given for the quadratic trend adjusted case.

The distributions of the symmetric estimators generalize to higher order processes in the same manner as the ordinary least squares estimators. Table 10.1.1 contains a data arrangement that can be used for the weighted regression estimation of the autoregressive model written as

$$Y_t=\theta_1Y_{t-1}+\theta_2Z_{t-1}+\theta_3Z_{t-2}+\cdots+\theta_pZ_{t-p+1}+e_t, \qquad (10.1.46)$$

where $Z_t=Y_t-Y_{t-1}$. If the model contains a mean function, the procedure is adjusted as described in Section 8.2.2. The hypothesis of a unit root is tested by testing the hypothesis that $\theta_1=1$.

Table 10.1.1. Data Arrangement for Regression Estimation of Autoregressive Parameters by the Weighted Symmetric Procedure

Weight	Dependent Variable	Parameter			
		θ_1	θ_2	$\cdots$	θ_p
w_{p+1}	Y_{p+1}	Y_p	Y_p-Y_{p-1}	$\cdots$	Y_2-Y_1
w_{p+2}	Y_{p+2}	Y_{p+1}	$Y_{p+1}-Y_p$	$\cdots$	Y_3-Y_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
w_{n-p}	Y_n	Y_{n-1}	$Y_{n-1}-Y_{n-2}$	$\cdots$	$Y_{n-p+1}-Y_{n-p}$
$1-w_{n-p-1}$	Y_{n-p}	Y_{n-p+1}	$Y_{n-p+1}-Y_{n-p+2}$	$\cdots$	$Y_{n-1}-Y_n$
$1-w_{n-p}$	Y_{n-p-1}	Y_{n-p}	$Y_{n-p}-Y_{n-p+1}$	$\cdots$	$Y_{n-2}-Y_{n-1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
$1-w_2$	Y_1	Y_2	Y_2-Y_3	$\cdots$	Y_p-Y_{p+1}

Theorem 10.1.9. Let the assumptions of Theorem 10.1.2 hold. Let $w_{ts} \equiv 0.5$, and let w_{tw} be defined in (8.2.15). Let $\hat{\boldsymbol{\theta}}_{gj}$ be the value of $\boldsymbol{\theta}$ that minimizes

$$\sum_{t=p+1}^{n} w_{tg}(y_{tj} - \theta_1 y_{t-1,j} - \theta_2 z_{t-1,j} - \cdots - \theta_p z_{t-p+1,j})^2$$
$$+ \sum_{t=1}^{n-p} (1 - w_{t+1,g})(y_{tj} - \theta_1 y_{t+1,j} + \theta_2 z_{t+2,j} + \cdots + \theta_p z_{t+p,j})^2 , \quad (10.1.47)$$

where $y_{tj} = Y_t - \hat{f}_{ytj}$, $z_{tj} = y_{tj} - y_{t-1,j}$, $\hat{f}_{yt1} = \bar{y}$, and $\hat{f}_{yt2}$ is the least squares estimated linear trend function. Let $\hat{\tau}_{gj}$, $g = s, w$, $j = 1, 2$, be the pivotal for $\theta_1 = 1$ associated with (10.1.47). Then the limiting distributions of the $\hat{\tau}_{gj}$ are those given in Theorem 10.1.8.

Proof. Omitted. ▲

The estimator of ρ for the autoregressive process (10.1.35) constructed by maximizing the normal stationary likelihood has a different limiting distribution than either that of ordinary least squares or those of the symmetric estimators when $\rho = 1$. The limiting distribution of that estimator, which we call the maximum likelihood estimator, has been given by Gonzalez-Farias (1992).

Theorem 10.1.10. Let the assumptions of Theorem 10.1.7 hold. Let $\hat{\rho}_m$ be the value of ρ that maximizes

$$L(Y; \rho, \sigma^2) = -0.5n \log 2\pi - 0.5n \log \sigma^2 + 0.5 \log(1 - \rho^2)$$
$$- (2\sigma^2)^{-1}\left[(1 - \rho^2)Y_1^2 + \sum_{t=2}^{n} (Y_t - \rho Y_{t-1})^2\right]. \quad (10.1.48)$$

Then

$$n(\hat{\rho}_m - 1) \xrightarrow{\mathcal{L}} 0.25G^{-1}\{(T^2 - 1) - [(T^2 - 1)^2 + 8G]^{1/2}\},$$

where G and T are defined in Theorem 10.1.1.

Proof. From (8.1.10), the cubic equation defining $(\rho - 1) = \delta$ can be written

$$f_n(\delta) = \delta^3 + (3 + c_1)\delta^2 + (3 + 2c_1 + c_2)\delta + (1 + c_1 + c_2 + c_3) = 0, \quad (10.1.49)$$

where $c_1 = -\tilde{\rho}$, $c_3 = (n-2)^{-1}n\tilde{\rho}$,

$$\tilde{\rho} = \left[(n-2)^{-1}(n-1)\sum_{t=2}^{n-1} Y_t^2\right]^{-1} \sum_{t=2}^{n} Y_{t-1}Y_t ,$$
$$c_2 = -(n-1)^{-1}(n+1) - \hat{\eta} ,$$

and

$$\hat{\eta} = (n-1)^{-1}\left(\sum_{t=2}^{n-1} Y_{t-1}^2\right)^{-1}(Y_1^2 + Y_n^2).$$

We observe that

$$\begin{aligned}
3 + c_1 &= 3 - \tilde{\rho} = 2 + O_p(n^{-1}), \\
3 + 2c_1 + c_2 &= -2(\tilde{\rho} - 1 + n^{-1}) + O_p(n^{-2}), \\
1 + c_1 + c_2 + c_3 &= 2n^{-1}(\tilde{\rho} - 1 + n^{-1}) - \hat{\eta} + O_p(n^{-3}).
\end{aligned}$$

It follows that the roots of (10.1.49) converge in probability to 0, 0, and -2. Hence, the root in $(-2, 0)$ is converging in probability to zero. Also,

$$\begin{aligned}
n^2\hat{\eta} &\xrightarrow{\mathscr{L}} G^{-1}T^2, \\
n(\tilde{\rho} - 1) &\xrightarrow{\mathscr{L}} G^{-1}[0.5(T^2 - 1) - G], \\
n^2(1 + c_1 + c_2 + c_3) &\xrightarrow{\mathscr{L}} -G^{-1}.
\end{aligned}$$

Let

$$g_n(\delta) = a_{0n}\delta^2 + a_{1n}\delta + a_{2n},$$

where $a_{0n} = 3 + c_1$, $a_{1n} = 3 + 2c_1 + c_2$, and $a_{2n} = 1 + c_1 + c_2 + c_3$. Because $n^2 a_{2n} \xrightarrow{\mathscr{L}} -G^{-1}$, the roots of $g_n(\delta) = 0$ are real and of opposite sign with probability approaching one as n increases. Letting $\tilde{\delta}_2$ denote the negative root of $g_n(\delta) = 0$, we have

$$\begin{aligned}
n\tilde{\delta}_2 &= n(2a_{0n})^{-1}[-a_{1n} - (a_{1n}^2 - 4a_{0n}a_{2n})^{1/2}] \\
&= 0.5n\{\tilde{\rho} - 1 + n^{-1} - [(\tilde{\rho} - 1 - n^{-1})^2 + 2\hat{\eta} - 4n^{-2}]^{1/2}\} + O_p(n^{-1}) \\
&\xrightarrow{\mathscr{L}} 0.25G^{-1}\{(T^2 - 1) - [(T^2 - 1)^2 + 8G]^{1/2}\}. \qquad (10.1.50)
\end{aligned}$$

Also,

$$\begin{aligned}
0 = f_n(\hat{\delta}_m) &= f_n(\tilde{\delta}_2) + f_n'(\tilde{\delta}_2)(\hat{\delta}_m - \tilde{\delta}_2) + f_n''(\delta^*)(\hat{\delta}_m - \tilde{\delta}_2)^2 \\
&= \tilde{\delta}_2^3 + (3\tilde{\delta}_2^2 + 2a_{0,n}\tilde{\delta}_2 + a_{1,n})(\hat{\delta}_m - \tilde{\delta}_2) + (6\delta^* + 2a_{0,n})(\hat{\delta}_m - \tilde{\delta}_2)^2,
\end{aligned}$$

where $\hat{\delta}_m$ is the unique root of $f_n(\delta) = 0$ in $(-2, 0)$, $f_n'(\cdot)$ is the derivative of $f_n(\cdot)$, $f_n''(\cdot)$ is the second derivative of $f_n(\cdot)$, and δ^* is between $\hat{\delta}_m$ and $\tilde{\delta}_2$. Because

$$n^2 f_n(\tilde{\delta}_2) = n^2\tilde{\delta}_2^3 < 0$$

and

$$n^2 f_n(0) = n^2 a_{2,n} < 0$$

for n large, $\hat{\delta}_m < \tilde{\delta}_2$ for n large. It follows that

$$n(\hat{\delta}_m - \tilde{\delta}_2) = \frac{-nb + [n^2 b^2 - 4n^2 \tilde{\delta}_2^3 (2a_{0,n} + 6\delta^*)]^{1/2}}{2(2a_{0,n} + 6\delta^*)} \xrightarrow{\mathscr{L}} 0,$$

where $b = 2a_{0,n}\tilde{\delta}_2 + a_{1,n} + 3\tilde{\delta}_2^2$. Therefore, the limiting distribution of $n\hat{\delta}_m$ is the same as the limiting distribution of $n\tilde{\delta}_2$. ▲

From (10.1.50), we see that the value of $\hat{\rho}_m - 1$ is close to the value of $\tilde{\rho}_w - 1$ unless $\tilde{\rho}_w - 1$ is positive, or unless $\tilde{\rho}_w - 1$ is negative and $2(n^2\hat{\eta} - 2)$ is large relative to $\tilde{\rho}_w - 1$. The limiting distribution of $2(n^2\hat{\eta} - 2)$ is the distribution of $2G^{-1}(T^2 - 2G)$, where $E\{T^2 - 2G\} = 0$. Empirical studies show that the 5% and 10% points of the distribution of $n(\hat{\rho}_m - 1)$ are very similar to those of the distribution of $n(\tilde{\rho}_w - 1)$.

If μ is unknown and the estimator of (μ, ρ, σ^2) is obtained by maximizing the likelihood (8.1.5), the limiting distribution of the estimator of ρ differs from that of Theorem 10.1.10. Let $\tilde{\rho}_{m1}$ denote the estimator of ρ that maximizes the likelihood (8.1.5). If Y_t satisfies the assumptions of Theorem 10.1.7, Gonzalez-Farias (1992) has shown that

$$n(\tilde{\rho}_{m1} - 1) \xrightarrow{\mathscr{L}} \zeta,$$

where ζ is the negative root of

$$a_4\zeta^4 + a_3\zeta^3 + a_2\zeta^2 + a_1\zeta + a_0 = 0,$$

$$[a_0, a_3, a_4] = [-4, -T^2 + 1 - 2TH + H^2 - 8(G - H^2), 2(G - H^2)],$$
$$a_1 = -4T^2 + 8 + 8TH - 8H^2 + 2(T - 2H)^2,$$
$$a_2 = 8(G - H^2) + 4T - 5 - 8TH + H^2,$$

and G, H, and T are defined in Theorem 10.1.6. Gonzalez-Farias (1992) also showed empirically that the limiting distribution of the estimator of ρ that maximizes (8.1.5) is nearly indistinguishable from the limiting distribution of the estimator that maximizes the known mean likelihood (10.1.48) where μ is replaced with $\bar{y}$.

Corollary 10.1.10. Let the assumptions of Theorem 10.1.7 hold. Let $\hat{\rho}_{m1}$ be the value of ρ that maximizes (10.1.48) with $y_{t1} = Y_t - \bar{y} = Y_t - \hat{f}_{yt1}$ replacing Y_t, and let $\hat{\rho}_{m2}$ be the value of ρ that maximizes (10.1.48) with $y_{t2} = Y_t - \hat{f}_{yt2}$ replacing Y_t, where $\hat{f}_{yt2}$ is the ordinary least squares estimator of the linear trend function. Then

$$n(\hat{\rho}_{mj} - 1) \xrightarrow{\mathscr{L}} 0.5\{\zeta_j + 1 - [(\zeta_j - 1)^2 + 2\eta_j - 4]^{1/2}\},$$

where ζ_j is the limiting distribution of $n(\bar{\rho}_{wj} - 1)$ obtained by multiplying the distribution of $\hat{\tau}_{wj}$ of Theorem 10.1.8 by $(G - H^2)^{-1/2}$ for $j = 1$ and by $(G - H^2 - 3K^2)^{-1/2}$ for $j = 2$,

$$\eta_1 = (G - H^2)^{-1}[H^2 + (T - H)^2],$$

and

$$\eta_2 = (G - H^2 - 3K^2)^{-1}[(H - 3K)^2 + (T - H - 3K)^2].$$

Proof. The results follow from (10.1.50). ▲

A pivotal statistic for the test of $\rho = 1$ can be constructed by dividing $\hat{\rho}_{mj} - 1$ by the estimated standard error of $\hat{\rho}_{mj}$. For higher order processes, the test is constructed by testing the hypothesis that the sum of the autoregressive coefficients is one.

There are several methods available for computing the estimated covariance matrix of the estimated coefficients. One procedure is to write the concentrated log likelihood of (8.4.5) in the form

$$\sum_{t=1}^{n} g_t^2(\mathbf{Y}; \boldsymbol{\theta}), \tag{10.1.51}$$

where $g_t(\mathbf{Y}; \boldsymbol{\theta}) = c(\boldsymbol{\theta})z_t(\mathbf{Y}; \boldsymbol{\theta})$, $z_t(\mathbf{Y}; \boldsymbol{\theta})$ are uncorrelated $(0, \sigma^2)$ random variables when the function is evaluated at the true $\boldsymbol{\theta}$, and $c(\boldsymbol{\theta})$ is a constant depending on $\boldsymbol{\theta}$. Then the maximum likelihood estimators are obtained by minimizing (10.1.51), and an estimator of the covariance matrix of the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_m$ is

$$\hat{\mathbf{V}}\{\hat{\boldsymbol{\theta}}_m\} = \left[\sum_{t=1}^{n} \mathbf{h}_t'(\mathbf{Y}; \hat{\boldsymbol{\theta}}_m)\mathbf{h}_t(\mathbf{Y}; \hat{\boldsymbol{\theta}}_m)\right]^{-1} \hat{\sigma}_m^2, \tag{10.1.52}$$

where $\mathbf{h}_t'(\mathbf{Y}; \hat{\boldsymbol{\theta}}_m)$ is the vector of derivatives of $g_t(\mathbf{Y}; \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ evaluated at the maximum likelihood estimator,

$$\hat{\sigma}_m^2 = (n - r)^{-1} \sum_{t=1}^{n} z_t^2(\mathbf{Y}; \boldsymbol{\theta}) \tag{10.1.53}$$

is the maximum likelihood estimator of σ^2 adjusted for degrees of freedom, and r is the number of parameters estimated.

Table 10.A.5 contains the percentiles of the test statistics based on the maximum likelihood estimator and the estimated covariance matrix (10.1.52). This table was constructed by the same procedures used to construct Table 10.A.4.

We have presented a number of test statistics for the unit root problem. The ordinary least squares estimators are easy to compute and can be computed with an ordinary regression program. They may be used in exploratory procedures even when other programs are available. The simple symmetric estimator is next in ease of computation and can also be computed with an ordinary least squares regression program.

If the objective is to test the hypothesis of a unit root against the alternative of a stationary process with unknown mean, the maximum likelihood and weighted symmetric procedures are recommended. Monte Carlo studies have demonstrated that these two procedures have nearly identical power against the stationary unknown mean alternative and that they are much more powerful than ordinary least squares in that case. The power of the test based on the simple symmetric estimator falls between that of ordinary least squares and that of maximum likelihood, being closer to that of maximum likelihood. See for example, Pantula, Gonzalez-Farias, and Fuller (1994) and Park and Fuller (1995).

If a test for a unit root against the alternative of a model with known mean is desired, the ordinary least squares procedure is recommended. See Park and Fuller (1995).

Example 10.1.2. We continue the analysis of the one-year treasury bill interest rate begun in Example 10.1.1. The simple symmetric estimator of the third order process is

$$\underset{}{\Delta \hat{Y}_t} = \underset{(0.062)}{0.106} - \underset{(0.011)}{0.020}\, Y_{t-1} + \underset{(0.065)}{0.354}\, \Delta Y_{t-1} - \underset{(0.066)}{0.089}\, \Delta Y_{t-2} .$$

The regression equation was computed using an ordinary regression program and the data arrangement of Table 10.1.1 with an intercept included in the regression. The estimated error variance is 0.083. The value of the statistic for a test of a unit root is -1.76, which is greater than the 0.10 tabular value of -2.34 given in Table 10.A.3.

If we use the weights of (8.2.15), we obtain

$$\Delta \hat{Y}_t = \underset{(0.062)}{0.090} - \underset{(0.011)}{0.015}\, Y_{t-1} + \underset{(0.065)}{0.360}\, \Delta Y_{t-1} - \underset{(0.066)}{0.092}\, \Delta Y_{t-2}$$

as the weighted symmetric least squares estimated equation. The test statistic for a unit root is -1.36, which is greater than the 0.10 tabular value of -2.23 given in Table 10.A.4.

The maximum likelihood estimates for the third order process are

$$(-\hat{\alpha}_1, -\hat{\alpha}_2, -\hat{\alpha}_3) = (\underset{(0.0650)}{1.3436}, \ \underset{(0.1053)}{-0.4510}, \ \underset{(0.0656)}{0.0915}),$$

where we used procedure ARIMA of SAS/ETS® to compute the maximum likelihood estimator. This program uses the Marquardt algorithm and equation (10.1.52) to compute the estimated covariance matrix of the coefficients. The sum of the estimated autoregressive coefficients is 0.9841, and the estimated standard error of the sum of the coefficients is 0.0119. Therefore, the maximum likelihood pivotal for testing the hypothesis of a unit root is -1.34. The computed value is greater than the 0.10 tabular value of -2.32 given in the second part of Table 10.A.5.

In this example, all procedures lead to essentially the same conclusions. As stated in the text, the weighted symmetric and maximum likelihood procedures are the recommended procedures when the mean is estimated. In this example, and in the majority of practical situations, the weighted symmetric estimator and the maximum likelihood estimator will both reject the hypothesis of a unit root or both accept the hypothesis when the test is against the stationary alternative. ▲▲

From the tabled distribution of $n(\hat{\rho}-1)$ of Table 10.A.1, it is clear that the least squares estimator of the unit root parameter of an autoregressive process is biased. Also, the estimator of the autoregressive parameter is biased in the stationary case. See Section 8.2. In the unit root case, the bias is order n^{-1}, the same order as the standard deviation of the estimator.

There are several possible situations associated with the use of data to investigate an autoregressive process. One may be willing, on the basis of subject matter knowledge, to specify that the process is stationary. Then the use of an estimation procedure, such as maximum likelihood based on the stationary likelihood, that restricts all roots of the process to be less than one in absolute value, is appropriate. In a second situation, one may have strong theoretical reasons for believing a unit root is present. If so, one may choose to test for a unit root and, if the test is consistent with the hypothesis of a unit root, estimate the process subject to the restriction of a unit root. In the third situation, a unit root is a distinct possibility, but not to the extent that it is a model to be maintained unless rejected by the data. In this third case, it is natural to adopt an estimator that is less biased in the vicinity of the unit root than are the ordinary estimators.

Assume that the model is a pth order autoregressive process. Assume we wish to restrict the estimator of the largest root to the interval $(-1, 1]$. Under the assumption that only a single root of one is possible and that there is no drift if a root is equal to one, an estimation procedure is the following.

1. Estimate the model

$$Y_t = \theta_0 + \theta_1 Y_{t-1} + \sum_{j=2}^{p} \theta_j \,\Delta Y_{t-j+1} + e_t \tag{10.1.54}$$

by weighted symmetric least squares, and compute the statistic $\hat{\tau}_{w1}$ for the test that $\theta_1 = 1$. This can be done by fitting the weighted least squares regression of y_t on $(y_{t-1}, \Delta Y_{t-1}, \ldots, \Delta Y_{t-1+p})$, where $y_t = Y_t - \bar{y}_n$.

2. Construct a new estimator of θ_1

$$\tilde{\theta}_1 = \hat{\theta}_1 + c(\hat{\tau}_{w1})[\hat{V}\{\hat{\theta}_1\}]^{0.5}, \tag{10.1.55}$$

where $\hat{\theta}_1$ is the estimator from step 1,

$$c(\hat{\tau}_{w1}) = \begin{cases} -\hat{\tau}_{w1} & \text{if} \quad \hat{\tau}_{w1} \geq -1.2\,, \\ 0.035672(\hat{\tau}_{w1} + 7.0)^2 & \text{if} \quad -7.00 < \hat{\tau}_{w1} \leq -1.2\,, \\ 0 & \text{if} \quad \hat{\tau}_{w1} \leq -7.00\,, \end{cases}$$

and $\hat{V}\{\hat{\theta}_1\}$ is the estimated variance of the weighted least squares estimator.

3. Estimate $(\theta_2, \ldots, \theta_p)$ by the regression of $y_t - \tilde{\theta}_1 y_{t-1}$ on ΔY_{t-j}, $j = 1, 2, \ldots, p-1$. Denote the estimators by $(\tilde{\theta}_2, \ldots, \tilde{\theta}_p)$.

In the presence of a unit root ($\theta_1 = 1$), the outlined estimation procedure is approximately median unbiased for θ_1 because the median of $\hat{\tau}_{w1}$ is about -1.20 when $\theta_1 = 1$. Of course, if θ_1 is restricted to the interval $(-1, 1]$, it is not possible to construct a reasonable estimator of θ_1 that is mean unbiased for all $\theta_1 \in (-1, 1]$. The function $c(\hat{\tau}_{w1})$ of (10.1.55) was chosen to be a smooth function of $\hat{\tau}_{w1}$ with value of 1.20 at $\hat{\tau}_{w1} = -1.20$. The modified estimator differs from the weighted least squares estimator if $\hat{\tau}_{w1} > -7.00$. This set of $\hat{\tau}_{w1}$ corresponds, approximately, to $\hat{\theta}_1 > (n+49)^{-1}(n-49)$. The weighted least squares estimator $\hat{\theta}_1$ is biased for most values of θ_1, but the bias is small relative to the standard error for θ_1 considerably less than one and n large.

The empirical properties of the weighted symmetric estimator and the estimator (10.1.55) are compared for the first order process in Table 10.1.2. The true model is

$$Y_t = \rho Y_{t-1} + e_t\,,$$

where $e_t \sim N(0, 1)$. The generated process is stationary if $\rho < 1$, and $Y_0 = 0$ if $\rho = 1$. The weighted least squares estimator, denoted by $\hat{\rho}_w$, is defined by (10.1.42) with Y_t replaced by $Y_t - \bar{y}$, provided the estimator is less than one, and is equal to one otherwise. The entries in the table are based on 3000 samples.

The estimator $\tilde{\rho}$ defined by (10.1.55) has a uniformly smaller median bias than the weighted least squares estimator $\hat{\rho}_w$. The mean bias is also uniformly smaller. However, the empirical variance of $\hat{\rho}_w$ is smaller than that of $\tilde{\rho}$ for ρ differing from one by more than $8n^{-1}$, and the mean square error of $\hat{\rho}_w$ is smaller than that of $\tilde{\rho}$ for ρ differing from one by more than four standard errors of the estimator. In those situations, the mean square error of $\tilde{\rho}$ exceeds that of $\hat{\rho}_w$ by modest percentages. The largest percentage difference in the table is about 8%. On the other hand, for all three values of n, the modified estimator has a mean square error less than one half that of $\hat{\rho}_w$ for $\rho = 1$. For a fixed ρ, such as 0.75, the mean square error multiplied by n declines as n increases for both estimators, primarily because of the decrease in bias.

The last two columns of Table 10.1.2 contain the empirical 2.5 percentiles of the statistics

$$\tilde{t}_1 = [\hat{V}(\hat{\rho}_w)]^{-1/2}(\tilde{\rho} - \rho)$$

Table 10.1.2. Empirical Properties of Estimator (10.1.55) and Weighted Least Squares Estimator for First Order Process

	Median		nMSE		Percentiles	
ρ	$\hat{\rho}_w$	$\tilde{\rho}$	$\hat{\rho}_w$	$\tilde{\rho}$	$\hat{t}_{1,025}$	$\tilde{t}_{1,025}$
			$n = 50$			
0.15	0.122	0.126	1.046	1.127	−2.25	−2.25
0.50	0.456	0.491	0.998	1.064	−2.36	−2.31
0.75	0.695	0.757	0.857	0.748	−2.52	−2.28
0.85	0.791	0.860	0.782	0.567	−2.58	−2.22
0.90	0.838	0.909	0.747	0.467	−2.64	−2.20
0.95	0.884	0.955	0.720	0.357	−2.71	−2.19
0.98	0.912	0.981	0.710	0.309	−2.82	−2.25
1.00	0.933	1.000	0.728	0.311	−2.83	−2.22
			$n = 100$			
0.50	0.479	0.483	0.836	0.887	−2.13	−2.13
0.75	0.724	0.747	0.630	0.624	−2.27	−2.16
0.85	0.822	0.855	0.521	0.452	−2.38	−2.13
0.90	0.870	0.905	0.459	0.352	−2.45	−2.11
0.95	0.919	0.956	0.395	0.241	−2.57	−2.10
0.98	0.948	0.984	0.363	0.170	−2.71	−2.15
0.99	0.957	0.992	0.360	0.156	−2.75	−2.16
1.00	0.965	1.000	0.379	0.159	−2.80	−2.17
			$n = 200$			
0.75	0.738	0.742	0.562	0.590	−2.29	−2.28
0.85	0.836	0.847	0.421	0.421	−2.33	−2.25
0.90	0.886	0.901	0.342	0.313	−2.40	−2.20
0.95	0.935	0.953	0.262	0.199	−2.52	−2.18
0.98	0.964	0.983	0.216	0.125	−2.69	−2.21
0.99	0.974	0.992	0.204	0.101	−2.74	−2.19
1.00	0.983	1.000	0.203	0.091	−2.80	−2.17

and

$$\hat{t}_1 = [\hat{V}(\hat{\rho}_w)]^{-1/2}(\hat{\rho}_w - \rho),$$

where $\hat{\rho}_w$ is the weighted symmetric estimator and $\tilde{\rho}$ is the estimator (10.1.55). Because $\tilde{\rho}$ has small median bias, the pivotal statistic $\tilde{t}_1$ has a median close to zero for all values of ρ.

Because the 2.5 percentile for $\hat{\tau}_{w1}$ is −2.80 for $n = 200$ and $\rho = 1$, the 2.5

percentile of $\tilde{t}_1$ is -2.17 for $n = 200$ and $\rho = 1$. The estimator $\hat{\rho}_w$ has a negative bias for positive ρ, and the 2.5 percentile is less than -2.0 for ρ's that are a considerable distance from one. The percentiles for $\hat{t}_1$ show a steady decrease to about -2.8 at $\rho = 1.00$.

The empirical 97.5 percentiles of $\tilde{t}_1$ and $\hat{t}_1$ are compared in Table 10.1.3. The percentiles of $\hat{t}_1$ decline as ρ moves towards one. The maximum value of the 97.5 percentile for $\tilde{t}_1$ is approximately equal to 2.30 for $1 - \rho = 12n^{-1}$. At least 2.5% of the estimates $\tilde{\rho}$ are equal to one for all three sample sizes when $1 - \rho = 6n^{-1}$. The majority of the 97.5 percentiles of $\tilde{t}_1$ fall below 2.30, and the majority of the 2.5 percentiles are above -2.30. Therefore, the interval

$$[\tilde{\rho} - 2.30(\hat{V}\{\hat{\rho}_w\})^{1/2}, \tilde{\rho} + 2.30(\hat{V}\{\hat{\rho}_w\})^{1/2}] \tag{10.1.56}$$

will have coverage greater than 95% for most parameter values and for sample sizes in the 50 to 200 range. Because ρ is restricted to $(-1, 1]$, the actual confidence interval is the intersection of (10.1.56) and $(-1, 1]$.

Example 10.1.3. We continue the analysis of Example 10.1.2 under the assumption that the largest root of the autoregressive process is in the interval $(-1, 1]$. The statistic $\hat{\tau}_{w1}$ from Example 10.1.2 is -1.36. The modified estimator of (10.1.55) is

$$\tilde{\theta}_1 = \hat{\theta}_{w1} + 0.013 = 0.997\,.$$

Regressing $y_t - 0.997y_{t-1}$ on ΔY_{t-1} and ΔY_{t-2}, we obtain the estimated equation,

$$\begin{array}{cccccc} \hat{y}_t = & 0.997\, y_{t-1} & + & 0.355\, \Delta Y_{t-1} & - & 0.103\, \Delta Y_{t-2}\,, \\ & (0.013) & & (0.065) & & (0.066) \end{array}$$

where $y_t = Y_t - \bar{y}_n$. The standard errors for the coefficients of ΔY_{t-1} and ΔY_{t-2} are those computed for the original weighted regression of Example 10.1.2. The standard error for the coefficient of Y_{t-1} is the standard error from Example 10.1.2 increased by 15%. With these standard errors, normal tables can be used for approximate tests and confidence intervals for the parameters of the equation. An

Table 10.1.3. Empirical 97.5 Percentiles of $\tilde{t}$ and $\hat{t}$

	$n = 50$		$n = 100$		$n = 200$	
$1 - \rho$	$\hat{t}_{1,975}$	$\tilde{t}_{1,975}$	$\hat{t}_{1,975}$	$\tilde{t}_{1,975}$	$\hat{t}_{1,975}$	$\tilde{t}_{1,975}$
$25n^{-1}$	1.49	2.02	1.58	2.17	1.55	2.17
$12n^{-1}$	1.26	2.27	1.34	2.29	1.36	2.32
$6n^{-1}$	1.04	$2.04^{\dagger}$	1.11	$2.08^{\dagger}$	1.05	$2.04^{\dagger}$

† More than 2.5% of the estimates equal one.

approximate 95% confidence interval for θ_1 under the assumption that the largest root is in $(-1, 1]$ is $[0.971, 1]$.

The characteristic equation associated with the modified estimates is

$$m^3 - 1.352m^2 + 0.458m - 0.103 = 0,$$

and the largest root is 0.996. Using a Taylor expansion and the estimated covariance matrix with the variance of the coefficient of y_{t-1} increased by 15%, the standard error of the largest root is 0.013. An approximate 95% confidence interval for the largest root is $[0.970, 1]$. ▲▲

10.1.4. Prediction for Unit Root Autoregressions

It is an important result that prediction procedures appropriate for stationary time series are also appropriate for time series with an autoregressive unit root. As in the stationary case, estimation of the parameters adds a term that is $O_p(n^{-1/2})$ to the prediction error when the autoregression contains a unit root.

Theorem 10.1.11. Let Y_t satisfy

$$\begin{aligned} Y_t &= \alpha_0 - \sum_{j=1}^{p} \alpha_j Y_{t-j} + e_t \\ &= \alpha_0 + \theta_1 Y_{t-1} + \sum_{j=2}^{p} \theta_j (Y_{t-j+1} - Y_{t-j}) + e_t, \end{aligned}$$

where $\theta_1 = -\Sigma_{j=1}^{p} \alpha_j$, $\theta_j = \Sigma_{i=j}^{p} \alpha_i$, $j = 2, 3, \ldots, p$, and e_t are iid$(0, \sigma^2)$ random variables. Let $Y_1, Y_2, \ldots, Y_p$ be fixed, and let $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\theta}}$ be the least squares estimators of $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_p)'$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)'$. Assume that one of the roots of the characteristic equation is one and that the other roots are less than one in absolute value. Let $\hat{Y}_{n+s}$ be the predictor of Y_{n+s} based upon $\hat{\boldsymbol{\alpha}}$. Then

$$Y_{n+s} - \hat{Y}_{n+s} = \sum_{j=0}^{s-1} w_j e_{n+s-j} + O_p(n^{-1/2}),$$

where the w_j satisfy the difference equation

$$w_j + \sum_{i=1}^{p} \alpha_i w_{j-i} = 0$$

subject to the initial conditions $w_0 = 1$ and

$$w_j + \sum_{i=1}^{j} \alpha_i w_{j-i} = 0, \qquad j = 1, 2, \ldots, p-1.$$

Proof. By Theorems 10.1.4 and 10.1.5, $\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} = O_p(n^{-1/2})$ and

$$\hat{\theta}_1 - \theta_1 = \begin{cases} O_p(n^{-1}) & \text{if} \quad \alpha_0 = 0 \,, \\ O_p(n^{-3/2}) & \text{if} \quad \alpha_0 \neq 0 \,. \end{cases}$$

It follows that

$$\begin{aligned} Y_{n+1} - \hat{Y}_{n+1} &= \alpha_0 - \hat{\alpha}_0 + (\theta_1 - \hat{\theta}_1)Y_n + \sum_{j=2}^{p} (\theta_j - \hat{\theta}_j)\,\Delta Y_{n-j+2} + e_{n+1} \\ &= e_{n+1} + O_p(n^{-1/2}) \end{aligned}$$

because $\Delta Y_t = O_p(1)$, $Y_n = O_p(n^{1/2})$ if $\alpha_0 = 0$, and $Y_n = O_p(n)$ if $\alpha_0 \neq 0$. Repeated substitution into the expression

$$\hat{Y}_{n+s} = \hat{\alpha}_0 + \hat{\theta}_1 \hat{Y}_{n+s-1} + \sum_{j=2}^{p} \hat{\theta}_j(\hat{Y}_{n-j+s+1} - \hat{Y}_{n-j+s})$$

yields the result for $s > 1$. ▲

It follows from Theorem 10.1.11 that the estimator of the prediction variance given in (8.5.14) for the stationary process can also be used for a process with a unit root.

Example 10.1.4. By Theorem 8.5.1 and Theorem 10.1.11, the prediction variance formula that ignores the effect of estimation error can be used for models whose autoregressive part has roots less than or equal to one in absolute value. In Examples 10.1.1 and 10.1.2, we found the behavior of the one-year treasury bill interest rate to be consistent with the hypothesis of a unit root. Least squares or maximum likelihood can be used to construct predictors that are appropriate whether or not the process has a unit root.

The one-, two-, and three-period forecasts computed with the third order process estimated by stationary maximum likelihood are 9.01, 8.95, and 8.87, respectively. The standard errors estimated using the approximation that ignores the estimation error are 0.291, 0.486, and 0.626 for the one-, two-, and three-period forecasts. The corresponding approximate standard errors computed by the least squares prediction formulas are 0.292, 0.493, and 0.641. The least squares standard errors were computed by fitting the model with three explanatory variables for the three predictions as described in Example 8.5.1. ▲▲

10.2. EXPLOSIVE AUTOREGRESSIVE TIME SERIES

We begin our study of explosive autoregressive time series with the first order process. Let

$$Y_t = \begin{cases} \theta_1 Y_{t-1} + e_t , & t = 1, 2, \ldots , \\ 0 , & t = 0 , \end{cases} \tag{10.2.1}$$

where $|\theta_1| > 1$ and $\{e_t\}_{t=-\infty}^{\infty}$ is a sequence of i.i.d. random variables with mean zero and positive variance σ^2. Equation (10.2.1) is understood to mean that Y_t is created by adding e_t to $\theta_1 Y_{t-1}$. As with other autoregressive processes, repeated substitution enables us to express Y_t as a weighted average of the e_t,

$$Y_t = \sum_{j=0}^{t-1} \theta_1^j e_{t-j} . \tag{10.2.2}$$

It follows that, for $|\theta_1| > 1$, the variance of Y_t is

$$V\{Y_t\} = (\theta_1^2 - 1)^{-1}(\theta_1^{2t} - 1)\sigma^2 \tag{10.2.3}$$

and the variance increases exponentially in t.

In expression (10.2.2), the term in e_1 is $\theta_1^{t-1} e_1$. When $|\theta_1| > 1$, the weight θ_1^{t-1} increases exponentially as t increases. Hence, the variance contribution from e_1 is of order θ_1^{2t}. If $|\theta_1| < 1$, the weight applied to the first observation declines to zero as t increases. If $|\theta_1| = 1$, the weight of the first observation becomes small relative to the standard deviation of the process.

To further understand the importance of the first few e_t in the model with $|\theta_1| > 1$, we observe that

$$\begin{aligned} Y_t &= \sum_{j=0}^{t-1} \theta_1^j e_{t-j} = \sum_{i=1}^{t} \theta_1^{t-i} e_i \\ &= \theta_1^t \sum_{i=1}^{t} \theta_1^{-i} e_i . \end{aligned} \tag{10.2.4}$$

Letting $X_t = \sum_{i=1}^{t} \theta_1^{-i} e_i$, we have

$$X_t \rightarrow X \qquad \text{a.s.} \tag{10.2.5}$$

as t increases, where $X = \sum_{i=1}^{\infty} \theta_1^{-i} e_i$. Thus for large t, the behavior of Y_t is essentially that of an exponential multiple of the random variable X.

The model (10.2.1) extended to include an intercept term and a general initial value is

$$Y_t = \begin{cases} \theta_0 + \theta_1 Y_{t-1} + e_t , & t = 1, 2, \ldots , \\ y_0 , & t = 0 . \end{cases} \tag{10.2.6}$$

The least squares estimators of θ_0 and θ_1 are

$$\begin{aligned} \hat{\theta}_1 &= \left[\sum_{t=1}^{n} (Y_{t-1} - \bar{y}_{(-1)})^2 \right]^{-1} \sum_{t=1}^{n} (Y_{t-1} - \bar{y}_{(-1)})(Y_t - \bar{y}_{(0)}) , \\ \hat{\theta}_0 &= \bar{y}_{(0)} - \hat{\theta}_1 \bar{y}_{(-1)} , \end{aligned} \tag{10.2.7}$$

where

$$\bar{y}_{(i)} = n^{-1} \sum_{t=1}^{n} Y_{t+i}, \qquad i = 0, -1.$$

If y_0 is fixed and $e_t \sim \mathrm{NI}(0, \sigma^2)$, the least squares estimators are the maximum likelihood estimators.

Rubin (1950) showed that $p\lim \hat{\theta}_1 = \theta_1$ for the least squares estimator of the model (10.2.1). White (1958, 1959) considered the asymptotic properties of the least squares statistics for (10.2.1) under the assumption that $e_t \sim \mathrm{NI}(0, \sigma^2)$. White used moment generating function techniques to obtain his results. Anderson (1959) extended White's results using the representation (10.2.4). Hasza (1977) studied estimators for the general model (10.2.6), and Theorem 10.2.1 is taken from that work.

Theorem 10.2.1. Let $\{Y_t\}_{t=0}^{\infty}$ be defined by (10.2.6) with $|\theta_1| > 1$ and $\{e_t\}$ a sequence of iid$(0, \sigma^2)$ random variables. Let the least squares estimators be defined by (10.2.7). The limiting distribution of $(\theta_1^2 - 1)^{-1}\theta_1^n(\hat{\theta}_1 - \theta_1)$ is that of $W_1 W_2^{-1}$, where W_1 and W_2 are independent random variables,

$$(W_1, W_2) = \left(\sum_{i=1}^{\infty} \theta_1^{-i} e_{-i}, \delta_0 + \sum_{j=1}^{\infty} \theta_1^{-j} e_j \right),$$

and $\delta_0 = y_0 + \theta_0(\theta_1 - 1)^{-1}$. The limiting distribution of $n^{1/2}(\hat{\theta}_0 - \theta_0)$ is that of a $N(0, \sigma^2)$ random variable, and $n^{1/2}(\hat{\theta}_0 - \theta_0)$ is independent of $\theta_1^n(\hat{\theta}_1 - \theta_1)$ in the limit. If $e_t \sim \mathrm{NI}(0, \sigma^2)$, $y_0 = 0$, and $\theta_0 = 0$, then $(\theta_1^2 - 1)^{-1}\theta_1^n(\hat{\theta}_1 - \theta_1)$ converges in distribution to a Cauchy random variable.

Proof. We may write Y_t as

$$\begin{aligned} Y_t &= \theta_1^t \left(y_0 + \theta_0 \sum_{j=1}^{t} \theta_1^{-j} + \sum_{j=1}^{t} \theta_1^{-j} e_j \right) \\ &= \theta_1^t X_t, \qquad t = 0, 1, \ldots, \end{aligned} \tag{10.2.8}$$

where

$$(X_t, \delta_{0t}) = \left(\delta_{0t} + \sum_{j=1}^{t} \theta_1^{-j} e_j, y_0 + \theta_0 \sum_{j=1}^{t} \theta_1^{-j} \right).$$

Now X_t converges a.s. as $t \to \infty$ to a random variable

$$X = \delta_0 + \sum_{j=1}^{\infty} \theta_1^{-j} e_j,$$

where $\delta_0 = y_0 + \theta_0(\theta_1 - 1)^{-1}$. Then $\theta_1^{-t} Y_t \to X$ a.s. as t increases. Note that

$$X - X_t = \sum_{j=t+1}^{\infty} (\theta_0 \theta_1^{-j} + \theta_1^{-j} e_j) = O_p(|\theta_1|^{-t}) .$$

The error in the least squares estimator of θ_1 is

$$\hat{\theta}_1 - \theta_1 = \frac{\sum_{t=1}^{n} Y_{t-1} e_t - n\bar{y}_{(-1)}\bar{e}}{\sum_{t=1}^{n} Y_{t-1}^2 - n\bar{y}_{(-1)}^2} , \tag{10.2.9}$$

where

$$\begin{aligned} \theta_1^{-2n} \sum_{t=1}^{n} Y_{t-1}^2 = \sum_{j=0}^{n-1} \theta_1^{-2j-2} X^2 + 2 \sum_{t=1}^{n} \theta_1^{2(t-n)-2} X(X_{t-1} - X) \\ + \sum_{t=1}^{n} \theta_1^{2(t-n)-2} (X_{t-1} - X)^2 . \end{aligned} \tag{10.2.10}$$

Using

$$E\{(X_{t-1} - X)^2\} = (1 - \theta_1^{-2})^{-1} \theta_1^{-2t} \sigma^2 + \theta_0^2 (\theta_1 - 1)^{-2} \theta_1^{-2(t-1)} = O(|\theta_1|^{-2t}) ,$$

$$\operatorname{Var}\left\{ \sum_{t=1}^{n} \theta_1^{2(t-n)-2} \sum_{j=t}^{\infty} \theta_1^{-j} e_j \right\} = O(|\theta_1|^{-2n}) ,$$

and

$$\sum_{t=1}^{n} \theta_1^{2(t-n)-2} (X_{t-1} - X)^2 = O_p(n|\theta_1|^{-2n}) , \tag{10.2.11}$$

we have

$$\theta_1^{-2n} \sum_{t=1}^{n} Y_{t-1}^2 = (\theta_1^2 - 1)^{-1} X^2 + O_p(|\theta_1|^{-n}) . \tag{10.2.12}$$

By similar arguments

$$\begin{aligned} \theta_1^{-n} \sum_{t=1}^{n} Y_{t-1} e_t &= X \sum_{t=1}^{n} \theta_1^{t-n-1} e_t + O_p(n|\theta_1|^{-n}) , \\ \theta_1^{-n} \bar{y}_{(-1)} &= n^{-1} (\theta_1 - 1)^{-1} X + O_p(|\theta_1|^{-n}) , \end{aligned} \tag{10.2.13}$$

and

$$[n^{1/2} \theta_1^{-n} \bar{y}_{(-1)} \bar{e},\ n\theta_1^{-2n} \bar{y}_{(-1)}^2] = O_p(n^{-1}) .$$

Therefore,

$$\theta_1^{-2n} \sum_{t=1}^{n} (Y_{t-1} - \bar{y}_{(-1)})^2 = (\theta_1^2 - 1)^{-1} X^2 + O_p(n^{-1})$$

and

$$\theta_1^{-n} \sum_{t=1}^{n} (Y_{t-1} - \bar{y}_{(-1)})(e_t - \bar{e}) = X \sum_{t=1}^{n} \theta_1^{t-n-1} e_t + O_p(n^{-1/2}) .$$

It follows that

$$(\theta_1^2 - 1)^{-1} \theta_1^n (\hat{\theta}_1 - \theta_1) = X^{-1} Z_n + O_p(n^{-1/2}) , \tag{10.2.14}$$

where $Z_n = \Sigma_{t=1}^n \theta_1^{t-n-1} e_t = \Sigma_{j=1}^n \theta_1^{-j} e_{n-j+1}$. Because

$$Z_n = \sum_{j<n/2} \theta_1^{-j} e_{n-j+1} + O_p(|\theta_1|^{-n})$$

and

$$X = \delta_0 + \sum_{j<n/2} \theta_1^{-j} e_j + O_p(|\theta_1|^{-n}) ,$$

Z_n and X are asymptotically independent.

Therefore, $(\theta_1^2 - 1)^{-1} \theta_1^n (\hat{\theta}_1 - \theta_1)$ has a nondegenerate limiting distribution. If the e_t's are $\mathrm{NI}(0, \sigma^2)$ random variables, then Z_n and X, both being linear combinations of the e_t's, are normally distributed. The mean of the Z_n is zero and the mean of X is δ_0, so that the limiting distribution of $(\theta_1^2 - 1)^{-1} \theta_1^n (\hat{\theta}_1 - \theta_1)$ is that of $W_1 W_2^{-1}$, where (W_1, W_2) is a normal vector. If $\delta_0 = 0$ and the e_t are normal, then the limiting distribution of $(\theta_1^2 - 1)^{-1} \theta_1^n (\hat{\theta}_1 - \theta_1)$ is that of the ratio of two independent standard normal variables, which is Cauchy.

Now, from (10.2.7),

$$\hat{\theta}_0 - \theta_0 = \bar{e}_{(0)} - (\hat{\theta}_1 - \theta_1)\bar{y}_{(-1)} = \bar{e}_{(0)} + O_p(n^{-1})$$

by (10.2.13) and (10.2.14). Also, $n^{1/2}\bar{e}_{(0)}$ is independent of X and of Z_n in the limit, because the elements of

$$\mathbf{L}_n = \left(n^{-1/2} \sum_{t=1}^{n-k_n} e_t, \sum_{j=1}^{k_n} \theta_1^{-j} e_{n-j+1} \right)$$

are independent and the vector $\mathbf{L}_n - (n^{1/2}\bar{e}_0, Z_n)$ converges in probability to zero for $n^{0\ 1} < k_n < n^{0.2}$, for example. ▲

The distributions of other regression statistics follow from Theorem 10.2.1. Let the regression residual mean square be

$$\hat{\sigma}^2 = (n-2)^{-1} \sum_{t=1}^{n} [Y_t - \bar{y}_{(0)} - \hat{\theta}_1(Y_{t-1} - \bar{y}_{(-1)})]^2 . \tag{10.2.15}$$

Corollary 10.2.1.1. Suppose the assumptions of Theorem 10.2.1 hold. Then

$$\operatorname*{plim}_{n\to\infty} \hat{\sigma}^2 = \sigma^2 .$$

Proof. Using the least squares definition of $\hat{\theta}_1$,

$$\hat{\sigma}^2 = (n-2)^{-1}\left[\sum_{t=1}^{n} (e_t - \bar{e})^2 - (\hat{\theta}_1 - \theta_1)^2 \sum_{t=1}^{n} (Y_{t-1} - \bar{y}_{(-1)})^2\right].$$

We have shown that $(\hat{\theta}_1 - \theta_1) = O_p(|\theta_1|^{-n})$ and that $\Sigma_{t=1}^{n}(Y_{t-1} - \bar{y}_{(-1)})^2 = O_p(|\theta_1|^{-2n})$. Because the e_t are iid$(0, \sigma^2)$, the mean square of the e_t converges in probability to σ^2. ▲

Corollary 10.2.1.2. Suppose that $e_t \sim \mathrm{NI}(0, \sigma^2)$. Then, under the model of Theorem 10.2.1,

$$\hat{\tau} = (\hat{\theta}_1 - \theta_1)\left[\hat{\sigma}^{-2} \sum_{t=1}^{n} (Y_{t-1} - \bar{y}_{(-1)})^2\right]^{1/2} \xrightarrow{\mathcal{L}} N(0, 1) . \tag{10.2.16}$$

Proof. We have

$$\begin{aligned}
&(\hat{\theta}_1 - \theta_1)\left(\sum_{t=1}^{n} (Y_{t-1} - \bar{y}_{(-1)})^2\right)^{1/2} \\
&\quad = [(\theta_1^2 - 1)X^{-1}Z_n][(\theta_1^2 - 1)^{-1/2}|X|] + O_p(n|\theta_1|^{-n}) \\
&\quad = (\operatorname{sgn} X)(\theta_1^2 - 1)^{1/2} Z_n + O_p(n|\theta_1|^{-n}) .
\end{aligned}$$

The random variables X and Z_n are asymptotically independent, and the variance of Z_n converges to $(\theta_1^2 - 1)^{-1}\sigma^2$. The result follows, because $\hat{\sigma}^2 \xrightarrow{P} \sigma^2$ by Corollary 10.2.1.1. ▲

Because the pivotal $\hat{\tau}$ is approximately $N(0, 1)$ in large samples, $\hat{\tau}$ may be used to test hypotheses concerning θ_1 and to set confidence intervals for θ_1 when the original errors are normally distributed.

We now extend these results to the pth order process with a single root larger than one in absolute value. Let

$$Y_t + \sum_{j=1}^{p} \alpha_j Y_{t-j} = \begin{cases} \theta_0 + e_t , & t = 1, 2, \ldots , \\ y_t , & t = -p+1, -p+2, \ldots, 0 , \end{cases} \tag{10.2.17}$$

where $(y_{-p+1}, y_{-p+2}, \ldots, y_0)$ is an initial vector and $\{e_t\}_{t=-\infty}^{\infty}$ is a sequence of independent identically distributed $(0, \sigma^2)$ random variables. Let $|m_1| > |m_2| \geq |m_3| \geq \cdots \geq |m_p|$ be the roots of the characteristic equation

$$m_p + \sum_{j=1}^{p} \alpha_j m^{p-j} = 0 . \tag{10.2.18}$$

Assume $|m_1| > 1$ and all other roots are less than one in absolute value. We can also write the first equation of model (10.2.17) as

$$Y_t = \theta_0 + m_1 Y_{t-1} + \sum_{j=2}^{p} \theta_j (Y_{t-j+1} - m_1 Y_{t-j}) + e_t , \tag{10.2.19}$$

where the roots of

$$m^{p-1} - \sum_{j=2}^{p} \theta_j m^{p-j} = \prod_{j=1}^{p-1} (m - m_{j+1}) = 0 \tag{10.2.20}$$

are the $p-1$ roots of (10.2.18) that are less than one in absolute value. We consider a conceptual regression with the true m_1, denoted by m_1^0, used to define the regression variables. Let

$$\hat{\boldsymbol{\theta}} = \left(\sum_{t=1}^{n} \mathbf{L}_t' \mathbf{L}_t \right)^{-1} \sum_{t=1}^{n} \mathbf{L}_t' Y_t , \tag{10.2.21}$$

where

$$\mathbf{L}_t = (1, Y_{t-1}, Y_{t-1} - m_1^0 Y_{t-2}, \ldots, Y_{t-p+1} - m_1^0 Y_{t-p}) ,$$

be the least squares estimator of $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_p)'$, where $\theta_1 = m_1$. The conclusion of Theorem 10.2.2 is analogous to that of Theorem 10.1.2 in that the limiting distribution of the coefficient of Y_{t-1} is closely related to the limiting distribution in the first order case and the other coefficients have the limiting distribution associated with the stationary process.

Two alternative assumptions for the initial conditions can be considered. In the first, the vector $(y_{-p+1}, y_{-p+2}, \ldots, y_0)$ is treated as fixed. In the second, we let

$$Y_t = \begin{cases} \theta_0 + \theta_1 Y_{t-1} + Z_t , & t = 1, 2, \ldots , \\ y_0 , & t = 0 , \end{cases} \tag{10.2.22}$$

where y_0 is fixed and $\{Z_t\}_{t=-\infty}^{\infty}$ is a stationary autoregressive process with characteristic equation (10.2.20). The limiting distributions are the same under the two formulations.

Theorem 10.2.2. Suppose the model (10.2.17), (10.2.22) holds with $|\theta_1^0| > 1$ and the roots of (10.2.20) less than one in absolute value, where $\boldsymbol{\theta}^0$ is the true parameter vector. Let $\hat{\boldsymbol{\theta}}$ be defined by (10.2.21), and let

$$\boldsymbol{\xi}_n = n^{1/2}[(\hat{\theta}_0 - \theta_0^0), (\hat{\theta}_2 - \theta_2^0), \ldots, (\hat{\theta}_p - \theta_p^0)]' .$$

Then $d_{n11}^{1/2}(\hat{\theta}_1 - \theta_1^0)$ converges to a random variable with mean zero and variance σ^2 and

$$\xi_n \xrightarrow{\mathscr{L}} N(\mathbf{0}, \mathbf{A}^{-1}\sigma^2),$$

where $\mathbf{A} = E\{(1, Z_{t-1}, \ldots, Z_{t-p})'(1, Z_{t-1}, \ldots, Z_{t-p})\}$ and $d_{n11} = \Sigma_{t=2}^n Y_{t-1}^2$. Furthermore, ξ_n and $d_{n11}^{1/2}(\hat{\theta}_1 - \theta_1)$ are independent in the limit. If $e_t \sim \mathrm{NI}(0, \sigma^2)$, then

$$d_{n11}^{1/2}(\hat{\theta}_1 - \theta_1^0) \xrightarrow{\mathscr{L}} N(0, \sigma^2).$$

Proof. We have

$$\mathbf{D}_n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) = \left(\mathbf{D}_n^{-1/2} \sum_{t=1}^n \mathbf{L}_t'\mathbf{L}_t\mathbf{D}_n^{-1/2}\right)^{-1} \mathbf{D}_n^{-1/2} \sum_{t=1}^n \mathbf{L}_t'e_t,$$

where

$$\mathbf{D}_n = \operatorname{diag}\left[n, \left(\sum_{t=2}^n Y_{t-1}^2\right), n, \ldots, n\right].$$

By the arguments used in Theorem 10.2.1

$$Y_t = (m_1^0)^t\left[y_0 + \theta_0^0 \sum_{j=1}^t (m_1^0)^{-j} + \sum_{j=1}^t (m_1^0)^{-j}Z_j\right] = (m_1^0)^t U_t$$

and U_t converges to

$$U = y_0 + \theta_0^0(m_1^0 - 1)^{-1} + \sum_{j=1}^\infty (m_1^0)^{-j}Z_j.$$

To simplify the notation, we now write m_1 for m_1^0. It follows from the arguments used to obtain (10.2.12) and (10.2.14) that

$$\frac{\Sigma_{t=1}^n Y_{t-1}e_t}{(\Sigma_{t=1}^n Y_{t-1}^2)^{1/2}} = \frac{\Sigma_{t=1}^n m_1^{-(n-t+1)}e_t}{(m_1^2 - 1)^{-1/2}} \operatorname{sgn} U + O_p(n|m_1|^{-n}).$$

Now,

$$\sum_{t=1}^n Y_{t-1}Z_{t-j} = \sum_{t=1}^n m_1^{t-1}UZ_{t-j} + \sum_{t=1}^n m_1^{t-1}(U_{t-1} - U)Z_{t-j}.$$

We have

$$E\{|U_t - U|\,|Z_{t-j}|\} = O(|m_1|^{-t})$$

and

$$E\left\{\sum_{t=1}^{n} m_1^{t-1}(U_{t-1} - U)Z_{t-j}\right\} = O(n|m_1|^{-n}).$$

Therefore,

$$m_1^{-n} \sum_{t=1}^{n} Y_{t-1}Z_{t-j} = U \sum_{t=1}^{n} m_1^{-n+t-1}Z_{t-j} + O_p(n|m_1|^{-n})$$

and

$$\left(n \sum_{t=1}^{n} Y_{t-1}^2\right)^{-1/2} \sum_{t=1}^{n} Y_{t-1}Z_{t-j} = O_p(n^{-1/2}).$$

It follows that

$$\mathbf{D}_n^{-1/2} \sum_{t=1}^{n} \mathbf{L}_t'\mathbf{L}_t\mathbf{D}_n^{-1/2} \longrightarrow \mathbf{A}^*,$$

where the row in $\mathbf{A}^*$ associated with $\hat{\theta}_1$ is $(0, 1, 0, \ldots, 0)$. The matrix obtained by deleting the second row and column of $\mathbf{A}^*$ is $\mathbf{A}$. By the arguments of Theorem 8.2.1,

$$n^{-1/2}\left[\sum_{t=1}^{n} e_t, \sum_{t=1}^{n} Z_{t-1}e_t, \ldots, \sum_{t=1}^{n} Z_{t-p}e_t\right]$$

converges to a normal vector random variable and we have the limiting result for $\boldsymbol{\xi}_n$. The remaining results follow from arguments used in the proof of Theorem 10.2.1. ▲

In Theorem 10.2.2 we obtained the limiting distribution of estimators where some of the explanatory variables were a function of the unknown true maximum root. These results can be used to construct confidence intervals for the largest root. This operation is illustrated in Example 10.2.1. In Corollary 10.2.2 we give the limiting distribution of the ordinary regression pivotal for the largest root.

Corollary 10.2.2. Let $(\tilde{\theta}_0, \tilde{m}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_p)$ be the nonlinear least squares estimator of the parameters of (10.2.19), where $|\tilde{m}_1|$ is at least as large as the largest absolute value of the roots of the estimated equation (10.2.20). Let the assumptions of Theorem 10.2.2 hold with $e_t \sim \mathrm{NI}(0, \sigma^2)$. Let

$$\tilde{\mathbf{V}} = \left[\sum_{t=1}^{n} \tilde{\mathbf{L}}_t'\tilde{\mathbf{L}}_t\right]^{-1} \tilde{\sigma}^2,$$

where $\tilde{\mathbf{L}}_t = (1,\ Y_{t-1} - \sum_{j=2}^{p} \tilde{\theta}_j Y_{t-j},\ Y_{t-1} - \tilde{m}_1 Y_{t-2}, \ldots, Y_{t-p+1} - \tilde{m}_1 Y_{t-p})$, and

$$\hat{\sigma}^2 = (n-p-1)^{-1} \sum_{t=1}^{n} \left[Y_t - \tilde{\theta}_0 - \tilde{m}_1 Y_{t-1} - \sum_{j=2}^{p} \tilde{\theta}_j (Y_{t-j+1} - \tilde{m}_1 Y_{t-j}) \right]^2 .$$

Then

$$\tilde{t} = \tilde{v}_{11}^{-1/2} (\tilde{m}_1 - m_1^0) \xrightarrow{\mathscr{L}} N(0, 1) ,$$

where $\tilde{v}_{11}$ is the second diagonal element of $\tilde{\mathbf{V}}$, the element associated with $\tilde{m}_1$.

Proof. The nonlinear least squares estimator for m_1, based on the model (10.2.19), differs from the root calculated with the characteristic equation of the linear least squares fit of (10.2.17) only in that the $\tilde{m}_1$ from the nonlinear fit is always a real number. By Theorem 10.2.2, the linear least squares estimators are consistent. Because the true largest root is real, the probability approaches one that the two estimation procedures agree. The estimated characteristic equation, based on the estimator of (10.2.21), is

$$m^p - (\hat{\theta}_1 + \hat{\theta}_2) m^{p-1} - \sum_{j=2}^{p} (\hat{\theta}_{j+1} - m_1^0 \hat{\theta}_j) m^{p-j} + m_1^0 \hat{\theta}_p = 0 . \quad (10.2.23)$$

The coefficients of Y_{t-j}, $j = 1, 2, \ldots, p$, obtained from the estimator (10.2.21) are the same as those obtained by the least squares fit of the model (10.2.17). Therefore, the roots of (10.2.23) are the same as the roots of the estimated characteristic equation associated with the least squares estimator of (10.2.17).

By a Taylor expansion, we have that the largest root of (10.2.23), denoted by $\hat{m}_1$, satisfies

$$\begin{aligned} \hat{m}_1 - m_1^0 = {} & \left[(m_1^0)^{p-1} - \sum_{j=2}^{p} \theta_j^0 (m_1^0)^{p-j} \right]^{-1} (m_1^0)^{p-1} (\hat{\theta}_1 - m_1^0) \\ & + o_p(|m_1^0|^{-n}) \end{aligned} \quad (10.2.24)$$

and $\tilde{m}_1 - m_1^0 = O_p(|m_1^0|^{-n})$. It follows that

$$\sum_{t=1}^{n} (Y_{t-j+1} - \tilde{m}_1 Y_{t-j})^2 = \sum_{t=1}^{n} (Y_{t-j+1} - m_1^0 Y_{t-j})^2 + O_p(1) .$$

Also,

$$\sum_{t=1}^{n} \left(Y_{t-1} - \sum_{j=2}^{p} \tilde{\theta}_j Y_{t-j} \right)^2 = \sum_{t=1}^{n} \left(Y_{t-1} - \sum_{j=2}^{p} \theta_j^0 Y_{t-j} \right)^2 + o_p(|m_1^0|^n) ,$$

where

$$\sum_{t=1}^{n}\left(Y_{t-1}-\sum_{j=2}^{p}\theta_j^0 Y_{t-j}\right)^2 = U^2\sum_{t=1}^{n}\left[(m_1^0)^{t-1}\left(1-\sum_{j=2}^{p}\theta_j^0(m_1^0)^{-j+1}\right)\right]^2 + O_p(|m_1^0|^n) \tag{10.2.25}$$

and U is defined in the proof of Theorem 10.2.2. By the arguments used in the proof of Theorem 10.2.2,

$$\left[n\sum_{t=1}^{n}\left(Y_{t-1}-\sum_{j=2}^{p}\theta_j^0 Y_{t-j}\right)^2\right]^{-1/2}\sum_{t=1}^{n}\left(Y_{t-1}-\sum_{j=2}^{p}\theta_j^0 Y_{t-j}\right)Z_{t-j} = O_p(n^{-1/2}).$$

Therefore,

$$\operatorname*{plim}_{n\to\infty}\sigma^{-2}\sum_{t=1}^{n}\left(Y_{t-1}-\sum_{j=2}^{p}\theta_j^0 Y_{t-j}\right)^2\tilde{v}_{11} = 1.$$

Because the multiplier for $\hat{\theta}_1 - m_1^0$ in (10.2.24) is the square root of the multiplier for $(m_1^0)^{2n}[(m_1^0)^2-1]U^2$ in (10.2.25), the result follows from the limiting distribution of $d_{n11}^{1/2}(\hat{\theta}_1-\theta_1^0)$ given in Theorem 10.2.2. ▲

Example 10.2.1. Engle and Kraft (1981) analyzed the logarithm of the implicit price deflator for the gross national product as an autoregressive time series. The analysis we present follows Fuller (1984). We simplify the model of Engle and Kraft and use data for the period 1955 first quarter through 1980 third quarter. If we use least squares to estimate an autoregressive equation of the third order, 100 observations are used in the regression. The estimated characteristic equation associated with the third order process is

$$m^3 - 1.429m^2 + 0.133m + 0.290 = 0,$$

and the largest root of this equation is 1.0178. Because the largest root is greater than one, the estimated model is explosive. Before accepting an explosive model, it is reasonable to test the hypothesis that the largest root is one. To test the hypothesis of a unit root, we regress the first difference on Y_{t-1} and the lagged first differences. The estimated equation is

$$\begin{aligned}\hat{Y}_t - Y_{t-1} = &-\underset{(0.0082)}{0.0211} + \underset{(0.0020)}{0.0054}\,Y_{t-1} + \underset{(0.098)}{0.423}\,(Y_{t-1}-Y_{t-2})\\ &+\underset{(0.099)}{0.290}\,(Y_{t-2}-Y_{t-3}),\end{aligned}$$

where the numbers in parentheses are the estimated standard errors obtained from the ordinary least squares regression program. By Theorem 10.1.4, the statistic

$$\hat{\tau}_\mu = (0.0020)^{-1}(0.0054) = 2.72$$

has the distribution tabulated in the second part of Table 10.A.2 when the largest root is one. According to that table, $\hat{\tau}_\mu$ will exceed 0.63 about 1% of the time when the true root is one. Therefore, the hypothesis of a unit root is easily rejected.

To set confidence limits for the largest root, we use Corollary 10.2.2. Let the coefficient of Y_{t-1} in the regression of $Y_t - m_1 Y_{t-1}$ on 1, Y_{t-1}, $Y_{t-1} - m_1 Y_{t-2}$, and $Y_{t-2} - m_1 Y_{t-3}$ be denoted by $\hat{b}$. If $m_1 > 1$ is the largest root of the characteristic equation and if all other roots are less than one in absolute value, then the limiting distribution of the statistic $\hat{t} = (\text{s.e. } \hat{b})^{-1}\hat{b}$, where s.e. $\hat{b}$ is the ordinary least squares standard error, is that of a $N(0, 1)$ random variable. Therefore, we define a confidence set for m_1 to be those m_1 such that the absolute value of the calculated $\hat{t}$ is less than the tabular value of Student's t for the desired confidence level. The confidence set can be constructed by fitting models for a range of m_1-values and choosing the values for which the test statistic is equal to the tabular value. For our data

$$\begin{aligned}\hat{Y}_t - 1.0091\,Y_{t-1} = &\underset{(0.0082)}{-\,0.0211} + \underset{(0.00139)}{0.00274}\,Y_{t-1} + \underset{(0.098)}{0.417}\,(Y_{t-1} - 1.0091\,Y_{t-2}) \\ &+ \underset{(0.099)}{0.288}\,(Y_{t-2} - 1.0091\,Y_{t-3})\end{aligned}$$

and

$$\begin{aligned}\hat{Y}_t - 1.0256\,Y_{t-1} = &\underset{(0.0082)}{-\,0.0211} - \underset{(0.00128)}{0.00254}\,Y_{t-1} + \underset{(0.098)}{0.406}\,(Y_{t-1} - 1.0256\,Y_{t-2}) \\ &+ \underset{(0.097)}{0.283}\,(Y_{t-2} - 1.0256\,Y_{t-3}).\end{aligned}$$

It follows that a 95% confidence interval for m_1 based on the large sample theory is (1.0091, 1.0256).

By Corollary 10.2.2, we can construct tests for the stationary part of the model. For example, the ordinary regression t-statistic for $Y_{t-3} - m_1 Y_{t-4}$ in the regression of $Y_t - m_1 Y_{t-1}$ on $(1, Y_{t-1}, Y_{t-1} - m_1 Y_{t-2}, Y_{t-2} - m_1 Y_{t-3}, Y_{t-3} - m_1 Y_{t-4})$ has a $N(0, 1)$ distribution in the limit. Because the t-statistic for the hypothesis that the coefficient for $Y_{t-3} - m_1 Y_{t-4}$ is zero is identical (for any $m_1 \neq 0$) to the t-statistic for the coefficient of Y_{t-4} in the regression of Y_t on $(1, Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-4})$, we have a test of the hypothesis that the process is third order against the hypothesis that it is fourth order. The fitted fourth order model, using the estimated value of m_1 is

$$\begin{aligned}\hat{Y}_t - 1.0178\,Y_{t-1} = &\underset{(0.0086)}{-\,0.0207} - \underset{(0.00114)}{0.00000}\,Y_{t-1} + \underset{(0.104)}{0.409}\,(Y_{t-1} - 1.0178\,Y_{t-2}) \\ &+ \underset{(0.108)}{0.280}\,(Y_{t-2} - 1.0178\,Y_{t-3}) + \underset{(0.104)}{0.012}\,(Y_{t-3} - 1.0178\,Y_{t-4}).\end{aligned}$$

Because the regression pivotal for Y_{t-4} is $t = (0.104)^{-1}0.012 = 0.12$, we easily accept the hypothesis that the process is third order. The argument extends to the use of an F to test that the vector of coefficients for a set of variables is zero. ▲▲

We now consider the problem of predicting Y_{n+s} for the explosive autoregressive model. We derive the results for the first order model

$$Y_t = \theta_1 Y_{t-1} + e_t ,$$

where the e_t are NI$(0, \sigma^2)$ random variables. Given $(Y_1, Y_2, \dots, Y_n)$, we take as our prediction of Y_{n+s}

$$\hat{Y}_{n+s} = \hat{\theta}_1^s Y_n , \tag{10.2.26}$$

where $s > 0$. The error in our prediction is

$$\hat{Y}_{n+s} - Y_{n+s} = (\hat{\theta}_1^s - \theta_1^s)Y_n - \sum_{j=1}^{s} \theta_1^{s-j} e_{n+j} ,$$

where

$$\hat{\theta}_1^s - \theta_1^s = s\theta_1^{s-1}(\hat{\theta}_1 - \theta_1) + O_p(|\theta_1|^{-2n}) .$$

Also,

$$\hat{\theta}_1 - \theta_1 = \theta_1^{-n}(\theta_1^2 - 1)X^{-1}Z_n + O_p(n|\theta_1|^{-2n}) ,$$

where $Z_n = \sum_{t=1}^{n} \theta_1^{t-n-1} e_t$ and $X = y_0 + \sum_{j=1}^{\infty} \theta_1^{-j} e_j$. Therefore,

$$\hat{Y}_{n+s} - Y_{n+s} = s\theta_1^{s-1}(\theta_1^2 - 1)Z_n - \sum_{j=1}^{s} \theta_1^{s-j} e_{n+j} + O_p(|\theta_1|)^{-n}) ,$$

where we have used $\theta_1^{-n} Y_n = X + O_p(\theta_1^{-n})$. Now Z_n and $\sum_{j=1}^{s} \theta_1^{s-j} e_{n+j}$ are independent,

$$\mathrm{Var}(Z_n) = \sigma^2(\theta_1^2 - 1)^{-1}(1 - \theta_1^{-2n}) ,$$

and

$$\mathrm{Var}\left(\sum_{j=1}^{s} \theta_1^{s-j} e_{n+j}\right) = \sigma^2(\theta_1^{2s} - 1)(\theta_1^2 - 1)^{-1} .$$

Therefore, the variance of the leading term in the prediction error is

$$\mathrm{Var}(\epsilon_n) = \sigma^2[s^2\theta_1^{2(s-1)}(\theta_1^2 - 1) + (\theta_1^2 - 1)^{-1}(\theta_1^{2s} - 1)] + O(|\theta_1|^{-2n}) , \tag{10.2.27}$$

where

$$\epsilon_n = s\theta_1^{s-1}(\theta_1^2 - 1)Z_n - \sum_{j=1}^{s} \theta_1^{s-j} e_{n+j} .$$

The variance of the prediction error may be estimated by

$$\hat{V}(\hat{Y}_{n+s} - Y_{n+s}) = \hat{\sigma}^2\left[\frac{\hat{\theta}_1^{2s} - 1}{\hat{\theta}_1^2 - 1} + \frac{(s\hat{\theta}_1^{s-1})^2 Y_n^2}{\sum_{t=1}^n Y_{t-1}^2}\right]. \tag{10.2.28}$$

For $s = 1$, this estimated variance reduces to the familiar regression formula for the variance of a prediction.

In the explosive case, unlike the case with $|\theta_1| \le 1$, the error in $\hat{\theta}_1$ contributes to the leading term of the variance of the prediction error. It follows that the usual approximations, such as (8.5.13), are not formally correct when the process has a root larger than one. However, the approximations based upon regression formulas remain appropriate for predictions with $s > 1$ and for higher order processes with a root greater than one in absolute value.

Example 10.2.2. In this example, we construct predictions using the model of Example 10.2.1 and the method of indicator variables described in Example 8.5.1. To obtain predictions for three periods, we write

$$\begin{aligned} Y_t = \theta_0 &+ \theta_1 X_{t1} + \theta_2 X_{t2} + \theta_3 X_{t3} + \theta_5 X_{t5} \\ &+ (\theta_6 - \theta_1\theta_5)X_{t6} + (\theta_7 - \theta_1\theta_6 - \theta_2\theta_5)X_{t7} + e_t, \end{aligned}$$

where the regression variables are defined in Table 8.5.1 of Example 8.5.1 and the estimator of $(\theta_5, \theta_6, \theta_7)$ is the predictor of $(Y_{n+1}, Y_{n+2}, Y_{n+3})$. Using a nonlinear least squares program and the data for 1955–1 through 1980–4, we obtain

$$\begin{aligned} (\hat{Y}_{81-1}, \hat{Y}_{81-2}, \hat{Y}_{81-3}) = (&4.5144, \quad 4.5410, \quad 4.5672), \\ (&0.0044) \quad (0.0074) \quad (0.0107) \end{aligned}$$

where the numbers in parentheses are the standard errors output by the nonlinear least squares program. There are 117 observations used in the regression, and the residual mean square is $\hat{\sigma}^2 = 0.0000179$. Using this estimate of σ^2, the approximation (8.5.13) gives 0.0042 as the standard error for the one-period prediction. Thus, for a root close to one the difference between the appropriate estimator and the approximation based on the stationary model is not large for a one-period prediction. ▲▲

10.3. MULTIVARIATE AUTOREGRESSIVE PROCESSES WITH UNIT ROOTS

10.3.1. Multivariate Random Walk

Before studying estimators for the general autoregressive model, we give the distribution of the least squares estimator for a multivariate version of the first order unit root process. Let

$$\mathbf{Y}_t = \begin{cases} \mathbf{Y}_{t-1} + \mathbf{e}_t, & t = 1, 2, \ldots, \\ \mathbf{Y}_0, & t = 0, \end{cases} \tag{10.3.1}$$

where $\mathbf{Y}_t$ is a k-dimensional column vector and the $\mathbf{e}_t$ are independent identically distributed $(\mathbf{0}, \boldsymbol{\Sigma}_{ee})$ random variables. It is always possible to transform the vector $\mathbf{Y}_t$ into a new vector $\mathbf{X}_t$ that satisfies an equation of the form (10.3.1) with $\boldsymbol{\Sigma}_{ee} = \mathbf{I}$. One transformation is

$$\mathbf{X}_t = \mathbf{T}\mathbf{Y}_t ,$$

where $\mathbf{T} = \boldsymbol{\Sigma}_{ee}^{-1/2} = \mathbf{Q}\mathbf{D}_e^{-1/2}\mathbf{Q}'$, $\mathbf{Q}'$ is the orthogonal matrix such that

$$\mathbf{Q}'\boldsymbol{\Sigma}_{ee}\mathbf{Q} = \mathbf{D}_e ,$$

$\mathbf{Q}'\mathbf{Q} = \mathbf{I}$, and $\mathbf{D}_e$ is diagonal.

We write the general first order autoregressive process as

$$\mathbf{Y}_t = \mathbf{H}_1\mathbf{Y}_{t-1} + \mathbf{e}_t \tag{10.3.2}$$

and consider the least squares estimator of $\mathbf{H}_1$,

$$\hat{\mathbf{H}}_1' = \left(\sum_{t=2}^{n} \mathbf{Y}_{t-1}\mathbf{Y}_{t-1}'\right)^{-1} \sum_{t=2}^{n} \mathbf{Y}_{t-1}\mathbf{Y}_t' . \tag{10.3.3}$$

If $\mathbf{H}_1 = \mathbf{I}$, the normalized error in the least squares estimator is

$$n(\hat{\mathbf{H}}_1' - \mathbf{I}) = \left(n^{-2}\sum_{t=2}^{n} \mathbf{Y}_{t-1}\mathbf{Y}_{t-1}'\right)^{-1} n^{-1}\sum_{t=2}^{n} \mathbf{Y}_{t-1}\mathbf{e}_t' .$$

The quantities in the expression converge in distribution to random variables that are analogous to the limit random variables of Theorem 10.1.1.

The first order vector autoregressive model with an intercept is

$$\mathbf{Y}_t = \mathbf{H}_0 + \mathbf{H}_1\mathbf{Y}_{t-1} + \mathbf{e}_t , \tag{10.3.4}$$

where $\mathbf{H}_0$ is a k-dimensional column vector. The ordinary least squares estimator of $\mathbf{H}_1$ for the model (10.3.4) is

$$\begin{aligned}\hat{\mathbf{H}}_{1,\mu}' = &\left[\sum_{t=2}^{n} (\mathbf{Y}_{t-1} - \bar{\mathbf{Y}}_{(-1)})(\mathbf{Y}_{t-1} - \bar{\mathbf{Y}}_{(-1)})'\right]^{-1} \\ &\times \sum_{t=2}^{n} (\mathbf{Y}_{t-1} - \bar{\mathbf{Y}}_{(-1)})(\mathbf{Y}_t - \bar{\mathbf{Y}}_{(0)})' ,\end{aligned} \tag{10.3.5}$$

where $(\bar{\mathbf{Y}}_{(-1)}, \bar{\mathbf{Y}}_{(0)}) = (n-1)^{-1}\sum_{t=2}^{n}(\mathbf{Y}_{t-1}, \mathbf{Y}_t)$.

Lemma 10.3.1. Let the model (10.3.1) hold. Let $\mathbf{Y}_0$ be a vector random variable with finite covariance matrix, perhaps the zero matrix, independent of $\{\mathbf{e}_t\}$. Let $\{\mathbf{e}_t\}_{t=1}^{\infty}$ be a sequence of random vectors, and let $\mathscr{A}_{t-1}$ be the sigma-field generated by $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{t-1}\}$. Assume

$$E\{(\mathbf{e}_t, \mathbf{e}_t\mathbf{e}_t') \mid \mathscr{A}_{t-1}\} = (\mathbf{0}, \boldsymbol{\Sigma}_{ee}) \qquad \text{a.s.}$$

and

$$E\{|\mathbf{e}_t|^{2+\delta} \mid \mathscr{A}_{t-1}\} < M < \infty \qquad \text{a.s.}$$

for some $\delta > 0$, or assume the $\mathbf{e}_t$ are iid$(\mathbf{0}, \mathbf{\Sigma}_{ee})$, where $\mathbf{\Sigma}_{ee}$ is positive definite. Let $\hat{\mathbf{H}}_1'$ be defined by (10.3.3), and $\hat{\mathbf{H}}_{1,\mu}$ by (10.3.5). Then

$$n\mathbf{\Sigma}_{ee}^{1/2}(\hat{\mathbf{H}}_1' - \mathbf{I})\mathbf{\Sigma}_{ee}^{-1/2} \xrightarrow{\mathscr{L}} \mathbf{G}^{-1}\mathbf{Y},$$

and

$$n\mathbf{\Sigma}_{ee}^{1/2}(\hat{\mathbf{H}}_{1,\mu}' - \mathbf{I})\mathbf{\Sigma}_{ee}^{-1/2} \xrightarrow{\mathscr{L}} (\mathbf{G} - \boldsymbol{\zeta}\boldsymbol{\zeta}')^{-1}(\mathbf{Y} - \boldsymbol{\zeta}\boldsymbol{\eta}'),$$

where

$$\mathbf{G} = \sum_{i=1}^{\infty} \gamma_i^2 \mathbf{U}_i\mathbf{U}_i' = \int_0^1 \mathbf{W}(r)\mathbf{W}'(r)\,dr,$$

$$\mathbf{Y} = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} 2(\gamma_i + \gamma_j)^{-1}\gamma_i^2\gamma_j\mathbf{U}_i\mathbf{U}_j' - 0.5\mathbf{I} = \int_0^1 \mathbf{W}(r)\,d\mathbf{W}'(r),$$

$$\boldsymbol{\zeta} = \sum_{i=1}^{\infty} 2^{1/2}\gamma_i^2\mathbf{U}_i = \int_0^1 \mathbf{W}(r)\,dr,$$

$$\boldsymbol{\eta} = \sum_{i=1}^{\infty} 2^{1/2}\gamma_i\mathbf{U}_i = \mathbf{W}(1),$$

$$\gamma_l = (-1)^{l+1}2[(2l-1)\pi]^{-1},$$

$\mathbf{U}_l = (U_{1l}, U_{2l}, \ldots, U_{kl})'$, $l = 1, 2, \ldots$, is a sequence of NI$(\mathbf{0}, \mathbf{I})$ vectors, and $\mathbf{W}(\mathbf{r})$ is the standard vector Wiener process.

Proof. The result follows immediately from Theorem 5.3.7. See Phillips and Durlauf (1986) and Nagaraj (1990). ▲

The extension of Theorem 5.3.7 to a moving average process will be used repeatedly in this section.

Lemma 10.3.2. Let $\{\mathbf{e}_t\}$ satisfy the assumptions of Theorem 5.3.7. Let

$$\mathbf{Y}_t = \sum_{i=1}^{t}\sum_{j=0}^{\infty} \mathbf{K}_j\mathbf{e}_{i-j},$$

where

$$\sum_{j=0}^{\infty} \|\mathbf{K}_j\| < \infty, \qquad \sum_{j=0}^{\infty} \mathbf{K}_j = \mathbf{C},$$

$\|\mathbf{K}_j\|$ is the maximum of the absolute values of the elements of $\mathbf{K}_j$, and $\mathbf{C}$ is a nonsingular real matrix. Then

$$n^{-3/2}\sum_{t=1}^{n}\mathbf{Y}_t \Rightarrow \mathbf{C}\mathbf{\Sigma}_{ee}^{1/2}\int_0^1 \mathbf{W}(u)\,du = \mathbf{C}\mathbf{\Sigma}_{ee}^{1/2}\boldsymbol{\zeta}\,,$$

$$n^{-2}\sum_{t=1}^{n}\mathbf{Y}_t\mathbf{Y}_t' \Rightarrow \mathbf{C}\mathbf{\Sigma}_{ee}^{1/2}\int_0^1 \mathbf{W}(u)\mathbf{W}'(u)\,du\,\mathbf{\Sigma}_{ee}^{1/2}\mathbf{C}' = \mathbf{C}\mathbf{\Sigma}_{ee}^{1/2}\mathbf{G}\mathbf{\Sigma}_{ee}^{1/2}\mathbf{C}'\,,$$

$$n^{-2}\sum_{t=1}^{n}(\mathbf{Y}_t-\bar{\mathbf{y}}_n)(\mathbf{Y}_t-\bar{\mathbf{y}}_n)' \Rightarrow \mathbf{C}\mathbf{\Sigma}_{ee}^{1/2}(\mathbf{G}-\boldsymbol{\zeta}\boldsymbol{\zeta}')\mathbf{\Sigma}_{ee}^{1/2}\mathbf{C}'\,,$$

$$n^{-1}\sum_{t=1}^{n}\mathbf{Y}_{t-1}\mathbf{e}_t' \Rightarrow \mathbf{C}\mathbf{\Sigma}_{ee}^{1/2}\int_0^1 \mathbf{W}(u)\,d\mathbf{W}'(u)\,\mathbf{\Sigma}_{ee}^{1/2} = \mathbf{C}\mathbf{\Sigma}_{ee}^{1/2}\mathbf{\Upsilon}\mathbf{\Sigma}_{ee}^{1/2}\,,$$

$$n^{-1}\sum_{t=1}^{n}(\mathbf{Y}_{t-1}-\bar{\mathbf{y}}_n)\mathbf{e}_t' \Rightarrow \mathbf{C}\mathbf{\Sigma}_{ee}^{1/2}[\mathbf{\Upsilon}-\boldsymbol{\zeta}\mathbf{W}'(1)]\mathbf{\Sigma}_{ee}^{1/2}\,,$$

where $\mathbf{G}$, $\mathbf{\Upsilon}$, $\boldsymbol{\zeta}$, and $\mathbf{W}(u)$ are defined in Lemma 10.3.1.

Proof. Using the arguments in the proof of Theorem 10.1.2, the sums of squares and products of the $\mathbf{Y}_t$ can be approximated by the sums of squares and products of $\mathbf{C}\sum_{j=1}^{t}\mathbf{e}_j$. The results then follow from Theorem 5.3.7. ▲

10.3.2. Vector Process with a Single Unit Root

In this section we examine estimation of the parameters of a vector autoregressive process whose characteristic equation has a unit root. We begin by considering least squares estimation of the parameters for two special equations. The first is

$$Y_t = \theta_1 Y_{t-1} + \boldsymbol{\beta}'\mathbf{X}_{t-1} + e_t\,, \qquad t = 1, 2, \ldots, \tag{10.3.6}$$

where the e_t are independent $(0, \sigma_{ee})$ random variables with bounded $2+\delta$ $(\delta > 0)$ moments and Y_t is a scalar. Assume $E\{e_t \mid \mathscr{A}_{t-1}\} = 0$, where $\mathscr{A}_{t-1}$ is the sigma-field generated by $[(Y_{t-1}, \mathbf{X}_{t-1}, e_{t-1}), (Y_{t-2}, \mathbf{X}_{t-2}, e_{t-2}), \ldots, (Y_0, \mathbf{X}_0, e_0)]$. Assume that $\mathbf{X}_t$ is a zero mean stationary k-dimensional pth order autoregressive process satisfying

$$\mathbf{X}_t + \sum_{j=1}^{p}\mathbf{A}_j\mathbf{X}_{t-j} = \boldsymbol{\epsilon}_t \tag{10.3.7}$$

where the $\boldsymbol{\epsilon}_t$ are independent $(\mathbf{0}, \mathbf{\Sigma}_{\epsilon\epsilon})$ random variables with bounded $2+\delta$ $(\delta > 0)$ moments.

If $\theta_1 = 1$, then from (10.3.6),

$$Y_t = Y_0 + \sum_{j=1}^{t}(\boldsymbol{\beta}'\mathbf{X}_{j-1} + e_j)\,. \tag{10.3.8}$$

Also, from (10.3.7) and Theorem 2.8.1, $\boldsymbol{\beta}'\mathbf{X}_{t-1}$ has a representation as an infinite moving average

$$\boldsymbol{\beta}'\mathbf{X}_{t-1} = \boldsymbol{\beta}' \sum_{j=0}^{\infty} \mathbf{K}_j \boldsymbol{\epsilon}_{t-1-j} . \tag{10.3.9}$$

Hence, by the proof of Theorem 10.1.2,

$$Y_t \doteq Y_0 + \sum_{j=1}^{t} u_j , \tag{10.3.10}$$

where $u_t = \mathbf{c}_k' \boldsymbol{\epsilon}_{t-1} + e_t$ and $\mathbf{c}_k' = \boldsymbol{\beta}' \sum_{j=0}^{\infty} \mathbf{K}_j$. These properties of the processes can be used to obtain the limiting distribution of the least squares estimator of $(\theta_1, \boldsymbol{\beta}')$.

Theorem 10.3.1. Let the model (10.3.6)–(10.3.7) hold with $\theta_1 = 1$, where $(e_t, \boldsymbol{\epsilon}_t')$ is a sequence of independent zero mean vectors with common covariance matrix and bounded $2 + \delta$ $(\delta > 0)$ moments. Let $\hat{\boldsymbol{\theta}}$ be the least squares estimator of $(\theta_1, \boldsymbol{\beta}')'$, where

$$\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\boldsymbol{\beta}}')' = \left(\sum_{t=2}^{n} \mathbf{F}_t' \mathbf{F}_t \right)^{-1} \sum_{t=2}^{n} \mathbf{F}_t' Y_t , \tag{10.3.11}$$

and $\mathbf{F}_t = (Y_{t-1}, \mathbf{X}_{t-1}')$. Let

$$t_1 = [\hat{V}\{\hat{\theta}_1\}]^{-1/2} (\hat{\theta}_1 - 1) ,$$

where $\hat{V}\{\hat{\theta}_1\}$ is the least squares estimator of the variance of $\hat{\theta}_1$. Then

$$\hat{\theta}_1 - 1 = O_p(n^{-1}) ,$$

$$t_1 \xrightarrow{\mathcal{L}} \rho_{ue} \hat{\tau}_l + (1 - \rho_{ue}^2)^{1/2} d ,$$

and

$$n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \boldsymbol{\Gamma}_{XX}^{-1} \sigma_{ee}) ,$$

where $\boldsymbol{\Gamma}_{XX} = E\{\mathbf{X}_t \mathbf{X}_t'\}$, $\hat{\tau} \xrightarrow{\mathcal{L}} \hat{\tau}_l$, $\hat{\tau}$ is defined in Corollary 10.1.1.2, $\hat{\tau}_l$ has the limiting distribution of Corollary 10.1.1.2, u_t is defined in (10.3.10), ρ_{ue} is the correlation between u_t and e_t, d is a $N(0, 1)$ random variable, and $\hat{\tau}_l$ is independent of d. If $(\theta_1, \boldsymbol{\beta}') = (1, \mathbf{0})$, then

$$n(\hat{\theta}_1 - 1) \xrightarrow{\mathcal{L}} (2G)^{-1}(T^2 - 1) ,$$

where (G, T) is defined in Theorem 10.1.1.

Proof. From representation (10.3.10) and Theorem 10.1.1, we have

$$n^{-2} \sum_{t=2}^{n} Y_{t-1}^2 \xrightarrow{\mathcal{L}} \sigma_{uu} G .$$

Also

$$\left(\sum_{t=2}^{n} Y_{t-1}^2\right)^{-1} \sum_{t=2}^{n} Y_{t-1}e_t = O_p(n^{-1}),$$

because

$$V\left\{\sum_{t=2}^{n} Y_{t-1}e_t\right\} = \sum_{t=2}^{n} E\{Y_{t-1}^2\}\sigma_{ee} = O(n^2).$$

Following the arguments used in the proof of Theorem 10.1.2, we have

$$\begin{aligned} n^{-3/2}\sum_{t=2}^{n} Y_{t-1}\mathbf{X}'_{t-1} &= O_p(n^{-1/2}), \\ \left(n^{-1}\sum_{t=2}^{n} Y_{t-1}e_t, n^{-1/2}\sum_{t=2}^{n} \mathbf{X}'_{t-1}e_t\right) &= O_p(1). \end{aligned} \tag{10.3.12}$$

Thus,

$$\begin{aligned} \mathbf{D}_n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) &= \left(\mathbf{D}_n^{-1}\sum_{t=2}^{n} \mathbf{F}_t'\mathbf{F}_t\mathbf{D}_n^{-1}\right)^{-1} \mathbf{D}_n^{-1}\sum_{t=2}^{n} \mathbf{F}_t'e_t \\ &= \mathbf{B}_n^{-1}\mathbf{D}_n^{-1}\sum_{t=2}^{n} \mathbf{F}_t'e_t + O_p(n^{-1/2}), \end{aligned}$$

where

$$\mathbf{B}_n = \text{block diag}\left(n^{-2}\sum_{t=2}^{n} Y_{t-1}^2, \mathbf{\Gamma}_{XX}\right)$$

and $\mathbf{D}_n = \text{diag}(n, n^{1/2}, n^{1/2}, \ldots, n^{1/2})$. Therefore,

$$n(\hat{\theta}_1 - 1) = \left(n^{-2}\sum_{t=2}^{n} Y_{t-1}^2\right)^{-1} n^{-1}\sum_{t=2}^{n} Y_{t-1}e_t + O_p(n^{-1/2}) = O_p(1).$$

If $\boldsymbol{\beta} = \mathbf{0}$, then $Y_t = \Sigma_{j=1}^{t} e_j$,

$$n(\hat{\theta}_1 - 1) = \left[\sum_{t=2}^{n}\left(\sum_{j=1}^{t-1} e_j\right)^2\right]^{-1} \sum_{t=2}^{n}\left(\sum_{j=1}^{t-1} e_j\right)e_t + O_p(n^{-1/2}),$$

and

$$n(\hat{\theta}_1 - 1) \xrightarrow{\mathscr{L}} (2G)^{-1}(T^2 - 1).$$

The distributional result for $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ follows because $n^{-1/2}\Sigma_{t=2}^{n} \mathbf{X}_{t-1}e_t$ satisfies the conditions of Theorem 5.3.3 with $Z_{tn} = n^{-1/2}\mathbf{X}_{t-1}e_t$.

To establish the limiting result for t_1, we see that

$$\begin{aligned} t_1 &= \left[s_e^2 \left(\sum_{t=2}^{n} Y_{t-1}^2 \right)^{-1} \right]^{-1/2} \left(\sum_{t=2}^{n} Y_{t-1}^2 \right)^{-1} \sum_{t=2}^{n} Y_{t-1} e_t + O_p(n^{-1/2}) \\ &= \sigma_{ee}^{-1/2} \left[\sum_{t=2}^{n} \left(\sum_{j=1}^{t-1} u_j \right)^2 \right]^{-1/2} \sum_{t=2}^{n} \left(\sum_{j=1}^{t-1} u_j \right) e_t + o_p(1) . \end{aligned}$$

Now

$$\sum_{t=2}^{n} \left(\sum_{j=1}^{t-1} u_j \right) e_t = \sum_{t=2}^{n} \left(\sum_{j=1}^{t-1} u_j \right) (b_{eu} u_t + d_t) ,$$

where $b_{eu} = \sigma_{uu}^{-1} \sigma_{ue}$, and $d_t = e_t - b_{eu} u_t$ is uncorrelated with u_t. By Corollary 10.1.1.1,

$$\sigma_{ee}^{-1/2} b_{eu} \left[\sum_{t=2}^{n} \left(\sum_{j=1}^{t-1} u_j \right)^2 \right]^{-1/2} \sum_{t=2}^{n} \left(\sum_{j=1}^{t-1} u_j \right) u_t \xrightarrow{\mathscr{L}} \rho_{ue} \hat{\tau}_l .$$

By Lemma 10.3.1,

$$n^{-1} \sigma_{uu}^{-1/2} \sigma_{ee}^{-1/2} \sum_{t=2}^{n} Y_{t-1} d_t \xrightarrow{\mathscr{L}} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} 2(\gamma_i + \gamma_j)^{-1} \gamma_i^2 \gamma_j U_{ui} U_{dj} = Y_{ud} ,$$

where the limit normal random variables (U_{ui}, U_{di}) are defined for the process (u_t, d_t) by the expression following (10.1.12). Because u_t and d_t are uncorrelated, U_{ui} and U_{dj} are independent. Therefore, the conditional distribution of Y_{ud} conditional on $\{U_{u1}, U_{u2}, \ldots\}$ is normal with mean zero and variance

$$\sum_{j=1}^{\infty} \left[\sum_{i=1}^{\infty} 2(\gamma_i + \gamma_j)^{-1} \gamma_i^2 \gamma_j U_{ui} \right]^2 = \sum_{i=1}^{\infty} \gamma_i^2 U_{ui}^2 .$$

Also,

$$\sigma_{dd}^{-1/2} \left(\sum_{i=1}^{\infty} \gamma_i^2 U_{ui}^2 \right)^{-1/2} Y_{ud} \sim N(0, 1)$$

for almost every $\{U_{u1}, U_{u2}, \ldots\}$. It follows that $(\Sigma_{i=1}^{\infty} \gamma_i^2 U_{ui}^2)^{-1/2} Y_{ud}$ is independent of $\hat{\tau}_l$, and

$$\left(\sum_{t=2}^{n} Y_{t-1}^2 \right)^{-1/2} \sum_{t=2}^{n} Y_{t-1} d_t \xrightarrow{\mathscr{L}} \sigma_{dd}^{1/2} \xi , \qquad (10.3.13)$$

where ξ is a $N(0, 1)$ random variable independent of $\hat{\tau}_l$. ▲

The $\mathbf{X}_{t-1}$ of the model (10.3.6) is a rather general vector. The vector could contain current values of an explanatory variable and might contain variables of the form $\Delta Y_{t-i} = Y_{t-i} - Y_{t-i-1}$ for $i \geq 1$. If $\mathbf{X}_{t-1}$ contains only ΔY_{t-i}, $i =$

$1, 2, \ldots, p-1$, then we obtain Theorem 10.1.2 as a special case of Theorem 10.3.1.

In Theorem 10.3.1, the model contains no intercept. If the model contains a constant term, then the pivotal statistic for the coefficient of lagged Y is a linear combination of $\hat{\tau}_\mu$ of Table 10.A.2 and a $N(0, 1)$ random variable.

Note that the usual regression statistics are appropriate, in large samples, for testing and confidence statements concerning the β-parameters. The usual pivotal statistic for the coefficient for Y_{t-1} has the distribution of Table 10.A.2 if $\rho_{ue} = 0$. This can happen if the only elements of $\mathbf{X}_{t-1}$ with nonzero coefficients are elements of the form ΔY_{t-j}, $j \geq 1$. If one is interested in testing $\rho = 1$ in a model such as (10.3.6), the statistic $\hat{\lambda}_k$ of Theorem 10.3.4 can be computed for the vector process $(Y_t, \mathbf{X}_t')$.

A special case of the model (10.3.6) is that in which Y_t is a function of a vector, say $\mathbf{X}_{1,t-1}$, and an error that is autoregressive with parameter θ_1. The model can be written

$$
\begin{aligned}
Y_t &= \boldsymbol{\beta}_1'\mathbf{X}_{1,t-1} + a_t, \\
a_t &= \theta_1 a_{t-1} + e_t,
\end{aligned}
$$

or as

$$Y_t = \boldsymbol{\beta}_1'\mathbf{X}_{1,t-1} - \theta_1\boldsymbol{\beta}_1'\mathbf{X}_{1,t-2} + \theta_1 Y_{t-1} + e_t .$$

Under rather general conditions for $\mathbf{X}_{1,t}$ and $\theta_1 = 1$, the nonlinear least squares estimator $(\hat{\boldsymbol{\beta}}_1', \hat{\theta}_1)$ satisfies

$$
\begin{aligned}
n^{1/2}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) &\xrightarrow{\mathscr{L}} N(\mathbf{0}, \mathbf{A}_{XX}^{-1}), \\
[\hat{V}\{\hat{\theta}_1\}]^{-1/2}(\hat{\theta}_1 - 1) &\xrightarrow{\mathscr{L}} \hat{\tau}_t,
\end{aligned}
$$

where $\hat{V}\{\hat{\theta}_1\}$ is the usual nonlinear least squares estimator of the variance of $\hat{\theta}_1$, and

$$\mathbf{A}_{XX} = E\{(\mathbf{X}_{t-1} - \theta_1\mathbf{X}_{t-2})(\mathbf{X}_{t-1} - \theta_1\mathbf{X}_{t-2})'\}.$$

We next consider an equation in which one of the explanatory variables satisfies an autoregressive equation with a unit root.

Theorem 10.3.2. Let

$$X_t = \beta_1 Y_{t-1} + \boldsymbol{\beta}_2'\mathbf{Z}_{t-1} + e_t, \tag{10.3.14}$$

$$Y_t + \sum_{j=1}^{p} \alpha_j Y_{t-j} = u_t, \tag{10.3.15}$$

$$\mathbf{Z}_t + \sum_{j=1}^{r} \mathbf{B}_j \mathbf{Z}_{t-j} = \mathbf{v}_t , \tag{10.3.16}$$

where $(e_t, \mathbf{v}_t', u_t)$ is a sequence of independent zero mean random vectors with covariance matrix $\mathbf{\Sigma}$ and bounded $2+\delta$ $(\delta > 0)$ moments. Let one of the roots of

$$m^p + \sum_{j=1}^{p} \alpha_j m^{p-j} = 0$$

be one, and let the remaining $p-1$ roots be less than one in absolute value. Let $\mathbf{Z}_t$ be a stationary process where the roots of

$$\left| \mathbf{I}m^r + \sum_{j=1}^{r} \mathbf{B}_j m^{r-j} \right| = 0$$

are less than one in absolute value. Let $\boldsymbol{\beta} = (\beta_1, \boldsymbol{\beta}_2')'$, and let

$$\hat{\boldsymbol{\beta}} = \left(\sum_{t=2}^{n} \mathbf{F}_{t-1}' \mathbf{F}_{t-1} \right)^{-1} \sum_{t=2}^{n} \mathbf{F}_{t-1}' X_t ,$$

where $\mathbf{F}_{t-1} = (Y_{t-1}, \mathbf{Z}_{t-1}')$. Let

$$t_{\beta 1} = [\hat{V}\{\hat{\beta}_1\}]^{-1/2} (\hat{\beta}_1 - \beta_1) ,$$

where $\hat{V}\{\hat{\beta}_1\}$ is the least squares estimator of the variance of $\hat{\beta}_1$. Then

$$\begin{aligned} \hat{\beta}_1 - \beta_1 &= O_p(n^{-1}) , \\ n^{1/2}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) &\xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{\Gamma}_{ZZ}^{-1} \sigma_{ee}) , \\ t_{\beta 1} &\xrightarrow{\mathcal{L}} \rho_{ue} \hat{\tau}_l + (1 - \rho_{ue}^2)^{1/2} d , \end{aligned} \tag{10.3.17}$$

where $\mathbf{\Gamma}_{ZZ} = E\{\mathbf{Z}_t \mathbf{Z}_t'\}$, $\hat{\tau} \xrightarrow{\mathcal{L}} \hat{\tau}_l$, $\hat{\tau}$ is defined in Corollary 10.1.1.2, ρ_{ue} is the correlation between u_t and e_t, u_t is defined in (10.3.15), d is a $N(0, 1)$ random variable, and $\hat{\tau}_l$ is independent of d.

Proof. From the proof of Theorem 10.1.2,

$$n^{-2} \sum_{t=2}^{n} Y_{t-1}^2 \xrightarrow{\mathcal{L}} c^2 \sigma_{uu} G ,$$

where $c = [1 + \Sigma_{i=1}^{p} (i-1)\alpha_i]^{-1}$, G is defined in Theorem 10.1.1, and

$$V\left\{ n^{-1} \sum_{t=2}^{n} Y_{t-1} e_t \right\} = n^{-2} \sum_{t=2}^{n} E\{Y_{t-1}^2\} \sigma_{ee} = O(1) .$$

The error in the least squares estimator is

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\sum_{t=2}^{n} \mathbf{F}'_{t-1}\mathbf{F}_{t-1} \right)^{-1} \sum_{t=2}^{n} \mathbf{F}'_t e_t$$
$$= \mathbf{B}_n^{-1}\mathbf{D}_n^{-1} \sum_{t=2}^{n} \mathbf{F}'_{t-1} e_t + O_p(n^{-1/2}),$$

where $\mathbf{D}_n = \operatorname{diag}(n, n^{1/2}, n^{1/2}, \ldots, n^{1/2})$,

$$\boldsymbol{B}_n = \text{block diag}\left(n^{-2} \sum_{t=2}^{n} Y_{t-1}^2, \boldsymbol{\Gamma}_{ZZ} \right),$$

and we have used

$$\left(n^{-1} \sum_{t=2}^{n} Y_{t-1}\mathbf{Z}_{t-1}, \sum_{t=2}^{n} \mathbf{D}_n^{-1}\mathbf{F}'_{t-1} e_t \right) = O_p(1),$$
$$n^{-1} \sum_{t=2}^{n} \mathbf{Z}_{t-1}\mathbf{Z}'_{t-1} - \boldsymbol{\Gamma}_{ZZ} = O_p(n^{-1/2}).$$

Thus,

$$n(\hat{\beta}_1 - \beta_1) = \left(n^{-2} \sum_{t=2}^{n} Y_{t-1}^2 \right)^{-1} n^{-1} \sum_{t=2}^{n} Y_{t-1} e_t + O_p(n^{-1/2}) = O_p(1).$$

The limiting distribution of $n^{-1/2} \sum_{t=2}^{n} \mathbf{Z}_{t-1} e_t$ follows from Theorem 5.3.3.

The derivation of the limiting distribution of $t_{\beta 1}$ follows that for t_1 of Theorem 10.3.1. ▲

Theorem 10.3.2 identifies the difficulties in making inferential statements about parameters of equations such as (10.3.15) where the explanatory variable is a time series. In Chapter 9 we assumed the explanatory variables to be fixed sequences. An operationally equivalent assumption is that the explanatory variables form a random sequence independent of the error sequence. In such cases, the t-statistics are $N(0, 1)$ random variables in the limit. If an explanatory variable is an autoregressive process, the distribution of the t-statistics depends on the roots of the process and on the correlation properties of the error processes. If the autoregressive process defining the explanatory variable has a unit root and if the error in the autoregressive process is correlated with the error in the equation, then by Theorem 10.3.2 the usual t-statistics have nonstandard limiting distributions. If the explanatory process is stationary, then the t-statistics for the parameters of the model (10.3.6) are $N(0, 1)$ random variables in the limit.

If it is known that the explanatory process has a unit root, then imposing that condition on the system (10.3.14), (10.3.15) produces efficient estimates of the parameters of interest, and the t-statistics converge in distribution to $N(0, 1)$ random variables. See Phillips (1991) and Exercise 10.12. However, in many practical situations it is not known that the Y_t-process of equation (10.3.15) has a unit root. In such situations, the investigator must recognize the potential effect of a unit root in the Y_t-process on inferences about β_1.

Example 10.3.1. In this example, we consider estimation for the model

$$X_t = \beta_0 + \beta_1 Y_{t-1} + e_t ,$$
$$Y_t = \theta_0 + \theta_1 Y_{t-1} + \theta_2 \Delta Y_{t-1} + u_t ,$$

where $(e_t, u_t)' \sim \mathrm{NI}(\mathbf{0}, \mathbf{\Sigma})$. Our situation is assumed to be that described in Section 10.1.3 in that we specify $\theta_1 \in (-1, 1]$ and $\theta_0 = 0$ if $\theta_1 = 1$. The data of Table 10.B.4 were artificially generated to satisfy the model. The ordinary least squares estimates of the two equations are

$$\begin{aligned} \hat{X}_t &= \underset{(0.099)}{12.992} + \underset{(0.0075)}{0.7783}\, y_{t-1} , \\ \hat{y}_t &= \underset{(0.0076)}{0.9798}\, y_{t-1} + \underset{(0.0597)}{0.7869}\, \Delta Y_{t-1} , \end{aligned}$$

where $y_t = Y_t - \bar{y}_n$ and the ordinary least squares standard errors are given in parentheses. The estimated covariance matrix of (e_t, u_t) is

$$\hat{V}\{(e_t, u_t)'\} = \begin{pmatrix} 0.9618 & 0.7291 \\ 0.7291 & 0.9795 \end{pmatrix} ,$$

and the estimated correlation between e_t and u_t is $\hat{\rho}_{12} = 0.751$. The weighted least squares estimate of θ_1 is $\hat{\theta}_{w1} = 0.9850$, and the test for a unit root is $\hat{\tau}_{w1} = -1.94$. The estimator of θ_1 constructed from (10.1.55) is $\tilde{\theta}_{w1} = 0.9921$.

To estimate the parameters of the X-equation, we use a system method of estimation such as full information maximum likelihood or three-stage least squares, available in SAS/ETS®. The two equations are estimated simultaneously, subject to the restriction that $\theta_1 = 0.9921$. The standard errors computed by the program should not be used as the standard errors of the estimators. They are only appropriate if it is known that $\theta_1 = 0.9921$. We suggest that the standard errors from the initial ordinary least squares regression be modified as in Example 10.1.3. Thus, the standard error for y_{t-1} in the regression of y_t on y_{t-1} is increased by 15%, and the standard error for y_{t-1} in the regression of X_t on $(1, y_{t-1})$ is multiplied by $1 + 0.15\hat{\rho}_{12}$. Our estimate of the model is

$$\begin{aligned} \hat{X}_t &= \underset{(0.099)}{12.937} + \underset{(0.0083)}{0.7876}\, y_{t-1} , \\ \hat{y}_t &= \underset{(0.0087)}{0.9921}\, y_{t-1} + \underset{(0.0597)}{0.7308}\, \Delta Y_{t-1} , \end{aligned}$$

where the numbers in parentheses are the modified ordinary least squares standard errors. ▲▲

We now investigate estimation for the vector autoregressive process. Let $\mathbf{Y}_t$ be a pth order autoregressive process of dimension k. Let

$$\mathbf{Y}_t + \sum_{i=1}^{p} \mathbf{A}_i \mathbf{Y}_{t-i} = \mathbf{e}_t, \qquad t = 1, 2, \ldots, \tag{10.3.18}$$

where the $\mathbf{e}_t$ are independent and identically distributed $(\mathbf{0}, \mathbf{\Sigma}_{ee})$ random vectors, and the $\mathbf{A}_i$ are $k \times k$ parameter matrices. The representation (10.3.18) was introduced in Section 2.8. Assume that the characteristic equation

$$\left| \mathbf{I}m^p + \sum_{i=1}^{p} \mathbf{A}_i m^{p-i} \right| = 0 \tag{10.3.19}$$

has g roots equal to one and that all remaining roots are less than one in absolute value. By analogy to (10.1.25), we can write

$$\mathbf{Y}_t = \mathbf{H}_1 \mathbf{Y}_{t-1} + \sum_{i=2}^{p} \mathbf{H}_i \, \Delta \mathbf{Y}_{t-i+1} + \mathbf{e}_t, \tag{10.3.20}$$

where $\Delta \mathbf{Y}_t = \mathbf{Y}_t - \mathbf{Y}_{t-1}$, $\mathbf{H}_1 = -\Sigma_{j=1}^{p} \mathbf{A}_j$, $\mathbf{H}_p = \mathbf{A}_p$, and

$$\mathbf{H}_{p-i} = \sum_{j=p-i}^{p} \mathbf{A}_j \tag{10.3.21}$$

for $i = 2, 3, \ldots, p-1$. We will study the process in which the portion of the Jordan canonical form of $\mathbf{H}_1$ associated with the unit roots is a g-dimensional identity matrix.

By Theorem 2.4.2, there exists a $\mathbf{Q}$ such that

$$\mathbf{Q}^{-1}\mathbf{H}_1\mathbf{Q} = \mathbf{\Lambda}, \tag{10.3.22}$$

where the elements of $\mathbf{\Lambda}$ are determined by the roots of $\mathbf{H}_1$. If the roots of $\mathbf{H}_1$ are distinct, $\mathbf{\Lambda}$ is a diagonal matrix with the roots on the diagonal. If $\mathbf{H}_1$ has repeated roots, the matrix $\mathbf{\Lambda}$ is of the form described in Theorem 2.4.2. Assume that $\mathbf{\Lambda}$ is block diagonal with the $g \times g$ identity matrix as the first block and $\mathbf{\Lambda}_{22}$ as the second block, where $\mathbf{\Lambda}_{22} - \mathbf{I}$ is nonsingular.

Let $\mathbf{T}$ be a nonsingular matrix whose first g rows are the first g rows of $\mathbf{Q}^{-1}$, the rows associated with the roots of one. The remaining rows of $\mathbf{T}$ are linear combinations of the rows of $\mathbf{Q}^{-1}$ chosen such that all elements of $\mathbf{T}$ are real. The matrix $\mathbf{T}$ can be chosen so that the upper left $g \times g$ matrix of $\mathbf{T}\mathbf{\Sigma}_{ee}\mathbf{T}'$ is an identity matrix. Let $\mathbf{X}_t = \mathbf{T}\mathbf{Y}_t$. Then

$$\mathbf{X}_t = \mathbf{\Phi}_1 \mathbf{X}_{t-1} + \sum_{i=2}^{p} \mathbf{\Phi}_i \, \Delta \mathbf{X}_{t-i+1} + \mathbf{a}_t , \tag{10.3.23}$$

where $\mathbf{\Phi}_i = \mathbf{T}\mathbf{H}_i\mathbf{T}^{-1}$, $i = 1, 2, \ldots, p$, and $\mathbf{a}_t = \mathbf{T}\mathbf{e}_t$. The matrix $\mathbf{\Phi}_1$ has a $g \times g$ identity matrix as the upper left block. The remaining elements of the first g rows and first g columns of $\mathbf{\Phi}_1$ are zero. It follows that the vector process

$$\overset{*}{\mathbf{X}}_t = (X_{1t} - X_{1,t-1}, X_{2t} - X_{2,t-1}, \ldots, X_{gt} - X_{g,t-1}, X_{g+1,t}, X_{g+2,t}, \ldots, X_{kt})'$$

is a vector autoregressive process with all roots less than one in absolute value.

A vector process $\mathbf{Y}_t$ for which there exists a transformation $\mathbf{T}$ such that $k - g$ components of the transformed vector are stationary and g components are unit root processes is called *cointegrated of order* $k - g$ in the econometrics literature. See Granger (1981) and Engle and Granger (1987). Granger's original definition of cointegration required every element of $\mathbf{Y}_t$ to contain a unit root. The last $k - g$ rows of $\mathbf{T}$ that define stationary processes are called *cointegrating vectors.*

A canonical form for the pth order process of interest is

$$\mathbf{Y}_t = \mathbf{H}_1 \mathbf{Y}_{t-1} + \sum_{i=2}^{p} \mathbf{H}_i \, \Delta \mathbf{Y}_{t-i+1} + \mathbf{e}_t , \tag{10.3.24}$$

where $(\mathbf{Y}'_{1t}, \mathbf{Y}'_{2t}) = \mathbf{Y}'_t$, $\mathbf{Y}_{1t}$ is the vector composed of the first g elements of the $\mathbf{Y}_t$, the upper left $g \times g$ portion of $\mathbf{\Sigma}_{ee}$ is $\mathbf{I}_g$, $\mathbf{H}_1 = \text{block diag}\{\mathbf{I}_g, \mathbf{H}_{1,22}\}$, and $\mathbf{H}_{1,22} - \mathbf{I}_{k-g}$ is nonsingular. The conformable partition of $\mathbf{e}_t$ of (10.3.24) is $(\mathbf{e}'_{1t}, \mathbf{e}'_{2t})'$, where $\mathbf{e}_{1t}$ is a g-dimensional vector. We can write

$$\mathbf{e}_{2t} = \mathbf{e}'_{1t}\mathbf{\Sigma}_{ee12} + \mathbf{a}_{2t} , \tag{10.3.25}$$

where $\mathbf{\Sigma}_{ee12}$ is the $g \times (k - g)$ upper right part of $\mathbf{\Sigma}_{ee}$, $\mathbf{a}_{2t}$ is uncorrelated with $\mathbf{e}_{1t}$ and

$$E\{\mathbf{a}_{2t}\mathbf{a}'_{2t}\} = \mathbf{\Sigma}_{aa22} . \tag{10.3.26}$$

The vector $(\Delta\mathbf{Y}'_{1t}, \mathbf{Y}'_{2t})$ is a stationary process and can be expressed as

$$(\Delta\mathbf{Y}'_{1t}, \mathbf{Y}'_{2t})' = \sum_{j=0}^{\infty} \mathbf{K}_j \mathbf{e}_{t-j} \tag{10.3.27}$$

by Theorem 2.8.1. Also,

$$\mathbf{Y}_{1t} = \sum_{i=1}^{t} \sum_{j=2}^{p} \mathbf{H}_{j,11} \, \Delta \mathbf{Y}_{1,i-j+1} + \sum_{i=1}^{t} \sum_{j=2}^{p} \mathbf{H}_{j,12} \, \Delta \mathbf{Y}_{2,i-j+1} + \sum_{i=1}^{t} \mathbf{e}_{1i} ,$$

where $\mathbf{H}_{j,11}$ is the upper left $g \times g$ block of $\mathbf{H}_j$ and $\mathbf{H}_{j,12}$ is the upper right $g \times (k - g)$ block of $\mathbf{H}_j$. Because $\Sigma_{i=1}^{t} \Delta\mathbf{Y}_{2i} = \mathbf{Y}_{2t} - \mathbf{Y}_{21}$ and $\Delta\mathbf{Y}_{1t}$ is stationary, we have

$$\mathbf{Y}_{1t} = \sum_{j=2}^{p} \mathbf{H}_{j,11}\mathbf{Y}_{1t} + \sum_{i=1}^{t} \mathbf{e}_{1i} + O_p(1)$$

and

$$\begin{aligned}\mathbf{Y}_{1t} &= \left(\mathbf{I} - \sum_{j=2}^{p} \mathbf{H}_{j,11}\right)^{-1} \sum_{i=1}^{t} \mathbf{e}_{1i} + O_p(1)\,, \\ &\overset{\text{def}}{=} \mathbf{C}_g \sum_{i=1}^{t} \mathbf{e}_{1i} + O_p(1)\,, \end{aligned} \tag{10.3.28}$$

where the $O_p(1)$ remainder is uncorrelated with $\mathbf{e}_s$ for $s > t$.

Let the ordinary least squares estimators of the parameters of the model (10.3.24) be

$$\hat{\mathbf{H}}' = \left(\sum_{t=p+1}^{n} \mathbf{L}'_{t-1}\mathbf{L}_{t-1}\right)^{-1} \sum_{t=p+1}^{n} \mathbf{L}'_{t-1}\mathbf{Y}'_t \tag{10.3.29}$$

and

$$\hat{\mathbf{\Sigma}}_{ee} = (n-p-pk)^{-1} \sum_{t=p+1}^{n} \hat{\mathbf{e}}_t\hat{\mathbf{e}}'_t\,, \tag{10.3.30}$$

where

$$\begin{aligned}\mathbf{L}_{t-1} &= (\mathbf{Y}'_{t-1}, \Delta\mathbf{Y}'_{t-1}, \ldots, \Delta\mathbf{Y}'_{t-p+1})\,, \\ \mathbf{H}' &= (\mathbf{H}_1, \mathbf{H}_2, \ldots, \mathbf{H}_p)'\,,\end{aligned}$$

and $\hat{\mathbf{e}}'_t = \mathbf{Y}'_t - \mathbf{L}_{t-1}\hat{\mathbf{H}}'$. The ordinary least squares estimator of the "covariance matrix" of vec $\hat{\mathbf{H}}'$ is

$$\hat{\mathbf{\Sigma}}_{ee} \otimes \left(\sum_{t=p+1}^{n} \mathbf{L}'_{t-1}\mathbf{L}_{t-1}\right)^{-1}. \tag{10.3.31}$$

This matrix will normalize the estimators, but it is not an estimated covariance matrix in the traditional sense because some of its elements cannot be standardized to converge to constants. This is analogous to the results for the univariate case discussed in Section 10.1. The matrix $\mathbf{V} = \sum_{t=p+1}^{n} \mathbf{L}'_{t-1}\mathbf{L}_{t-1}$ is $kp \times kp$, and we denote the ijth $k \times k$ block of the matrix by $\mathbf{V}_{ij}$.

From the derivation of the canonical form (10.3.24), we know that the least squares estimator is invariant to linear transformations in the sense that if $\mathbf{T}$ is nonsingular, then the least squares estimators of the coefficients for the transformed model based on $\mathbf{TY}_t$ are $\mathbf{T}\hat{\mathbf{H}}_i\mathbf{T}^{-1}$, $i = 1, 2, \ldots, p$, where $\hat{\mathbf{H}}_i$, $i = 1, 2, \ldots, p$, are the least squares estimators for the original model. Also, the estimator of the error covariance matrix for the transformed problem is $\mathbf{T}\hat{\mathbf{\Sigma}}_{ee}\mathbf{T}'$, where $\hat{\mathbf{\Sigma}}_{ee}$ is the regression residual mean square for the original problem.

We now give the limiting distribution of the estimators for the canonical model.

Theorem 10.3.3. Let $\mathbf{Y}_t$ be a k-dimensional pth order autoregressive process written in the canonical form (10.3.24). Assume the $\mathbf{e}_t$ are iid$(\mathbf{0}, \mathbf{\Sigma}_{ee})$ random variables or that the $\mathbf{e}_t$ are martingale differences satisfying the conditions of Lemma 10.3.1. Partition $\mathbf{Y}'_t$ as $(\mathbf{Y}'_{1t}, \mathbf{Y}'_{2t})$, where $\mathbf{Y}_{1t}$ is the g-dimensional vector with g unit roots. Assume that $(\Delta\mathbf{Y}'_{1t}, \mathbf{Y}'_{2t})$ is a stationary process with $E\{\Delta\mathbf{Y}_{1t}\} = \mathbf{0}$ and $\mathbf{Y}_{1,0} = \mathbf{0}$. Let the ordinary least squares estimators be defined by (10.3.29) and (10.3.30). Then

$$\mathbf{D}_n(\hat{\mathbf{H}}' - \mathbf{H}') \xrightarrow{\mathscr{L}} \begin{pmatrix} \mathbf{C}_g^{-1\prime}\mathbf{G}_{gg}^{-1}(\mathbf{Y}_{11}, \mathbf{Y}_{11}\mathbf{\Sigma}_{ee12} + \mathbf{Y}_{12}\mathbf{\Sigma}_{aa22}^{1/2}) \\ \mathbf{\Psi}_{2.} \end{pmatrix},$$

where $\mathbf{C}_g = (\mathbf{I} - \Sigma_{j=2}^{p} \mathbf{H}_{j,11})^{-1}$ is defined in (10.3.28), $\mathbf{a}_{2t}$ is defined in (10.3.25),

$$\mathbf{D}_n = \operatorname{diag}(n, n, \ldots, n, n^{1/2}, n^{1/2}, \ldots, n^{1/2}),$$

$(\mathbf{G}, \mathbf{Y})$ is defined in Lemma 10.3.1, $\mathbf{G}_{gg}$ is the upper left $g \times g$ portion of $\mathbf{G}$, $\mathbf{Y}_{1.} = (\mathbf{Y}_{11}, \mathbf{Y}_{12})$ is the matrix composed of the first g rows of $\mathbf{Y}$, the elements of $\mathbf{\Psi}_{2.}$ are zero mean normal random variables with

$$V\{\operatorname{vec} \mathbf{\Psi}_{2.}\} = \mathbf{\Sigma}_{ee} \otimes \mathbf{\Sigma}_{LL22}^{-1},$$

and

$$\mathbf{\Sigma}_{LL22} = E\{(\mathbf{Y}'_{2,t-1}, \Delta\mathbf{Y}'_{t-1}, \ldots, \Delta\mathbf{Y}'_{t-p+1})'(\mathbf{Y}'_{2,t-1}, \Delta\mathbf{Y}'_{t-1}, \ldots, \Delta\mathbf{Y}'_{t-p+1})\}.$$

Also, $\hat{\mathbf{\Sigma}}_{ee}$ converges in probability to $\mathbf{\Sigma}_{ee}$.

Proof. By (10.3.28) and by the arguments used in the proof of Theorem 10.1.2,

$$\mathbf{D}_n^{-1} \sum_{t=p+1}^{n} \mathbf{L}'_{t-1}\mathbf{L}_{t-1}\mathbf{D}_n^{-1} \xrightarrow{\mathscr{L}} \begin{pmatrix} \mathbf{C}_g\mathbf{G}_{gg}\mathbf{C}'_g & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{LL22} \end{pmatrix}.$$

Likewise, by Theorem 5.3.7 and Theorem 8.2.3,

$$\mathbf{D}_n^{-1} \sum_{t=p+1}^{n} \mathbf{L}'_{t-1}\mathbf{e}'_t \xrightarrow{\mathscr{L}} \begin{pmatrix} \mathbf{C}_g(\mathbf{Y}_{11}, \mathbf{Y}_{11}\mathbf{\Sigma}_{ee12} + \mathbf{Y}_{12}\mathbf{\Sigma}_{aa11}^{1/2}) \\ \mathbf{\Phi}_{2.} \end{pmatrix},$$

where $\mathbf{\Phi}_{2.}$ is a matrix of zero mean normal random variables with covariance matrix

$$V\{\operatorname{vec} \mathbf{\Phi}_{2.}\} = \mathbf{\Sigma}_{ee} \otimes \mathbf{\Sigma}_{LL22}.$$

The covariance between $n^{-1/2} \Sigma\, \Delta\mathbf{Y}_{t-j}\, \mathbf{e}'_t$, $j \geq 1$, and $n^{-1} \Sigma\, \mathbf{Y}_{t-1}\mathbf{e}'_t$ goes to zero as n increases and the limit random variables are independent. See Lai and Wei (1985a) and Chan and Wei (1988). Using

$$\mathbf{D}_n(\hat{\mathbf{H}}' - \mathbf{H}') = \left(\mathbf{D}_n^{-1} \sum_{t=p+1}^{n} \mathbf{L}_{t-1}'\mathbf{L}_{t-1}\mathbf{D}_n^{-1}\right)^{-1} \mathbf{D}_n^{-1} \sum_{t=p+1}^{n} \mathbf{L}_{t-1}'\mathbf{e}_t',$$

we obtain the limiting distribution.

The estimated covariance matrix $\hat{\mathbf{\Sigma}}_{ee}$ converges to $\mathbf{\Sigma}_{ee}$ because

$$\begin{aligned}[n-(p+1)k]^{-1} &\sum_{t=p+1}^{n} (\mathbf{Y}_t' - \mathbf{L}_{t-1}\hat{\mathbf{H}}')'(\mathbf{Y}_t' - \mathbf{L}_{t-1}\hat{\mathbf{H}}') \\ &= [n-(p+1)k]^{-1} \sum_{t=p+1}^{n} \mathbf{e}_t\mathbf{e}_t' + O_p(n^{-1}).\end{aligned} \qquad \blacktriangle$$

Theorem 10.3.3 gives the distribution for the parameters of the canonical form (10.3.24). The coefficient matrices for any other representation are linear combinations of the elements of the canonical coefficient matrices. Hence, the limiting distributions are defined in terms of those of Theorem 10.3.3. As in the univariate case, the limiting distribution of coefficients associated with unit roots differ from the limiting distribution of coefficients associated with the stationary portion of the model.

The following corollary to Theorem 10.3.3 was derived for $g = 1$ by Fountis and Dickey (1989).

Corollary 10.3.3. Let $\mathbf{Y}_t$ be a k-dimensional pth order autoregressive process with g unit roots and all other roots less than one in absolute value. Assume the process can be transformed to the canonical form (10.3.24) with g unit roots. Let the assumptions on $\mathbf{e}_t$ of Theorem 10.3.3 hold. Let $\hat{m}_1 \geq \hat{m}_2 \geq \cdots \geq \hat{m}_g$ be the g roots of

$$|\mathbf{I}m^p + \hat{\mathbf{A}}_1 m^{p-1} + \cdots + \hat{\mathbf{A}}_p| = 0,$$

with largest absolute values, where the $\hat{\mathbf{A}}_i$, $i = 1, 2, \ldots, p$, are the least squares estimators. Then the limiting distribution of $n(\hat{m}_i - 1)$, $i = 1, 2, \ldots, g$, is the distribution of the g roots of

$$|\mathbf{I}m - \mathbf{Y}_{1.}'\mathbf{G}^{-1}| = 0,$$

where $(\mathbf{G}, \mathbf{Y})$ is the $g \times 2g$ matrix defined in Lemma 10.3.1.

Proof. The roots of the characteristic equation are unchanged by nonsingular transformations and we assume, without loss of generality, that the equation is in the canonical form (10.3.24).

By Theorem 10.3.3, the least squares coefficients are converging to the true values. Therefore, by the results of Section 5.8, the roots of the characteristic equation are converging to the true roots. We expand the largest roots about one and write $\hat{m}^r = 1 + r\delta + o_p(\delta)$. Then the local approximation to the determinantal equation becomes

$$|\mathbf{I}(1+p\delta)+\hat{\mathbf{A}}_1[1+(p-1)\delta]+\hat{\mathbf{A}}_2[1+(p-2)\delta]+\cdots+\hat{\mathbf{A}}_p|=0.$$

Using the definitions of $\hat{\mathbf{H}}_i$ in terms of the $\hat{\mathbf{A}}_i$, we have

$$\left|[1+(p-1)\delta](\mathbf{I}-\hat{\mathbf{H}}_1)+\left(\mathbf{I}-\sum_{j=2}^{p}\hat{\mathbf{H}}_j\right)\delta\right|=0. \tag{10.3.32}$$

The lower right portion of $\mathbf{I}-\hat{\mathbf{H}}_1$ is converging to a nonsingular matrix. The matrix $\mathbf{I}-\Sigma_{j=2}^{p}\hat{\mathbf{H}}_j$ is converging to a matrix, say $\mathbf{M}$, with $\mathbf{C}_g^{-1}$ as the upper left $g\times g$ block. Therefore, we multiply the matrix in (10.3.32) on the left by the nonsingular matrix

$$\begin{pmatrix}\mathbf{C}_g & \mathbf{0}\\ -\mathbf{M}_{21}\mathbf{C}_g & \mathbf{I}\end{pmatrix},$$

where $\mathbf{M}_{21}$ is the lower left $(k-g)\times g$ portion of $\mathbf{M}$. The values of δ that satisfy the resulting determinantal equation are equal to the values that satisfy (10.3.32). Using the fact that the first g columns of $\hat{\mathbf{H}}_1-\mathbf{I}$ are $O_p(n^{-1})$, the g smallest roots of (10.3.32) are converging to the g smallest roots of

$$|\mathbf{C}_g(\hat{\mathbf{H}}_{1,11}-\mathbf{I}_g)-\mathbf{I}_g\delta|=0.$$

The result follows from the limiting distribution of $\hat{\mathbf{H}}_{1,11}-\mathbf{I}_g$ given in Theorem 10.3.3. ▲

We now turn to the problem of estimating the parameters of (10.3.18) subject to the restriction that one of the roots of (10.3.19) is equal to one. If we write the model in the form (10.3.20), the determinant of $\mathbf{H}_1-\mathbf{I}$ is zero under the null model and there exists a vector $\boldsymbol{\kappa}_1$, with $\boldsymbol{\kappa}_1'\boldsymbol{\kappa}_1\neq 0$, such that

$$\boldsymbol{\kappa}_1'(\mathbf{H}_1-\mathbf{I})=\mathbf{0}. \tag{10.3.33}$$

Therefore, if the first observations are treated as fixed, the maximum likelihood estimator of $(\mathbf{H}_1,\ldots,\mathbf{H}_p)$ is obtained by imposing the restriction (10.3.33) on the least squares estimator. Assume the first element of $\boldsymbol{\kappa}_1$, denoted by κ_{11}, is not zero, and let $\kappa_{11}^{-1}\boldsymbol{\kappa}_1'=(1,-\mathbf{B}_{0,12})$. If we multiply (10.3.20) by $\kappa_{11}^{-1}\boldsymbol{\kappa}_1'$, we have

$$\Delta Y_{1t}=\mathbf{B}_{0,12}\,\Delta\mathbf{Y}_{2t}+\sum_{i=1}^{p-1}\mathbf{B}_{i+1,1.}\,\Delta\mathbf{Y}_{t-i}+v_{1t}, \tag{10.3.34}$$

where $\mathbf{B}_{i+1,1.}=\kappa_{11}^{-1}\boldsymbol{\kappa}_1'\mathbf{H}_{i+1}$, $\mathbf{Y}_t'=(Y_{1t},\mathbf{Y}_{2t}')$, and $v_{1t}=\kappa_{11}^{-1}\boldsymbol{\kappa}_1'\mathbf{e}_t$.

Let $\boldsymbol{\theta}=(\mathbf{B}_{0,12},\mathbf{B}_{2,1.},\ldots,\mathbf{B}_{p,1.})$. Let $\hat{\boldsymbol{\pi}}=(\hat{\boldsymbol{\pi}}_1,\hat{\boldsymbol{\pi}}_2,\hat{\boldsymbol{\pi}}_3)$ be the matrix of regression coefficients obtained in the regression of $[\Delta Y_{1t},\Delta\mathbf{Y}_{2t}',(\Delta\mathbf{Y}_{t-1}',\ldots,\Delta\mathbf{Y}_{t-p+1}')]$ on $(\mathbf{Y}_{t-1}',\Delta\mathbf{Y}_{t-1}',\ldots,\Delta\mathbf{Y}_{t-p+1}')$. Of course, the elements of $\hat{\boldsymbol{\pi}}_3$ are zeros and ones. Let $\hat{\lambda}_k$ be the smallest root of

$$\left| \hat{\boldsymbol{\pi}}' \sum_{t=p+1}^{n} \mathbf{L}'_{t-1}\mathbf{L}_{t-1}\hat{\boldsymbol{\pi}} - \lambda\hat{\boldsymbol{\Sigma}}_{\epsilon\epsilon} \right| = 0\,, \tag{10.3.35}$$

where

$$\hat{\boldsymbol{\Sigma}}_{\epsilon\epsilon} = \text{block diag}(\hat{\boldsymbol{\Sigma}}_{ee}, \mathbf{0}_{k(p-1),k(p-1)})\,,$$

$\mathbf{0}_{k(p-1),k(p-1)}$ is a $k(p-1) \times k(p-1)$ matrix of zeros, $\hat{\boldsymbol{\Sigma}}_{ee}$ is defined in (10.3.30), and $\mathbf{L}_{t-1} = (\mathbf{Y}'_{t-1}, \Delta\mathbf{Y}'_{t-1}, \ldots, \Delta\mathbf{Y}'_{t-p+1})$. Then, the maximum likelihood estimator of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = \left[(\hat{\boldsymbol{\pi}}_2, \hat{\boldsymbol{\pi}}_3)' \sum_{t=p+1}^{n} \mathbf{L}'_{t-1}\mathbf{L}_{t-1}(\hat{\boldsymbol{\pi}}_2, \hat{\boldsymbol{\pi}}_3) - \hat{\lambda}_k\hat{\boldsymbol{\Sigma}}_{\epsilon\epsilon 22}\right]^{-1}$$

$$\times \left[(\hat{\boldsymbol{\pi}}_2, \hat{\boldsymbol{\pi}}_3)' \sum_{t=p+1}^{n} \mathbf{L}'_{t-1}\mathbf{L}_{t-1}\hat{\boldsymbol{\pi}}_1 - \hat{\lambda}_k\hat{\boldsymbol{\Sigma}}_{\epsilon\epsilon 21}\right], \tag{10.3.36}$$

where $\hat{\boldsymbol{\Sigma}}_{\epsilon\epsilon 22}$ is the lower right $(kp-1) \times (kp-1)$ submatrix of $\hat{\boldsymbol{\Sigma}}_{\epsilon\epsilon}$, and $\hat{\boldsymbol{\Sigma}}_{\epsilon\epsilon 21}$ is the lower left $(kp-1) \times 1$ submatrix of $\hat{\boldsymbol{\Sigma}}_{\epsilon\epsilon}$.

The vector $\hat{\boldsymbol{\theta}}$ is called the limited information maximum likelihood estimator in the econometric literature. Hence, $\hat{\boldsymbol{\theta}}$ can be calculated by any of the many econometric computer packages. The limiting distributions of the test statistic and the limiting distribution of the coefficients associated with the conditional likelihood were obtained by Johansen (1988) and Ahn and Reinsel (1990). We give a proof that differs somewhat from the proofs of those authors.

Theorem 10.3.4. Suppose the model (10.3.18) holds with exactly one root of (10.3.19) equal to one and exactly one root of $\mathbf{H}_1$ equal to one. Assume the $\mathbf{e}_t$ are iid$(\mathbf{0}, \boldsymbol{\Sigma}_{ee})$ random variables or that the $\mathbf{e}_t$ are martingale differences satisfying the conditions of Lemma 10.3.1. Assume that $\mathbf{Y}_t$ is such that $\boldsymbol{\kappa}_1$ can be normalized to give equation (10.3.34). Let $\hat{\boldsymbol{\theta}}$ of (10.3.36) be the estimator of $\boldsymbol{\theta}$. Let

$$\hat{\mathbf{V}}_{\theta\theta} = \left(\sum_{t=p+1}^{n} \hat{\boldsymbol{\psi}}'_t\hat{\boldsymbol{\psi}}_t\right)^{-1}(1, -\hat{\mathbf{B}}_{0,12})\hat{\boldsymbol{\Sigma}}_{ee}(1, -\hat{\mathbf{B}}_{0,12})'\,,$$

where $\hat{\boldsymbol{\Sigma}}_{ee}$ is defined in (10.3.30),

$$\boldsymbol{\psi}_t = (\Delta Y_{2t}, \Delta Y_{3t}, \ldots, \Delta Y_{kt}, \Delta\mathbf{Y}'_{t-1}, \Delta\mathbf{Y}'_{t-2}, \ldots, \Delta\mathbf{Y}'_{t-p+1})\,,$$

and $\hat{\boldsymbol{\psi}}_t$ is the predicted value of $\boldsymbol{\psi}_t$ from the regression of $\boldsymbol{\psi}_t$ on $(\mathbf{Y}'_{t-1}, \Delta\mathbf{Y}'_{t-1}, \ldots, \Delta\mathbf{Y}'_{t-p+1})$. Then

$$\hat{\mathbf{V}}_{\theta\theta}^{-1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I})$$

and

$$\hat{\lambda}_k \xrightarrow{\mathcal{L}} \hat{\tau}_l^2\,,$$

where $\hat{\tau} \xrightarrow{\mathcal{L}} \hat{\tau}_l$, $\hat{\tau}$ is defined in Corollary 10.1.1.2, the distribution of $\hat{\tau}_l$ is given in Corollary 10.1.1.2, and $\hat{\lambda}_k$ is defined in (10.3.35).

Proof. Let $\ddot{\mathbf{Y}}_{t-1}$ be the deviation from fit for the regression of $\mathbf{Y}'_{t-1}$ on $(\Delta\mathbf{Y}'_{t-1}, \ldots, \Delta\mathbf{Y}'_{t-p+1})$. Then $\hat{\lambda}_k$ is also the smallest root of

$$\left|(\hat{\boldsymbol{\pi}}_{11}, \hat{\boldsymbol{\pi}}_{12})' \sum_{t=p+1}^{n} \ddot{\mathbf{Y}}_{t-1}\ddot{\mathbf{Y}}'_{t-1}(\hat{\boldsymbol{\pi}}_{11}, \hat{\boldsymbol{\pi}}_{12}) - \lambda\hat{\boldsymbol{\Sigma}}_{ee}\right| = 0,$$

where $(\hat{\boldsymbol{\pi}}_{11}, \hat{\boldsymbol{\pi}}_{12})$ is the upper $k \times k$ portion of $(\hat{\boldsymbol{\pi}}_1, \hat{\boldsymbol{\pi}}_2)$. Note that $(\hat{\boldsymbol{\pi}}_{11}, \hat{\boldsymbol{\pi}}_{12}) = \hat{\mathbf{H}}'_1 - \mathbf{I}$.

Let $\mathbf{T}_{11}$ be a transformation matrix such that the unit root is associated with the first element of $\mathbf{X}_t = \mathbf{T}_{11}\mathbf{Y}_t$. That is, $\mathbf{X}_t$ satisfies the equation

$$\mathbf{X}_t = \mathbf{H}_{x1}\mathbf{X}_{t-1} + \sum_{i=2}^{p} \mathbf{H}_{xi}\,\Delta\mathbf{X}_{t-i+1} + \mathbf{T}_{11}\mathbf{e}_t,$$

where $\mathbf{H}_{xi} = \mathbf{T}_{11}\mathbf{H}_i\mathbf{T}_{11}^{-1}$, $\mathbf{H}_i$ is defined in (10.3.21), and the first row and first column of $\mathbf{H}_{x1} - \mathbf{I}$ are vectors of zeros. Then

$$\begin{aligned}
&(\hat{\boldsymbol{\pi}}_{11}, \hat{\boldsymbol{\pi}}_{12})' \sum_{t=p+1}^{n} \ddot{\mathbf{Y}}_{t-1}\ddot{\mathbf{Y}}'_{t-1}(\hat{\boldsymbol{\pi}}_{11}, \hat{\boldsymbol{\pi}}_{12}) \\
&\quad = (\hat{\boldsymbol{\pi}}_{11}, \hat{\boldsymbol{\pi}}_{12})'\mathbf{T}_{11}^{-1} \sum_{t=p+1}^{n} \ddot{\mathbf{X}}_{t-1}\ddot{\mathbf{X}}'_{t-1}\mathbf{T}_{11}^{-1\prime}(\hat{\boldsymbol{\pi}}_{11}, \hat{\boldsymbol{\pi}}_{12}) \\
&\quad = \hat{\boldsymbol{\xi}}' \sum_{t=p+1}^{n} \ddot{\mathbf{X}}_{t-1}\ddot{\mathbf{X}}'_{t-1}\hat{\boldsymbol{\xi}},
\end{aligned}$$

where $\hat{\boldsymbol{\xi}} = (\hat{\boldsymbol{\xi}}_1, \hat{\boldsymbol{\xi}}_2) = \mathbf{T}_{11}^{-1\prime}(\hat{\boldsymbol{\pi}}_{11}, \hat{\boldsymbol{\pi}}_{12})$, and $\ddot{\mathbf{X}}_{t-1}$ is the deviation from fit for the regression of $\mathbf{X}_{t-1}$ on $(\Delta\mathbf{Y}'_{t-1}, \ldots, \Delta\mathbf{Y}'_{t-p+1})$. Also,

$$\left(\hat{\boldsymbol{\xi}}' \sum_{t=p+1}^{n} \ddot{\mathbf{X}}_{t-1}\ddot{\mathbf{X}}'_{t-1}\hat{\boldsymbol{\xi}} - \hat{\lambda}_k\hat{\boldsymbol{\Sigma}}_{ee}\right)(1, -\hat{\mathbf{B}}_{0,12})' = \mathbf{0}, \qquad (10.3.37)$$

where $\hat{\lambda}_k$ is the smallest root of (10.3.35) and the roots of (10.3.35) are the roots of

$$\left|\hat{\boldsymbol{\xi}}' \sum_{t=p+1}^{n} \ddot{\mathbf{X}}_{t-1}\ddot{\mathbf{X}}'_{t-1}\hat{\boldsymbol{\xi}} - \hat{\lambda}_k\hat{\boldsymbol{\Sigma}}_{ee}\right| = 0.$$

Now, $(\sum_{t=p+1}^{n} \ddot{X}_{1,t-1}^2)^{-1} = O_p(n^{-2})$, $\hat{\boldsymbol{\xi}}_1 = O_p(n^{-1})$, and

$$\sum_{t=p+1}^{n} \ddot{X}_{1,t-1}\,\Delta Y_{jt}\left(\sum_{t=p+1}^{n} \ddot{X}_{1,t-1}^2\right)^{-1} \sum_{t=p+1}^{n} \ddot{X}_{1,t-1}\,\Delta Y_{it} = O_p(1)$$

for $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, k$, by the arguments used in the proof of Theorem 10.1.2. Therefore, the elements of the first row and column of $\hat{\boldsymbol{\xi}}' \Sigma \ddot{\mathbf{X}}_{t-1}\ddot{\mathbf{X}}'_{t-1}\hat{\boldsymbol{\xi}}$ are $O_p(1)$. Also, because $(X_{2t}, \ldots, X_{kt})$ is stationary with positive definite covariance matrix, the matrix of mean squares and products of $(\ddot{X}_{2t}, \ldots, \ddot{X}_{kt})$ is an estimator of a nonsingular matrix. Also, by the assumption that $\mathbf{H}_1 - \mathbf{I}$ has one zero root, $\boldsymbol{\xi}_2$ is of rank $k-1$ and

$$(n-p)^{-1}(\hat{\boldsymbol{\pi}}_2, \hat{\boldsymbol{\pi}}_3)' \sum_{t=p+1}^{n} \mathbf{L}'_{t-1}\mathbf{L}_{t-1}(\hat{\boldsymbol{\pi}}_2, \hat{\boldsymbol{\pi}}_3) \xrightarrow{P} \bar{\mathbf{M}},$$

where $\bar{\mathbf{M}}$ is a nonsingular matrix. It follows that $\hat{\lambda}_k = O_p(1)$. Also see Theorem 2.4.1 of Fuller (1987, p. 151).

From the definition of $\hat{\boldsymbol{\theta}}$,

$$\begin{aligned}
\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} &= \bar{\mathbf{M}}^{-1}(n-p)^{-1}(\boldsymbol{\pi}_2, \boldsymbol{\pi}_3)' \sum_{t=p+1}^{n} \mathbf{L}'_{t-1}\mathbf{L}_{t-1}[\hat{\boldsymbol{\pi}}_1 - (\hat{\boldsymbol{\pi}}_2, \hat{\boldsymbol{\pi}}_3)\boldsymbol{\theta}] + o_p(n^{-1/2}) \\
&= \bar{\mathbf{M}}^{-1}(n-p)^{-1}(\boldsymbol{\pi}_2, \boldsymbol{\pi}_3)' \sum_{t=p+1}^{n} \mathbf{L}'_{t-1}\mathbf{e}'_t\mathbf{B}'_{0,1} + o_p(n^{-1/2}),
\end{aligned}$$

where $\mathbf{B}_{0,1.} = (1, -\mathbf{B}_{0,12})$ and $\boldsymbol{\pi}$ is the probability limit of $\hat{\boldsymbol{\pi}}$. Because $(\boldsymbol{\pi}_2, \boldsymbol{\pi}_3)'\mathbf{L}'_{t-1}$ is a stationary process, the asymptotic normality of $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ follows from Theorem 5.3.4. See Theorem 8.2.1.

Note that $\mathbf{L}_{t-1}(\hat{\boldsymbol{\pi}}_2, \hat{\boldsymbol{\pi}}_3)$ is the predicted value for $\boldsymbol{\psi}_t$ obtained in the regression of $\boldsymbol{\psi}_t$ on $\mathbf{L}_{t-1}$. Because $\hat{\boldsymbol{\pi}}$ is converging to $\boldsymbol{\pi}$, $\hat{\mathbf{B}}_{0,12}$ is converging to $\mathbf{B}_{0,12}$, and $\hat{\boldsymbol{\Sigma}}_{ee}$ is converging to $\boldsymbol{\Sigma}_{ee}$, we have

$$(n-p)\hat{\mathbf{V}}_{\theta\theta} \xrightarrow{P} \bar{\mathbf{M}}^{-1}\mathbf{B}_{0,1}\boldsymbol{\Sigma}_{ee}\mathbf{B}'_{0,1.},$$

and the distribution result for $\hat{\mathbf{V}}_{\theta\theta}^{-1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is established.

We now obtain the limiting distribution of $\hat{\lambda}_k$. Because the roots are unchanged by linear transformation, we assume, without loss of generality, that the model is in the canonical form. In the canonical form, $\mathbf{B}_{0,1.} = (1, 0, \ldots, 0)$. From (10.3.37),

$$\hat{\lambda}_k = (\hat{\mathbf{B}}_{0,1.}\hat{\boldsymbol{\Sigma}}_{ee}\hat{\mathbf{B}}'_{0,1.})^{-1}\hat{\mathbf{B}}_{0,1.}\hat{\boldsymbol{\xi}}'\ddot{\mathbf{S}}_{XX}\hat{\boldsymbol{\xi}}\hat{\mathbf{B}}'_{0,1.},$$

where $\ddot{\mathbf{S}}_{XX} = \Sigma_{t=p+1}^{n}\ddot{\mathbf{X}}_{t-1}\ddot{\mathbf{X}}'_{t-1}$.

Using $(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)\mathbf{B}'_{0,1.} = \mathbf{0}$, $\mathbf{B}_{0,12} = \mathbf{0}$, $(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) = p\lim(\hat{\boldsymbol{\xi}}_1, \hat{\boldsymbol{\xi}}_2)$, and a Taylor expansion, we have

$$\begin{aligned}
\hat{\boldsymbol{\xi}}_2\hat{\mathbf{B}}'_{0,12} &= \hat{\boldsymbol{\xi}}_2\mathbf{B}'_{0,12} + \boldsymbol{\xi}_2(\hat{\mathbf{B}}'_{0,12} - \mathbf{B}'_{0,12}) + O_p(n^{-1}) \\
&= \boldsymbol{\xi}_2\hat{\mathbf{B}}'_{0,12} + O_p(n^{-1}),
\end{aligned}$$

where

$$\hat{\mathbf{B}}'_{0,12} = (\boldsymbol{\xi}'_2 \ddot{\mathbf{S}}_{XX} \boldsymbol{\xi}_2)^{-1} \boldsymbol{\xi}'_2 \sum_{t=p+1}^{n} \ddot{\mathbf{X}}_{t-1} e_{1t} + o_p(n^{-1/2})$$

and the first row of $\boldsymbol{\xi}_2$ is a zero vector. It follows that

$$\begin{aligned}\hat{\lambda}_k &= \sigma_{ee11}^{-1} (\hat{\boldsymbol{\xi}}_1 - \boldsymbol{\xi}_2 \hat{\mathbf{B}}'_{0,12})' \ddot{\mathbf{S}}_{XX} (\hat{\boldsymbol{\xi}}_1 - \boldsymbol{\xi}_2 \hat{\mathbf{B}}'_{0,12}) + o_p(1) \\ &= \sigma_{ee11}^{-1} \left(\sum_{t=p+1}^{n} \ddot{\mathbf{X}}_{t-1} e_{1t} \right)' [\ddot{\mathbf{S}}_{XX}^{-1} - \boldsymbol{\xi}_2 (\boldsymbol{\xi}'_2 \ddot{\mathbf{S}}_{XX} \boldsymbol{\xi}_2)^{-1} \boldsymbol{\xi}'_2] \sum_{t=p+1}^{n} \ddot{\mathbf{X}}_{t-1} e_{1t} + o_p(1) .\end{aligned}$$

The standardized matrix

$$\ddot{\mathbf{D}}_{XX}^{-1/2} [\ddot{\mathbf{S}}_{XX}^{-1} - \boldsymbol{\xi}_2 (\boldsymbol{\xi}'_2 \ddot{\mathbf{S}}_{XX} \boldsymbol{\xi}_2)^{-1} \boldsymbol{\xi}'_2] \ddot{\mathbf{D}}_{XX}^{-1/2} ,$$

where $\ddot{\mathbf{D}}_{XX}^{-1} = \operatorname{diag} \ddot{\mathbf{S}}_{XX}$, is converging to a matrix with $(1, 0, \ldots, 0)$ as the first row. Therefore,

$$\hat{\lambda}_k = \sigma_{ee11}^{-1} \left(\sum_{t=p+1}^{n} \ddot{X}_{1,t-1}^2 \right)^{-1} \left(\sum_{t=p+1}^{n} \ddot{\mathbf{X}}_{1,t-1} e_{1t} \right)^2 + o_p(1)$$

and $\hat{\lambda}_k \xrightarrow{\mathscr{L}} \hat{\tau}_t^2$. See the proof of Theorem 10.1.2. ▲

The results extend to the model with an intercept.

Corollary 10.3.4. Let the model

$$\mathbf{Y}_t = \mathbf{H}_0 + \mathbf{H}_1 \mathbf{Y}_{t-1} + \sum_{i=2}^{p} \mathbf{H}_i \, \Delta \mathbf{Y}_{t-i+1} + e_t , \qquad t = 2, 3, \ldots, n ,$$

be estimated by ordinary least squares where $\mathbf{H}_0$ is a k-dimensional column vector. Let $\boldsymbol{\kappa}_1^{0\prime} \mathbf{H}_0^0 = \mathbf{0}$, where $\boldsymbol{\kappa}_1^{0\prime} (\mathbf{H}_1^0 - \mathbf{I}) = \mathbf{0}$, and $\boldsymbol{\kappa}_1^0$ and $\mathbf{H}_1^0$ are the true parameters of the process. Let all statistics be defined as in Theorem 10.3.4 except that the regression equations contain an intercept. Let the error assumptions of Theorem 10.3.4 hold. Let $\hat{\boldsymbol{\theta}} = (\hat{\mathbf{B}}_{0,12}, \hat{\zeta}_1, \hat{\mathbf{B}}_{2,1.}, \ldots, \hat{\mathbf{B}}_{p,1.})$, where $\hat{\zeta}_1$ is the estimated intercept for the model (10.3.34) expanded to include an intercept. Let $\hat{\mathbf{V}}_{\theta\theta}$ be defined as in Theorem 10.3.4 with

$$\boldsymbol{\psi}_t = (\Delta Y_{2t}, \Delta Y_{3t}, \ldots, \Delta Y_{kt}, 1, \Delta \mathbf{Y}'_{t-1}, \Delta \mathbf{Y}'_{t-2}, \ldots, \Delta \mathbf{Y}'_{t-p}) .$$

Then

$$\hat{\mathbf{V}}_{\theta\theta}^{-1/2} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathscr{L}} N(\mathbf{0}, \mathbf{I})$$

and

$$\hat{\lambda}_k \xrightarrow{\mathscr{L}} \hat{\tau}_{\mu,t}^2 ,$$

where $\hat{\tau}_\mu \xrightarrow{\mathcal{L}} \hat{\tau}_{\mu l}$, the limit random variable $\hat{\tau}_{\mu l}$ has the distribution given in Theorem 10.1.3, and $\hat{\lambda}_k$ is defined by (10.3.35) with $\mathbf{L}_{t-1}$ expanded to include a one.

Proof. Omitted. ▲

One method of estimating the entire system (10.3.20) subject to the restriction (10.3.33) is to use a nonlinear estimation program. The computations are illustrated in Example 10.3.1 of the next subsection.

10.3.3. Vector Process with Several Unit Roots

We now extend our discussion to the estimation of the multivariate process (10.3.18) subject to the restriction that exactly g roots of (10.3.19) are equal to one, under the assumption that the part of the Jordan canonical form of $\mathbf{H}_1$ associated with the unit roots is the g-dimensional identity matrix. This means that there are g vectors $\boldsymbol{\kappa}_i$ such that

$$\boldsymbol{\kappa}_i'(\mathbf{H}_1 - \mathbf{I}) = \mathbf{0}\,, \tag{10.3.38}$$

where the $\boldsymbol{\kappa}_i$ define a g-dimensional subspace.

Let

$$\overset{*}{\mathbf{S}}_{hh} = (\hat{\mathbf{H}}_1 - \mathbf{I})\mathbf{V}_{11}^{-1}(\hat{\mathbf{H}}_1 - \mathbf{I})'\,, \tag{10.3.39}$$

where $\mathbf{V}_{11}$ is the portion of the inverse of

$$\sum_{t=p+1}^{n} (\mathbf{Y}_{t-1}', \Delta\mathbf{Y}_{t-1}', \ldots, \Delta\mathbf{Y}_{t-p+1}')'(\mathbf{Y}_{t-1}', \Delta\mathbf{Y}_{t-1}', \ldots, \Delta\mathbf{Y}_{t-p+1}')$$

associated with $\mathbf{Y}_{t-1}$. Let $\hat{\boldsymbol{\kappa}}_i$ be the characteristic vectors of $\overset{*}{\mathbf{S}}_{hh}$ in the metric $\hat{\boldsymbol{\Sigma}}_{ee}$, and let

$$\hat{\mathbf{K}} = (\hat{\boldsymbol{\kappa}}_1, \hat{\boldsymbol{\kappa}}_2, \ldots, \hat{\boldsymbol{\kappa}}_k)\,,$$

where $\hat{\boldsymbol{\Sigma}}_{ee}$ is defined in (10.3.30). That is,

$$(\overset{*}{\mathbf{S}}_{hh} - \hat{\lambda}_i\hat{\boldsymbol{\Sigma}}_{ee})\hat{\boldsymbol{\kappa}}_i = 0\,, \qquad i = 1, 2, \ldots, k\,, \tag{10.3.40}$$

where $\hat{\lambda}_i$, $i = 1, 2, \ldots, k$, are the roots of

$$|\overset{*}{\mathbf{S}}_{hh} - \lambda\hat{\boldsymbol{\Sigma}}_{ee}| = 0 \tag{10.3.41}$$

and $\hat{\mathbf{K}}'\hat{\boldsymbol{\Sigma}}_{ee}\hat{\mathbf{K}} = \mathbf{I}$.

For normal e_t, the method of maximum likelihood defines the estimated g-dimensional subspace with the g vectors associated with the g smallest roots of (10.3.41). The vectors $\hat{\boldsymbol{\kappa}}_i$, $i = k - g + 1, k - g + 2, \ldots, k$, define a g-dimensional subspace, but any set of g vectors of rank g formed as linear combinations of the $\hat{\kappa}_i$ can be used to define the subspace.

We have solved the problem of estimating those linear combinations of the original variables that are unit root processes. One could ask for those linear combinations of the original variables that define stationary processes, the cointegrating vectors. The cointegrating vectors are obtained in Example 10.3.2 using a maximum likelihood computer program. The cointegrating vectors can be computed directly using a determinantal equation different than (10.3.41). Define $\mathbf{S}_{00}$, $\mathbf{S}_{10}$, and $\mathbf{S}_{01}$ by $\mathbf{S}_{01} = \mathbf{S}'_{10}$,

$$\mathbf{H}'_1 - \mathbf{I} = \mathbf{V}_{11}\mathbf{S}_{10}, \tag{10.3.42}$$

and

$$d_f\hat{\mathbf{\Sigma}}_{ee} = \mathbf{S}_{00} - \mathbf{S}_{01}\mathbf{V}_{11}\mathbf{S}_{10}, \tag{10.3.43}$$

where d_f is the degrees of freedom for $\hat{\mathbf{\Sigma}}_{ee}$. For the first order model, $\mathbf{V}_{11}^{-1} = \Sigma_{t=2}^{n}\mathbf{Y}_{t-1}\mathbf{Y}'_{t-1}$ and

$$(\mathbf{S}_{00}, \mathbf{S}_{01}) = \left(\sum_{t=2}^{n} \Delta\mathbf{Y}_t\Delta\mathbf{Y}'_t, \sum_{t=2}^{n} \Delta\mathbf{Y}_t\mathbf{Y}'_{t-1}\right).$$

Then, the $k-g$ cointegrating vectors are the vectors associated with the $k-g$ largest roots of

$$|\mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01} - \nu\mathbf{V}_{11}^{-1}| = 0. \tag{10.3.44}$$

The roots $\hat{\nu}$ of (10.3.44) are related to the roots of (10.3.41) by

$$\hat{\nu} = (1 + d_f^{-1}\hat{\lambda})d_f^{-1}\hat{\lambda} \tag{10.3.45}$$

and $\hat{\lambda} = (1-\hat{\nu})^{-1}\hat{\nu}d_f$.

Given that (10.3.38) holds, there is some arrangement of the elements of $\mathbf{Y}_t$ such that

$$(\mathbf{I}, -\mathbf{B}_{0,12})(\mathbf{H}_1 - \mathbf{I}) = \mathbf{0},$$

where $(\mathbf{I}, -\mathbf{B}_{0,12})$ is a $g \times k$ matrix obtained as a nonsingular transformation of $(\boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2, \ldots, \boldsymbol{\kappa}_g)'$. Thus, if we let

$$\mathbf{B}_0 = \begin{pmatrix} \mathbf{I}, & -\mathbf{B}_{0,12} \\ \mathbf{0}, & \mathbf{I} \end{pmatrix},$$

we can multiply the model (10.3.20) by $\mathbf{B}_0$ to obtain

$$\mathbf{B}_0 \Delta \mathbf{Y}_t = \mathbf{B}_1(\mathbf{Y}'_{1,t-1}, \mathbf{Y}'_{2,t-1})' + \sum_{j=1}^{p-1} \mathbf{B}_{j+1} \Delta \mathbf{Y}_{t-j} + \mathbf{v}_t \,, \qquad (10.3.46)$$

where $\mathbf{v}_t = \mathbf{B}_0 \mathbf{e}_t$, and the first g rows of $\mathbf{B}_1$ are composed of zeros. Then the maximum likelihood estimator of $\mathbf{B}_{0,12}$ is

$$\hat{\mathbf{B}}_{0,12} = \hat{\mathbf{\Sigma}}^{-1}_{\pi\pi 22} \hat{\mathbf{\Sigma}}_{\pi\pi 21} \,, \qquad (10.3.47)$$

where

$$\hat{\mathbf{\Sigma}}_{\pi\pi} = \hat{\mathbf{\Sigma}}_{ee} \hat{\mathbf{K}}[\mathrm{Diag}(\hat{\lambda}_1 - 1, \hat{\lambda}_2 - 1, \ldots, \hat{\lambda}_{k-g} - 1, 0, 0, \ldots, 0)] \hat{\mathbf{K}}' \hat{\mathbf{\Sigma}}_{ee} \,,$$

$\hat{\mathbf{\Sigma}}_{\pi\pi 22}$ is the lower right $(k-g) \times (k-g)$ submatrix of $\hat{\mathbf{\Sigma}}_{\pi\pi}$, $\hat{\mathbf{K}}$ is the matrix of vectors $\hat{\boldsymbol{\kappa}}_i$, and $\hat{\boldsymbol{\kappa}}_i$, $i = 1, 2, \ldots, k$, are the characteristic vectors in the metric $\hat{\mathbf{\Sigma}}_{ee}$ associated with equation (10.3.40). A derivation of this estimator in a different context is given in Fuller (1987, Theorems 4.1.1 and 4.1.2). The estimators of $\mathbf{B}_{j,1.}$, $j = 2, 3, \ldots, p$, are given by the regression of $\hat{\mathbf{B}}_0 \Delta \mathbf{Y}_t$ on $(\Delta \mathbf{Y}'_{t-1}, \Delta \mathbf{Y}'_{t-2}, \ldots, \Delta \mathbf{Y}'_{t-p+1})$.

We now give the limiting distributions for some of the statistics associated with maximum likelihood estimation.

Theorem 10.3.5. Assume that the model (10.3.19), (10.3.46) holds, where $\mathbf{e}_t$ are iid$(\mathbf{0}, \mathbf{\Sigma}_{ee})$ random vectors, g of the roots of $\mathbf{H}_1$ are one, and $k - g$ of the roots are less than one in absolute value. Assume that the part of the Jordan canonical form associated with the unit roots is diagonal. Let $\hat{\mathbf{H}}_1$ be the least squares estimator of $\mathbf{H}_1$ obtained in the least squares fit of equation (10.3.20). Let $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_k$ be the roots of the equation

$$|(\hat{\mathbf{H}}_1 - \mathbf{I})\mathbf{V}_{11}^{-1}(\hat{\mathbf{H}}_1 - \mathbf{I})' - \lambda \hat{\mathbf{\Sigma}}_{ee}| = 0 \,,$$

where $\hat{\mathbf{\Sigma}}_{ee}$ is the least squares estimator of $\mathbf{\Sigma}_{ee}$, and $\mathbf{V}_{11}$ is defined with (10.3.39). Then the distribution of $(\hat{\lambda}_{k-g+1}, \hat{\lambda}_{k-g+2}, \ldots, \hat{\lambda}_k)$ converges to the distribution of the roots $(\hat{\nu}_1, \hat{\nu}_2, \ldots, \hat{\nu}_g)$ of

$$|\mathbf{Y}'_{gg} \mathbf{G}_{gg}^{-1} \mathbf{Y}_{gg} - \nu \mathbf{I}_g| = 0 \,,$$

where $\mathbf{G}_{gg}$ and $\mathbf{Y}_{gg}$ are $g \times g$ matrices defined in Lemma 10.3.1.

If the statistics are computed using deviations from the mean, the distribution of $(\hat{\lambda}_{k-g+1}, \hat{\lambda}_{k-g+2}, \ldots, \hat{\lambda}_k)$ converges to the distribution of the roots $(\hat{\nu}_1, \hat{\nu}_2, \ldots, \hat{\nu}_g)$ of

$$|(\mathbf{Y}_{gg} - \boldsymbol{\zeta}\boldsymbol{\eta}')'(\mathbf{G}_{gg} - \boldsymbol{\zeta}\boldsymbol{\zeta}')^{-1}(\mathbf{Y}_{gg} - \boldsymbol{\zeta}\boldsymbol{\eta}') - \nu \mathbf{I}_g| = 0 \,,$$

where $\boldsymbol{\zeta}$ and $\boldsymbol{\eta}$ are g-dimensional vectors defined in Lemma 10.3.1.

Let $\hat{\mathbf{B}}_{0,12}$ be defined by (10.3.47), let $\mathbf{B}_{j,1.}$ be the $g \times k$ matrix composed of the first g rows of $\mathbf{B}_j$, $j = 2, 3, \ldots, p$, of (10.3.46), and let $\hat{\mathbf{B}}_{j,1.}$, $j = 2, 3, \ldots, p$, be the associated estimators of $\mathbf{B}_{j,1.}$, $j = 2, 3, \ldots, p$. Then

$$n^{1/2}\operatorname{vec}\{(\hat{\mathbf{B}} - \mathbf{B})'\} \xrightarrow{\mathscr{L}} N(\mathbf{0}, \mathbf{V}_{BB}),$$

where $\mathbf{V}_{BB} = \mathbf{\Sigma}_{vv} \otimes \bar{\mathbf{M}}^{-1}$, $\bar{\mathbf{M}} = E\{\bar{\boldsymbol{\psi}}_t' \bar{\boldsymbol{\psi}}_t\}$,

$$\mathbf{\Sigma}_{vv} = (\mathbf{I}, -\mathbf{B}_{0,12})\mathbf{\Sigma}_{ee}(\mathbf{I}, -\mathbf{B}_{0,12})',$$

$$\bar{\boldsymbol{\psi}}_t = (\Delta\bar{\mathbf{Y}}_{2t}', \Delta\mathbf{Y}_{t-1}', \ldots, \Delta\mathbf{Y}_{t-p+1}),$$

$$\Delta\bar{\mathbf{Y}}_t = (\mathbf{H}_1 - \mathbf{I})\mathbf{Y}_{t-1} + \sum_{j=1}^{p-1} \mathbf{H}_{j+1}\,\Delta\mathbf{Y}_{t-j},$$

and

$$\hat{\mathbf{B}} - \mathbf{B} = (\hat{\mathbf{B}}_{0,12} - \mathbf{B}_{0,12}, \hat{\mathbf{B}}_{2,1.} - \mathbf{B}_{2,1.}, \ldots, \hat{\mathbf{B}}_{p,1.} - \mathbf{B}_{p,1.}).$$

Proof. We outline the proof for the no-intercept model. The model can be transformed to the canonical form (10.3.24) and the roots of (10.3.41) computed with the canonical form are the same as those computed with the original variables.

Let $\ddot{\mathbf{Y}}_{t-1}$ be the column vector of residuals obtained in the regression of $\mathbf{Y}_{t-1}$ on $(\Delta\mathbf{Y}_{t-1}', \ldots, \Delta\mathbf{Y}_{t-p+1}')$, and let s_{ii} be the regression residual mean square for the regression of ΔY_{it} on $(\mathbf{Y}_{t-1}', \Delta\mathbf{Y}_{t-1}', \ldots, \Delta\mathbf{Y}_{t-p+1}')$. Then, in the set of ordinary regression statistics, the coefficients for $\mathbf{Y}_{t-1}$ in the regression of ΔY_{it} on $(\mathbf{Y}_{t-1}', \Delta\mathbf{Y}_{t-1}', \ldots, \Delta\mathbf{Y}_{t-p+1}')$ are the regression coefficients for the regression of ΔY_{it} on $\ddot{\mathbf{Y}}_{t-1}$ and

$$V_{11}s_{ii} = \left(\sum_{t=p+1}^{n} \ddot{\mathbf{Y}}_{t-1}\ddot{\mathbf{Y}}_{t-1}'\right)^{-1} s_{ii}$$

is the estimated covariance matrix for that vector of coefficients. For the model (10.3.26) in canonical form,

$$\mathbf{D}_n^{-1} \sum_{t=p+1}^{n} \ddot{\mathbf{Y}}_{t-1}\ddot{\mathbf{Y}}_{t-1}'\mathbf{D}_n^{-1} \Rightarrow \begin{pmatrix} \mathbf{C}_g\mathbf{G}_{gg}\mathbf{C}_g' & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_{22} \end{pmatrix},$$

where $\mathbf{C}_g\mathbf{G}_{gg}\mathbf{C}_g'$ is $g \times g$, $\mathbf{M}_{22}$ is $(k-g) \times (k-g)$,

$$\mathbf{M}_{22} = E\{(\ddot{Y}_{g+1,t}, \ddot{Y}_{g+2,t}, \ldots, \ddot{Y}_{kt})'(\ddot{Y}_{g+1,t}, \ddot{Y}_{g+2,t}, \ldots, \ddot{Y}_{kt})\},$$

and $\mathbf{D}_n = \operatorname{diag}(n, n, \ldots, n, n^{1/2}, n^{1/2}, \ldots, n^{1/2})$. Also,

$$\mathbf{V}_{11}^{1/2} \sum_{t=p+1}^{n} \ddot{\mathbf{Y}}_{t-1}\mathbf{e}_t' \stackrel{\text{def}}{=} \mathbf{R}_n = \begin{pmatrix} \mathbf{R}_{n11} & \mathbf{R}_{n12} \\ \mathbf{R}_{n21} & \mathbf{R}_{n22} \end{pmatrix}$$
$$\Rightarrow \begin{pmatrix} (\mathbf{C}_g\mathbf{G}_{gg}\mathbf{C}_g')^{-1/2}\mathbf{C}_g\overset{*}{\mathbf{Y}}_{gg} & (\mathbf{C}_g\mathbf{G}_{gg}\mathbf{C}_g')^{-1/2}\mathbf{C}_g\overset{*}{\mathbf{Y}}_{12} \\ \mathbf{\Psi}_{21} & \mathbf{\Psi}_{22} \end{pmatrix},$$

where $\overset{*}{\mathbf{Y}}$ is the limit random variable for $n^{-1}\sum_{t=1}^{n}(\sum_{j=1}^{t-1}\mathbf{e}_j)\mathbf{e}_t'$, $\overset{*}{\mathbf{Y}}_{gg}$ is the upper left $g \times g$ submatrix of $\overset{*}{\mathbf{Y}}$, $\overset{*}{\mathbf{Y}}_{12}$ is the upper right $g \times (k-g)$ submatrix of $\overset{*}{\mathbf{Y}}$, and the elements of $\mathbf{\Psi}_{ij}$ are normal random variables. See the proof of Theorem 10.3.1. Then

$$(\hat{\mathbf{H}}_1 - \mathbf{I}) \sum_{t=p+1}^{n} \ddot{\mathbf{Y}}_{t-1}\ddot{\mathbf{Y}}_{t-1}'(\hat{\mathbf{H}}_1 - \mathbf{I})' = (\mathbf{R}_n + \overset{*}{\mathbf{M}}_n)'(\mathbf{R}_n + \overset{*}{\mathbf{M}}_n),$$

where

$$\overset{*}{\mathbf{M}}_n = n^{1/2}\begin{pmatrix} 0 & 0 \\ 0 & \mathbf{M}_{22}^{-1/2} \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 0 & \mathbf{H}_{1,22}' - \mathbf{I} \end{pmatrix}.$$

The nonsingular matrix $n^{1/2}\mathbf{M}_{22}(\mathbf{H}_{1,22}' - \mathbf{I})$ dominates $\mathbf{R}_{n22}$, which is $O_p(1)$. Thus,

$$(\mathbf{R}_n + \overset{*}{\mathbf{M}}_n)'(\mathbf{R}_n + \overset{*}{\mathbf{M}}_n) = \begin{pmatrix} \mathbf{R}_{n11}'\mathbf{R}_{n11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{0} & \mathbf{R}_{n21}' \\ \mathbf{0} & \overset{*}{\mathbf{M}}_{n22}' \end{pmatrix}\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{R}_{n21} & \overset{*}{\mathbf{M}}_{n22} \end{pmatrix}$$
$$+ \text{remainder},$$

where the remainder terms are of small order relative to the included terms. Now $\overset{*}{\mathbf{M}}_{n22}'\overset{*}{\mathbf{M}}_{n22}$ is a nonsingular matrix multiplied by n, $\mathbf{R}_{n11}$ is $O_p(1)$, $\mathbf{R}_{n21}$ is $O_p(1)$, and the upper left $g \times g$ portion of $\hat{\mathbf{\Sigma}}_{ee}$ converges to $\mathbf{I}_g$. Therefore, the g smallest roots of (10.3.41) converge to the roots of

$$|\mathbf{Y}_{gg}(\mathbf{C}_g\mathbf{C}_g'\mathbf{G}_{gg}\mathbf{C}_g')^{-1}\mathbf{C}_g\mathbf{Y}_{gg} - \nu\mathbf{I}_g| = 0.$$

Because the elements of $\hat{\mathbf{\Sigma}}_{ee}$ are small relative to $\overset{*}{\mathbf{M}}_n$ and the g smallest roots are $O_p(1)$, the limiting behavior of $\hat{\mathbf{\Sigma}}_{\pi\pi}$ of (10.3.47) is that of

$$\begin{pmatrix} \mathbf{0} & \mathbf{R}_{n21}' \\ \mathbf{0} & \overset{*}{\mathbf{M}}_{n22}' \end{pmatrix}\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{R}_{n21} & \overset{*}{\mathbf{M}}_{n22} \end{pmatrix}.$$

Therefore, the limiting behavior of $n^{1/2}(\hat{\mathbf{B}}_{0,12} - \mathbf{B}_{0,12})$ is the limiting behavior of

$$\begin{aligned} n^{1/2}(\overset{*}{\mathbf{M}}_{n22}'\overset{*}{\mathbf{M}}_{n22})^{-1}\overset{*}{\mathbf{M}}_{n22}'\mathbf{R}_{n21} &= n^{1/2}\overset{*}{\mathbf{M}}_{n22}^{-1}\mathbf{R}_{n21} \\ &= (\mathbf{H}_{1,22} - \mathbf{I})^{-1}\mathbf{M}_{22}^{-1/2}\mathbf{\Psi}_{21}. \end{aligned}$$

In the canonical form, $\hat{\mathbf{B}}_{0,12}$ is estimating the zero matrix with an error that is converging to a matrix of normal random variables. Because the maximum likelihood estimators are invariant to linear transformations, the distribution for the estimators of the parameters of the original model are the appropriate transformation of the canonical $\hat{\mathbf{B}}_{0,12}$. The limiting distribution of the vector of other elements of $\hat{\mathbf{B}}$ is obtained by the arguments used in the proof of Theorem 10.3.4. ▲

Table 10.A.6 of Appendix 10.A contains percentiles to be used in testing for unit roots in the characteristic equation (10.3.19). The entries in the table are the percentiles associated with

$$\overset{*}{\lambda}_i = (1 + d_f^{-1}\hat{\lambda}_i)^{-1}\hat{\lambda}_i , \tag{10.3.48}$$

where $\hat{\lambda}_i$, $i = 1, 2, \ldots, k$, are the roots of the determinantal equation (10.3.41), and d_f is the degrees of freedom for $\hat{\mathbf{\Sigma}}_{ee}$. Monte Carlo studies conducted by Hoen Jin Park (personal communication) indicate that in small samples with $g > 1$ that statistics based upon $\overset{*}{\lambda}_i$ have power superior to the corresponding test based on $\hat{\lambda}_i$.

The first row of Table 10.A.6 is the distribution of the test statistic for testing the hypothesis of a single unit root against the alternative of no unit roots, the second row is for the hypothesis of two unit roots against the alternative of no unit roots, the third row is for the hypothesis of two unit roots against the alternative of one unit root, the fourth row is for the hypothesis of three unit roots against the alternative of no unit roots, etc. The test statistic for row 1 is $\overset{*}{\lambda}_k$, the test statistic for row 2 is $\overset{*}{\lambda}_{k-1} + \overset{*}{\lambda}_k$, the test statistic for row 3 is $\overset{*}{\lambda}_k$, the test statistic for row 4 is $\overset{*}{\lambda}_{k-2} + \overset{*}{\lambda}_{k-1} + \overset{*}{\lambda}_k$, the test statistic for row 5 is $\overset{*}{\lambda}_{k-1} + \overset{*}{\lambda}_k$, etc. The numbers at the left are the degrees of freedom associated with $\hat{\mathbf{\Sigma}}_{ee}$.

Table 10.A.7 contains the percentiles of analogous statistics for the model (10.3.20) estimated with an intercept. Table 10.A.8 contains the percentiles of analogous statistics for the model estimated with a time trend.

The tables were constructed by Heon Jin Park using the Monte Carlo method. The limiting percentiles were calculated by simulation using the infinite series representations of Lemma 10.3.1. The percentiles were smoothed using a function of d_f, where d_f is the degrees of freedom. The standard errors of the entries in the tables are a function of the size of the entries and range from 0.001 for the smallest entries to 0.25 for the largest entries.

We now discuss the estimation of $\mathbf{H}_1 - \mathbf{I}$ of equation (10.3.20) subject to the linear constraints associated with the presence of unit roots. Assume that the least squares statistics have been computed, and let $\hat{\mathbf{H}}_1$ be the least squares estimator of $\mathbf{H}_1$. Let $\mathbf{V}_{11}$ be the part of the inverse of the matrix of sums of squares and products associated with $\mathbf{Y}_{t-1}$. Each column of $\hat{\mathbf{H}}_1'$ is a vector of regression coefficients and the ordinary least squares estimator of the "covariance matrix" of a column is $\mathbf{V}_{11}$ multiplied by the error variance. Thus, arguing in a nonrigorous manner, the rows of

$$\mathbf{V}_{11}^{-1/2}(\hat{\mathbf{H}}_1' - \mathbf{H}_1')$$

have the property that they are estimated to be uncorrelated.

Under the hypothesis of g unit roots in the original system, equivalent to g zero roots for $\mathbf{H}_1 - \mathbf{I}$, there is a $\mathbf{B}_{0,1.}$ such that

$$\mathbf{V}_{11}^{-1/2}(\mathbf{H}_1' - \mathbf{I})\mathbf{B}_{0,1.}' = \mathbf{0}. \tag{10.3.49}$$

The estimator of $\mathbf{B}_{0,1.}$, up to a linear transformation, is given by the g characteristic vectors of (10.3.40) associated with the g smallest roots. Let

$$\hat{\mathbf{R}}_1' = \mathbf{V}_{11}^{-1/2}(\hat{\mathbf{H}}_1' - \mathbf{I})$$

and

$$\mathbf{r}_1' = \mathbf{V}_{11}^{-1/2}(\mathbf{H}_1' - \mathbf{I}).$$

Now each column of $\mathbf{v} = \mathbf{B}_{0,1.}\hat{\mathbf{R}}_1$ contains g linear contrasts that are estimating zero and $\hat{\mathbf{B}}_{0,1.}\hat{\mathbf{R}}_1$ is an estimator of those contrasts. Therefore, an improved estimator of a column of $\mathbf{r}_1$ is obtained by subtracting from the column of $\hat{\mathbf{R}}_1$ an estimator of the error in $\hat{\mathbf{R}}_1$ constructed with the estimator of the g linear contrasts. Thus, the estimator is

$$\begin{aligned}\tilde{\mathbf{r}}_{1,i.}' &= \hat{\mathbf{R}}_{1,i.}' - \hat{\mathbf{v}}_i'\hat{\mathbf{\Sigma}}_{vv}^{-1}\hat{\mathbf{\Sigma}}_{ve} \\ &= \hat{\mathbf{R}}_{1,i.}'(\mathbf{I} - \hat{\mathbf{B}}_{0,1.}'\hat{\mathbf{\Sigma}}_{vv}^{-1}\hat{\mathbf{\Sigma}}_{ve}), \qquad i = 1, 2, \ldots, k, \end{aligned} \tag{10.3.50}$$

where $\hat{\mathbf{v}}_i' = \hat{\mathbf{R}}_{1,i.}'\hat{\mathbf{B}}_{0,1.}'$, $\hat{\mathbf{\Sigma}}_{vv} = \hat{\mathbf{B}}_{0,1}\hat{\mathbf{\Sigma}}_{ee}\hat{\mathbf{B}}_{0,1.}'$, and $\hat{\mathbf{\Sigma}}_{ve} = \hat{\mathbf{B}}_{0,1.}\hat{\mathbf{\Sigma}}_{ee}$. It follows that the estimator of $\mathbf{H}_1' - \mathbf{I}$ is

$$\tilde{\mathbf{H}}_1' - \mathbf{I} = \mathbf{V}_{11}^{1/2}\tilde{\mathbf{r}}_1' = (\hat{\mathbf{H}}_1' - \mathbf{I})(\mathbf{I} - \hat{\mathbf{B}}_{0,1.}'\hat{\mathbf{\Sigma}}_{vv}^{-1}\hat{\mathbf{\Sigma}}_{ve}). \tag{10.3.51}$$

The estimated covariance matrix of the estimators $\tilde{\mathbf{r}}_{1,i.}$, $i = 1, 2, \ldots, k$, is

$$\hat{\mathbf{\Sigma}}_{rr} = \hat{\mathbf{\Sigma}}_{ee} - \hat{\mathbf{\Sigma}}_{ev}\hat{\mathbf{\Sigma}}_{vv}^{-1}\hat{\mathbf{\Sigma}}_{ve}.$$

Therefore, the estimated covariance matrix of $\operatorname{vec}(\tilde{\mathbf{H}}_1 - \mathbf{I})$ is

$$\hat{V}\{\operatorname{vec}(\tilde{\mathbf{H}}_1 - \mathbf{I})\} = (\mathbf{V}_{11}^{1/2} \otimes \mathbf{I})(\mathbf{I} \otimes \hat{\mathbf{\Sigma}}_{rr})(\mathbf{V}_{11}^{1/2} \otimes \mathbf{I})'.$$

The estimator of $\mathbf{H}_1 - \mathbf{I}$ is seen to be the least squares estimator adjusted by the estimators of zero, where the coefficient matrix is the matrix of regression coefficients for the regression of the error in $\hat{\mathbf{H}}_1 - \mathbf{I}$ on the estimators of zero. This formulation can be used to construct the maximum likelihood estimators of the entire model. Let

$$\mathbf{d} = \operatorname{vec}\{(\hat{\mathbf{H}}_1 - \mathbf{I})'(\mathbf{I}, -\hat{\mathbf{B}}_{0,12})'\}$$

be the kg-dimensional vector of estimators of zero. Then

$$\text{vec}\{(\tilde{\mathbf{H}}_1 - \mathbf{I}, \tilde{\mathbf{H}}_2, \ldots, \tilde{\mathbf{H}}_p)'\} = \text{vec}\{(\hat{\mathbf{H}}_1 - \mathbf{I}, \hat{\mathbf{H}}_2, \ldots, \hat{\mathbf{H}}_p)'\} - \mathbf{d}'\hat{\mathbf{\Sigma}}_{dd}^{-1}\hat{\mathbf{\Sigma}}_{d,\text{vec}\mathbf{H}},$$

where $\hat{\mathbf{\Sigma}}_{dd}$ is the estimated covariance matrix of $\mathbf{d}$ and $\hat{\mathbf{\Sigma}}_{d,\text{vec}\mathbf{H}}$ is the $gk \times (p \times k)$ estimated covariance matrix of $\mathbf{d}$ and $\text{vec}\,\hat{\mathbf{H}}'$. The estimated covariance matrix of the estimators,

$$\hat{\mathbf{V}}\{\text{vec}\,\tilde{\mathbf{H}}'\} = \hat{\mathbf{\Sigma}}_{\text{vec}\,\mathbf{H},\,\text{vec}\mathbf{H}} - \hat{\mathbf{\Sigma}}_{\text{vec}\,\mathbf{H},\,d}\hat{\mathbf{\Sigma}}_{dd}^{-1}\hat{\mathbf{\Sigma}}_{d,\text{vec}\,\mathbf{H}}, \tag{10.3.52}$$

follows from regression theory.

If the original vector autoregressive process is

$$\mathbf{Y}_t = \mathbf{H}_0 + \mathbf{H}_1\mathbf{Y}_{t-1} + \mathbf{H}_2\,\Delta\mathbf{Y}_{t-1} + \cdots + \mathbf{H}_p\,\Delta\mathbf{Y}_{t-p+1} + \mathbf{e}_t, \tag{10.3.53}$$

we may wish to estimate the vector of intercepts $\mathbf{H}_0$ with the restriction that

$$\mathbf{B}_{0,1}\mathbf{H}_0 = \mathbf{0}. \tag{10.3.54}$$

This is the restriction that the linear combinations of the original variables that are unit root processes are processes with no drift. The estimator of $\mathbf{B}_{0,1.}$ is given by the eigenvectors associated with the smallest roots of

$$|(\hat{\mathbf{H}}_0, \hat{\mathbf{H}}_1 - \mathbf{I})\mathbf{V}_{**}^{-1}(\hat{\mathbf{H}}_0, \hat{\mathbf{H}}_1 - \mathbf{I})' - \lambda\hat{\mathbf{\Sigma}}_{ee}| = 0, \tag{10.3.55}$$

where $\mathbf{V}_{**}$ is the portion of the estimated covariance matrix of the least squares estimators associated with the vector of regressor variables $(1, \mathbf{Y}'_{t-1})$.

In Example 10.3.2 we demonstrate how existing software for simultaneous equation systems can be used to construct the estimators and estimated standard errors of the estimators.

Example 10.3.2. To illustrate the testing and estimation methods for multivariate autoregressive time series, we use the data on U.S. interest rates studied by Stock and Watson (1988). The three series are the federal funds rate, denoted by Y_{1t}; the 90-day treasury bill rate, denoted by Y_{2t}; and the one-year treasury bill rate, denoted by Y_{3t}. The data are 236 monthly observations for the period, January 1960 through August 1979. The one-year treasury bill interest rate was studied as a univariate series in Examples 10.1.1 and 10.1.2.

Following the procedures of Example 10.1.1, a second order process fit to the differences yields

$$\begin{aligned}
\Delta^2\hat{Y}_{1t} &= \underset{(0.025)}{0.017} - \underset{(0.075)}{0.550}\,\Delta Y_{1,t-1} - \underset{(0.065)}{0.161}\,\Delta^2 Y_{1,t-1}, \\
\Delta^2\hat{Y}_{2t} &= \underset{(0.021)}{0.021} - \underset{(0.082)}{0.782}\,\Delta Y_{2,t-1} - \underset{(0.065)}{0.025}\,\Delta^2 Y_{2,t-1}, \\
\Delta^2\hat{Y}_{3t} &= \underset{(0.019)}{0.018} - \underset{(0.076)}{0.768}\,\Delta Y_{3,t-1} + \underset{(0.065)}{0.105}\,\Delta^2 Y_{3,t-1}.
\end{aligned} \tag{10.3.56}$$

The $\hat{\tau}_\mu$ statistics for the hypothesis of a second unit root under the maintained hypothesis of a single unit root are -7.37, -9.52, and -10.13, respectively. Thus, in all cases, the hypothesis of a second unit root is rejected at the one percent level.

Third order autoregressive models fit by ordinary least squares to each series gives

$$\begin{aligned}
\Delta\hat{Y}_{1t} &= \underset{(0.062)}{0.108} - \underset{(0.010)}{0.017}\, Y_{1,t-1} + \underset{(0.065)}{0.297}\, \Delta Y_{1,t-1} + \underset{(0.066)}{0.177}\, \Delta Y_{1,t-2}\,, \\
\Delta\hat{Y}_{2t} &= \underset{(0.062)}{0.078} - \underset{(0.012)}{0.011}\, Y_{2,t-1} + \underset{(0.066)}{0.200}\, \Delta Y_{2,t-1} + N(\underset{(0.066)}{0.035}\, \Delta Y_{2,t-2}\,, \qquad (10.3.57) \\
\Delta\hat{Y}_{3t} &= \underset{(0.062)}{0.082} - \underset{(0.011)}{0.012}\, Y_{3,t-1} + \underset{(0.065)}{0.343}\, \Delta Y_{3,t-1} - \underset{(0.066)}{0.094}\, \Delta Y_{3,t-2}\,.
\end{aligned}$$

The three residual mean squares are 0.143, 0.098, and 0.083 for the federal funds rate, 90-day treasury rate, and one-year treasury rate, respectively. There are 233 observations in the regression, and the residual mean square has 229 degrees of freedom. The $\hat{\tau}_\mu$ statistics for the hypothesis of a unit root in the univariate processes are -1.62, -0.98, and -1.09. The 10% point for $\hat{\tau}_\mu$ for a sample with 229 degrees of freedom is about -2.57. Hence, the null hypothesis of a unit root is not rejected for each series. Models were also fit with time as an additional explanatory variable. In all cases, the hypothesis of a unit root was accepted at the 10% level. On the basis of these results, we proceed as if each series were an autoregressive process with a single unit root and other roots less than one in absolute value.

We write the multivariate process as

$$\mathbf{Y}_t = \mathbf{H}_0 + \mathbf{H}_1\mathbf{Y}_{t-1} + \mathbf{H}_2\,\Delta\mathbf{Y}_{t-1} + \mathbf{H}_3\,\Delta\mathbf{Y}_{t-2} + \mathbf{e}_t\,, \qquad (10.3.58)$$

or as

$$\Delta\mathbf{Y}_t = \mathbf{H}_0 + (\mathbf{H}_1 - \mathbf{I})\mathbf{Y}_{t-1} + \mathbf{H}_2\,\Delta\mathbf{Y}_{t-1} + \mathbf{H}_3\,\Delta\mathbf{Y}_{t-2} + \mathbf{e}_t\,, \qquad (10.3.59)$$

where $\mathbf{Y}_t = (Y_{1t}, Y_{2t}, Y_{3t})'$ is the column vector of interest rates, $\mathbf{H}_0$ is a three-dimensional column vector, and $\mathbf{H}_i$, $i = 1, 2, 3$, are 3×3 matrices of parameters. If the multivariate process has a single unit root, the rank of $\mathbf{H}_1 - \mathbf{I}$ is 2. The rank of $\mathbf{H}_1 - \mathbf{I}$ is one if the multivariate process has two unit roots.

To investigate the number of unit roots associated with the multivariate process, we compute the roots of

$$|(\hat{\mathbf{H}}_1 - \mathbf{I})\mathbf{V}_{11}^{-1}(\hat{\mathbf{H}}_1 - \mathbf{I})' - \lambda\hat{\mathbf{\Sigma}}_{ee}| = 0\,, \qquad (10.3.60)$$

where $\hat{\mathbf{H}}_1$ is the least squares estimate of $\mathbf{H}_1$, $\hat{\mathbf{\Sigma}}_{ee}$ is the least squares estimate of $\mathbf{\Sigma}_{ee}$, and $\mathbf{V}_{11}\hat{\sigma}_{eeii}$ is the least squares estimated covariance matrix of the ith row of $\hat{\mathbf{H}}_1$. The matrix $(\hat{\mathbf{H}}_1 - \mathbf{I})\mathbf{V}_{11}^{-1}(\hat{\mathbf{H}}_1 - \mathbf{I})'$ can be computed by subtracting the residual sum of squares and products matrix for the full model from that for the reduced model

$$\Delta \mathbf{Y}_t = \mathbf{H}_0 + \mathbf{H}_2 \, \Delta \mathbf{Y}_{t-1} + \mathbf{H}_3 \, \Delta \mathbf{Y}_{t-2} + \mathbf{e}_t \, . \qquad (10.3.61)$$

The full model has nine explanatory variables $(\mathbf{Y}'_{t-1}, \Delta \mathbf{Y}'_{t-1}, \Delta \mathbf{Y}'_{t-2})$ plus the intercept and is fitted to 233 observations. Hence, the estimator of $\boldsymbol{\Sigma}_{ee}$ has 223 degrees of freedom. The reduced model contains $(\Delta \mathbf{Y}_{t-1}, \Delta \mathbf{Y}_{t-2})$ and the intercept. The matrices are

$$\hat{\mathbf{H}}'_1 - \mathbf{I} = \begin{pmatrix} -0.167 & 0.043 & 0.020 \\ (0.043) & (0.034) & (0.033) \\ 0.314 & -0.245 & -0.058 \\ (0.121) & (0.098) & (0.094) \\ -0.109 & 0.189 & 0.019 \\ (0.098) & (0.080) & (0.077) \end{pmatrix},$$

$$\hat{\boldsymbol{\Sigma}}_{ee} = \begin{pmatrix} 0.1304 & 0.0500 & 0.0516 \\ 0.0500 & 0.0856 & 0.0712 \\ 0.0516 & 0.0712 & 0.0788 \end{pmatrix}, \qquad (10.3.62)$$

$$(\hat{\mathbf{H}}_1 - \mathbf{I})\mathbf{V}_{11}^{-1}(\hat{\mathbf{H}}_1 - \mathbf{I})' = \begin{pmatrix} 2.0636 & -0.4693 & -0.1571 \\ -0.4693 & 0.5912 & 0.1752 \\ -0.1571 & 0.1752 & 0.1574 \end{pmatrix}. \qquad (10.3.63)$$

The standard errors as output by an ordinary regression program are given in parentheses below the estimated elements of $(\hat{\mathbf{H}}_1 - \mathbf{I})'$. The estimates of $\mathbf{H}_1 - \mathbf{I}$ are not normally distributed in the limit when the $\mathbf{Y}_t$-process contains unit roots.

The three roots of (10.3.60) are 1.60, 12.67, and 32.30. The test statistic tabled by Park is

$$\sum_{i=k-g+1}^{k} (1 + d_f^{-1} \hat{\lambda}_i)^{-1} \hat{\lambda}_i = \sum_{i=k-g+1}^{k} \overset{*}{\lambda}_i \, ,$$

where d_f is the degrees of freedom for $\hat{\boldsymbol{\Sigma}}_{ee}$, $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_k$, and g is the hypothesized number of unit roots in the process. The test statistic for the hypothesis of a single unit root is $\overset{*}{\lambda}_3 = 1.59$. Because the median of the statistic for a single root is 2.43, the hypothesis of one unit root is easily accepted. The test for two unit roots against the alternative of one unit root is based on the second smallest root $\overset{*}{\lambda}_2 = 11.99$. On the basis of this statistic, the hypothesis of two unit roots is accepted at the 10% level because the 10% point for the largest of two $\hat{\lambda}$ given two unit roots is about 12.5. The hypothesis of three unit roots is rejected because $\overset{*}{\lambda}_1 = 28.21$, exceeds the 1% tabular value for the largest of three $\overset{*}{\lambda}$ given three unit roots. For illustrative purposes, we estimate the process as a multivariate process with two unit roots and diagonal Jordan canonical form.

The eigenvectors associated with the determinantal equation (10.3.60) are

$$\begin{aligned}\hat{\boldsymbol{\kappa}}_1 &= (0.4374, 0.1042, 3.1472)' ,\\ \hat{\boldsymbol{\kappa}}_2 &= (1.8061, 5.3600, -6.3725)' ,\\ \hat{\boldsymbol{\kappa}}_3 &= (2.6368, -4.3171, 1.8279)' .\end{aligned} \tag{10.3.64}$$

If the vector process contains two unit roots, the two time series defined by

$$(\mathbf{W}_{1t}, \mathbf{W}_{2t}) = (\hat{\boldsymbol{\kappa}}_1'\mathbf{Y}_t, \hat{\boldsymbol{\kappa}}_2'\mathbf{Y}_t) , \tag{10.3.65}$$

are estimated to be the unit root processes. Furthermore, the errors in these two processes are estimated to be uncorrelated. Thus, it is estimated that

$$(\mathbf{H}_1 - \mathbf{I})(\hat{\boldsymbol{\kappa}}_1, \hat{\boldsymbol{\kappa}}_2) = (\mathbf{0}, \mathbf{0}) \tag{10.3.66}$$

and that

$$(\hat{\boldsymbol{\kappa}}_1, \hat{\boldsymbol{\kappa}}_2)'\boldsymbol{\Sigma}_{ee}(\hat{\boldsymbol{\kappa}}_1, \hat{\boldsymbol{\kappa}}_2) = \mathbf{I} .$$

The determinantal equation (10.3.44) that can be used to define the cointegrating vectors is

$$\left| \begin{pmatrix} 15.00 & 6.47 & 7.18 \\ 6.47 & 5.26 & 5.50 \\ 7.18 & 5.50 & 6.69 \end{pmatrix} - \nu \begin{pmatrix} 1261.6 & 897.0 & 834.9 \\ 897.0 & 679.4 & 639.1 \\ 834.9 & 639.1 & 615.8 \end{pmatrix} \right| = 0 .$$

The roots are 0.1264, 0.0537, and 0.0071, and the associated vectors are

$$\begin{aligned}&(-0.104, 0.313, -0.188)' ,\\ &(-0.056, -0.106, 0.195)' ,\\ &(0.004, 0.056, -0.025)' .\end{aligned} \tag{10.3.67}$$

If we assume that there are two unit roots, then the first vector of (10.3.67) defines the single cointegrating vector.

We note that if (10.3.66) holds,

$$(\mathbf{H}_1 - \mathbf{I})(\hat{\boldsymbol{\kappa}}_1, \hat{\boldsymbol{\kappa}}_2)\mathbf{K}_{11}^{-1} = (\mathbf{0}, \mathbf{0}) \tag{10.3.68}$$

for any nonsingular 2×2 matrix $\mathbf{K}_{11}$. In particular, we could let

$$\mathbf{K}_{11} = \begin{pmatrix} \hat{\kappa}_{12} & \hat{\kappa}_{22} \\ \hat{\kappa}_{13} & \hat{\kappa}_{23} \end{pmatrix} . \tag{10.3.69}$$

The $\mathbf{K}_{11}$ of (10.3.69) is nonsingular if Y_{2t} and Y_{3t} are both unit root processes and are not both functions of a single common unit root process. Therefore, in practice, care should be exercised in the choice of normalization. For our example, we assume that

$$(\mathbf{H}_1 - \mathbf{I})'\boldsymbol{\beta} = \mathbf{0},$$

where

$$\boldsymbol{\beta}' = \begin{pmatrix} -\beta_{21} & 1 & 0 \\ -\beta_{31} & 0 & 1 \end{pmatrix}. \tag{10.3.70}$$

We create a square nonsingular matrix by augmenting $\boldsymbol{\beta}$ with a column $(0, 0, 1)'$ to obtain

$$\mathbf{B}' = \begin{pmatrix} 1 & 0 & 0 \\ -\beta_{21} & 1 & 0 \\ -\beta_{31} & 0 & 1 \end{pmatrix}.$$

If we multiply (10.3.59) by $\mathbf{B}'$, we obtain the restricted model

$$\begin{aligned} \Delta Y_{1t} &= \theta_{11} Y_{1,t-1} + \theta_{12} Y_{2,t-1} + \theta_{13} Y_{3,t-1} + \mathbf{W}_t \boldsymbol{\theta}_{1w} + u_{1t}, \\ \Delta Y_{2t} &= \beta_{21}\, \Delta Y_{1t} + \mathbf{W}_t \boldsymbol{\theta}_{2w} + u_{2t}, \\ \Delta Y_{3t} &= \beta_{31}\, \Delta Y_{1t} + \mathbf{W}_t \boldsymbol{\theta}_{3w} + u_{3t}, \end{aligned} \tag{10.3.71}$$

where $\mathbf{W}_t = (1, \Delta \mathbf{Y}'_{t-1}, \Delta \mathbf{Y}'_{t-2})$. This model is in the form of a set of simultaneous equations. A number of programs are available to estimate the parameters of such models. We use the full information maximum likelihood option of SAS/ETS® to estimate the parameters. The estimates are

$$\begin{aligned} \Delta \hat{Y}_{1t} &= -\underset{(0.039)}{0.130}\, Y_{1,t-1} + \underset{(0.110)}{0.396}\, Y_{2,t-1} - \underset{(0.080)}{0.238}\, Y_{3,t-1} + \mathbf{W}_t \hat{\boldsymbol{\theta}}_{1w}, \\ \Delta \hat{Y}_{2t} &= -\underset{(0.319)}{0.487}\, \Delta Y_{1t} + \mathbf{W}_t \hat{\boldsymbol{\theta}}_{2w}, \\ \Delta \hat{Y}_{3t} &= -\underset{(0.238)}{0.122}\, \Delta Y_{1t} + \mathbf{W}_t \hat{\boldsymbol{\theta}}_{3w}, \end{aligned} \tag{10.3.72}$$

where we have ommitted the numerical values of $(\hat{\boldsymbol{\theta}}_{1w}, \hat{\boldsymbol{\theta}}_{2w}, \hat{\boldsymbol{\theta}}_{3w})$ to simplify the display. The vector $(-0.130, 0.396, -0.238)$ applied to $\mathbf{Y}_t$ defines a process that is estimated to be stationary. The vector is a multiple of the cointegrating vector given as the first vector of (10.3.67).

The estimator of $\mathbf{H}_1' - \mathbf{I}$ estimated subject to the restriction that the rank of $\mathbf{H}_1 - \mathbf{I}$ is one is

$$\tilde{\mathbf{H}}_1' - \mathbf{I} = \begin{pmatrix} \underset{(0.039)}{-0.130} & \underset{(0.030)}{0.063} & \underset{(0.027)}{0.016} \\ \underset{(0.110)}{0.396} & \underset{(0.091)}{-0.192} & \underset{(0.087)}{-0.048} \\ \underset{(0.080)}{-0.238} & \underset{(0.058)}{0.116} & \underset{(0.052)}{0.029} \end{pmatrix}. \tag{10.3.73}$$

The standard errors of the estimates are given in parentheses below the estimates. The estimates are not all normally distributed. However, the difference between the estimate and the true value divided by the standard error converges to a $N(0, 1)$ random variable for a correctly specified model. Therefore, the usual regression testing and confidence interval procedures can be applied to the restricted estimates under the assumption that the process contains two unit roots. The tests are subject to the usual preliminary test bias if one estimates the number of unit roots on the basis of hypothesis tests.

The two time series defined by

$$Y_{2t} + 0.487Y_{1t}$$

and

$$Y_{3t} + 0.122Y_{1t}$$

are processes estimated to have unit roots. These two processes are linear combinations of the two processes in (10.3.65). That is, the two processes are obtained by transforming the first two vectors of (10.3.64),

$$\begin{pmatrix} 0.437 & 1.806 \\ 0.104 & 5.360 \\ 3.147 & -6.372 \end{pmatrix} \begin{pmatrix} 0.104 & 5.360 \\ 3.147 & -6.372 \end{pmatrix}^{-1} = \begin{pmatrix} 0.487 & 0.122 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}. \qquad \blacktriangle\blacktriangle$$

10.4. TESTING FOR A UNIT ROOT IN A MOVING AVERAGE MODEL

The distributional results developed for the autoregressive process with a unit root can be used to test the hypothesis that a moving average process has a unit root. Let the model be

$$Y_t = e_t + \beta e_{t-1}, \qquad t = 1, 2, \ldots, \tag{10.4.1}$$

where the e_t are iid$(0, \sigma^2)$ random variables. We assume e_0 is unknown and to be estimated. We have

$$\begin{aligned} Y_1 &= \beta e_0 + e_1, \\ Y_t &= f_t(\mathbf{Y}; \beta, e_0) + e_t, \qquad t = 2, 3, \ldots, \end{aligned} \tag{10.4.2}$$

where

$$f_t(\mathbf{Y}; \beta, e_0) = -\sum_{j=1}^{t-1} (-\beta)^j Y_{t-j} - (-\beta)^t e_0.$$

On the basis of equation (10.4.2) we define the function $e_t = e_t(\mathbf{Y}; \beta, e_0)$, where

$$e_t(\mathbf{Y}; \beta, e_0) = \sum_{j=0}^{t-1} (-\beta)^j Y_{t-j} + (-\beta)^t e_0.$$

Then

$$\frac{\partial}{\partial \beta} e_t(\mathbf{Y}; \beta, e_0) = -\sum_{j=0}^{t-1} j(-\beta)^{j-1} Y_{t-j} - t(-\beta)^{t-1} e_0$$

and

$$\frac{\partial e_t(\mathbf{Y}; \beta, e_0)}{\partial e_0} = (-\beta)^t .$$

If $\beta = -1$, we have from (10.4.1)

$$\sum_{j=1}^{t} Y_j = -e_0 + e_t . \tag{10.4.3}$$

Therefore, treating e_0 as a fixed parameter to be estimated,

$$\hat{e}_0 = -n^{-1} \sum_{t=1}^{n} \sum_{j=1}^{t} Y_j \tag{10.4.4}$$

is the best linear unbiased estimator of e_0 when $\beta = -1$, and the error in $\hat{e}_0$ is $-\bar{e}_n = -n^{-1} \sum_{t=1}^{n} e_t$.

We shall construct a test of the hypothesis that $\beta = -1$ by considering one step of a Gauss–Newton iteration for (β, e_0) with $(-1, \hat{e}_0)$ as the initial value for (β, e_0). The dependent variable in such an iteration is

$$e_t(\mathbf{Y}; -1, \hat{e}_0) = \sum_{j=1}^{t} Y_j + \hat{e}_0 . \tag{10.4.5}$$

Under the null hypothesis,

$$e_t(\mathbf{Y}; -1, \hat{e}_0) = e_t - \bar{e}_n .$$

Let $W_t(\mathbf{Y}; \beta, e_0)$ denote the partial derivative of $f(\mathbf{Y}; \beta, e_0)$ with respect to β, and note that $W_t(\mathbf{Y}; \hat{\beta}, \hat{e}_0)$ is the negative of the partial derivative of $e_t(\mathbf{Y}; \beta, e_0)$ with respect to β evaluated at $(\hat{\beta}, \hat{e}_0)$. Then, under the null,

$$W_t(\mathbf{Y}; -1, \hat{e}_0) = -t\bar{e}_n + \sum_{j=0}^{t-1} e_j , \qquad t = 2, 3, \ldots, n . \tag{10.4.6}$$

The $W_t(\mathbf{Y}; -1, \hat{e}_0)$ satisfy the recursive relationship

$$\begin{aligned}
&W_1(\mathbf{Y}; -1, \hat{e}_0) = \hat{e}_0 , \\
&W_2(\mathbf{Y}; -1, \hat{e}_0) = Y_1 + 2\hat{e}_0 , \\
&W_t(\mathbf{Y}; -1, \hat{e}_0) = Y_{t-1} + 2W_{t-1}(\mathbf{Y}; -1, \hat{e}_0) - W_{t-2}(\mathbf{Y}; -1, \hat{e}_0) , \\
&\qquad\qquad t = 3, 4, \ldots ,
\end{aligned}$$

and the $e_t(\mathbf{Y}; -1, \hat{e}_0)$ satisfy the recursive relationship

$$e_1(\mathbf{Y}; -1, \hat{e}_0) = Y_1 + \hat{e}_0 ,$$
$$e_t(\mathbf{Y}; -1, \hat{e}_0) = Y_t + e_{t-1}(\mathbf{Y}; -1, \hat{e}_0) , \qquad t = 2, 3, \ldots .$$

The Gauss–Newton iteration consists of regressing $e_t(\mathbf{Y}; -1, \hat{e}_0)$ on $W_t(\mathbf{Y}; -1, \hat{e}_0)$ and a second variable that is identically equal to negative one. The regression equation is

$$e_t(\mathbf{Y}; -1, \hat{e}_0) = W_t(\mathbf{Y}; -1, \hat{e}_0)\,\Delta\beta - \Delta e_0 + e_t , \tag{10.4.7}$$

and the estimator of $\Delta\beta$ is

$$\Delta\hat{\beta} = \left[\sum_{t=1}^{n} (\hat{W}_t - \bar{w}_n)^2\right]^{-1} \sum_{t=1}^{n} (\hat{W}_t - \bar{w}_n)\hat{e}_t ,$$

where $\hat{W}_t = W_t(\mathbf{Y}; -1, \hat{e}_0)$, $\bar{w}_n = n^{-1}\sum_{t=1}^{n} \hat{W}_t$, and $\hat{e}_t = e_t(\mathbf{Y}; -1, \hat{e}_0)$. The pivotal statistic associated with (10.4.7) is

$$\overset{*}{\tau} = \left\{\left[\sum_{t=1}^{n} (\hat{W}_t - \bar{w}_n)^2\right]^{-1} s^2\right\}^{-1/2} \Delta\hat{\beta} , \tag{10.4.8}$$

where

$$s^2 = (n-2)^{-1} \sum_{t=1}^{n} (\hat{e}_t + \Delta\hat{e}_0 - \hat{W}_t\,\Delta\hat{\beta})^2$$

and $\Delta\hat{e}_0$ is the estimated regression coefficient for the regression equation (10.4.7). The limiting distribution of $n\,\Delta\hat{\beta}$ follows from the results of Section 10.1.1.

Lemma 10.4.1. Let Y_t satisfy the model (10.4.1) with $\beta^0 = -1$. Then

$$n\,\Delta\hat{\beta} \xrightarrow{\mathscr{L}} -0.5[G - H^2 - TK + (12)^{-1}T^2]^{-1} ,$$
$$\overset{*}{\tau} \xrightarrow{\mathscr{L}} -0.5[G - H^2 - TK + (12)^{-1}T^2]^{-1/2} ,$$

where $\Delta\hat{\beta}$ is defined in (10.4.7), G, T, and H are defined in Theorem 10.1.3, and K is defined in Theorem 10.1.6.

Proof. Under the null model, the regression coefficient estimating the change in β is

$$\Delta\hat{\beta} = \frac{\sum_{t=1}^{n} [X_{t-1} - \bar{x}_{(-1)} - (t - \bar{t})\bar{e}_n]e_t}{\sum_{t=1}^{n} [X_{t-1} - \bar{x}_{(-1)} - (t - \bar{t})\bar{e}_n]^2} , \tag{10.4.9}$$

where $X_t = \sum_{j=0}^{t} e_j$ and $\bar{x}_{(-1)} = n^{-1}\sum_{t=1}^{n} X_{t-1}$. Now

$$\sum_{t=1}^{n}\left(\sum_{j=0}^{t-1} e_j\right)e_t = 0.5\left(X_n^2 - \sum_{t=0}^{n} e_t^2\right),$$

$$-\bar{e}_n \sum_{t=1}^{n} (t-\bar{t})e_t = -0.5n^{-1}(n-1)X_n^2 + n\bar{x}_{(-1)}\bar{e}_n,$$

and

$$\sum_{t=1}^{n} [X_{t-1} - (t-\bar{t})\bar{e}_n - \bar{x}_{(-1)}]e_t = -0.5\sum_{t=1}^{n} (e_t - \bar{e}_n)^2 .$$

Hence,

$$n\,\Delta\hat{\beta} = \frac{-0.5n^{-1}\sum_{t=1}^{n}(e_t - \bar{e}_n)^2}{G_n - H_n^2 - T_nK_n + (12)^{-1}T_n^2}, \tag{10.4.10}$$

where

$$(G_n, T_n, H_n) = \left(n^{-2}\sum_{t=2}^{n} X_{t-1}^2,\, n^{-1/2}X_n,\, n^{-3/2}\sum_{t=2}^{n} Y_{t-1}\right)$$

and

$$K_n = n^{-5/2}\sum_{j=1}^{n-1} (n-j)(j-1)e_j .$$

The result for $n\,\Delta\hat{\beta}$ follows by the arguments of Theorem 10.1.1. Because $\Delta\hat{\beta} = O_p(n^{-1})$ and $\Delta\hat{e}_0 = O_p(n^{-1/2})$, s^2 converges to σ^2 and we have the result for $\overset{*}{\tau}$. ▲

Percentiles for the test statistic $\overset{*}{\tau}$ are given in Table 10.A.12 of Appendix 10.A. If the true β^0 is greater than -1, the Gauss–Newton procedure should produce a positive $\Delta\hat{\beta}$ to move the estimate away from -1. However, under the null model all values of $\Delta\hat{\beta}$ are negative. Therefore, the null hypothesis is rejected for the least negative of the values, the right tail of Table 10.A.12.

Consider the testing situation in which the null model is

$$Y_t = e_t - e_{t-1}, \qquad t = 1, 2, \ldots, \tag{10.4.11}$$

and the alternative model is

$$Y_t = \mu + \beta e_{t-1} + e_t, \qquad t = 1, 2, \ldots, \tag{10.4.12}$$

where $|\beta| < 1$. Under the alternative model

$$Y_t = f_t(\mathbf{Y};\, \mu, \beta, e_0) + e_t, \tag{10.4.13}$$

where

$$f_t(\mathbf{Y}; \mu, \beta, e_0) = \mu \sum_{j=0}^{t-1} (-\beta)^j - \sum_{j=1}^{t-1} (-\beta)^j Y_{t-j} - (-\beta)^t e_0$$

for $t = 2, 3, \ldots$. The partial derivatives are such that one step of a Gauss–Newton iteration for (μ, β, e_0) with $(0, -1, \hat{e}_0)$ as the initial value is given by the regression of $e_t(\mathbf{Y}; 0, -1, \hat{e}_0)$ defined in (10.4.5) on $(t, W_t, -1)$, where $\hat{e}_0$ is defined in (10.4.4) and $W_t = W_t(\mathbf{Y}; 0, -1, \hat{e}_0)$ is defined in (10.4.6). Under the null, the distribution of the regression pivotal for $W_t(\mathbf{Y}; 0, -1, \hat{e}_0)$ is that tabulated as $\hat{\tau}_\tau$ in Table 10.A.2.

Lemma 10.4.2. Let the model (10.4.1) hold with $\beta = -1$. Let $\Delta\hat{\beta}$ be the coefficient of W_t in the regression of $e_t(\mathbf{Y}; 0, -1, \hat{e}_0)$ on $(t, W_t, -1)$, and let $\hat{\tau}_\tau$ be the corresponding regression pivotal. Then

$$\hat{\tau}_\tau \xrightarrow{\mathscr{L}} 0.5(G - H^2 - 3K^2)^{-1/2}[(T - 2H)(T - 6K) - 1],$$

where G, T, H, and K are defined in Theorem 10.1.6.

Proof. Under the null model

$$\Delta\hat{\beta} = \frac{\sum_{t=1}^{n} [X_{t-1} - \bar{x}_{(-1)} - \hat{b}(t - \bar{t})]e_t}{\sum_{t=1}^{n} [X_{t-1} - \bar{x}_{(-1)} - \hat{b}(t - \bar{t})]^2},$$

where

$$\hat{b} = \left[\sum_{t=1}^{n} (t - \bar{t})^2\right]^{-1} \sum_{t=1}^{n} (t - \bar{t}) X_{t-1}$$

and $X_t = \Sigma_{j=0}^{t} e_j$. This is the same as the expression for $\hat{\theta}_1$ in the first order model (10.1.34), and the result follows from Theorem 10.1.6. ▲

A testing situation in which the null model contains a constant term is also of interest. The null model is

$$Y_t = \mu + e_t - e_{t-1}, \qquad t = 1, 2, \ldots, \tag{10.4.14}$$

and the alternative model is (10.4.12) with $|\beta| < 1$. The test statistic is constructed in a similar manner to that for the null model (10.4.11). The initial values for μ and e_0 are

$$\tilde{\mu} = \left[\sum_{t=1}^{n} (t - \bar{t})^2\right]^{-1} \sum_{t=1}^{n} (t - \bar{t}) \sum_{j=1}^{t} Y_j,$$

and

$$\tilde{e}_0 = -n^{-1} \sum_{t=1}^{n} \sum_{j=1}^{t} Y_j + 0.5\tilde{\mu}(n+1).$$

The value for $\tilde{e}_1$ is $Y_1 + \tilde{e}_0 - \tilde{\mu}$, we have $\tilde{e}_t = \Sigma_{j=1}^{t} Y_j + \tilde{e}_0 - t\tilde{\mu}$ for $t = 2, 3, \ldots, n$, and

$$W_t(\mathbf{Y}; \tilde{\mu}, -1, \tilde{e}_0) = -0.5t(t-1)\tilde{\mu} + \sum_{j=1}^{t-1} jY_{t-j} + t\tilde{e}_0 .$$

The test statistic $\overset{*}{\tau}_{\mu}$ is the regression pivotal for $\Delta\beta$ associated with the regression equation

$$e_t(\mathbf{Y}; \tilde{\mu}, -1, \tilde{e}_0) = (t-1)\,\Delta\mu + W_t(\mathbf{Y}; \tilde{\mu}, -1, \tilde{e}_0)\,\Delta\beta - \Delta e_0 + e_t . \tag{10.4.15}$$

The limiting distribution of the test statistic is given in Lemma 10.4.3. The distribution of the test statistic is tabled in the second part of Table 10.A.12.

Lemma 10.4.3. Let the model (10.4.12) hold with $\beta = -1$. Let $\Delta\hat{\beta}$ be the regression coefficient for $W_t(\mathbf{Y}; \tilde{\mu}, -1, \tilde{e}_0)$ in the regression (10.4.15), and let $\overset{*}{\tau}_{\mu}$ be the corresponding regression pivotal. Then

$$(n\,\Delta\hat{\beta}, \overset{*}{\tau}_{\mu}) \xrightarrow{\mathscr{L}} -0.5(D_{\tau}^{-1}, D_{\tau}^{-1/2}),$$

where $D_{\tau} = G - H^2 - 3K^2 + 0.5(T-2H)^2 - (T-2H)(12L - 2H - 3K)$,

$$L = \int_0^1 \int_0^t \int_0^s W(r)\,dr\,ds\,dt ,$$

G, H, and K are defined in Theorem 10.1.6, and $W(r)$ is the standard Wiener process.

Proof. Omitted. See Arellano and Pantula (1995). ▲

The testing procedure can be extended to the autoregressive moving average process. Let the model be

$$Y_t + \sum_{j=1}^{p} \alpha_j Y_{t-j} = e_t + \sum_{i=1}^{q} \beta_i e_{t-i} \tag{10.4.16}$$

or

$$Y_t + \sum_{j=1}^{p} \alpha_j Y_{t-j} = e_t + \zeta_1 e_{t-1} + \sum_{i=2}^{q} \zeta_i (e_{t-i+1} - e_{t-i}),$$

where the roots of

$$m^p + \sum_{j=1}^{p} \alpha_j m^{p-j} = 0$$

are less than one in absolute value, $\zeta_1 = \Sigma_{i=1}^{q} \beta_i$, and $\zeta_j = -\Sigma_{i=j}^{p} \beta_i$, $j = 2, 3, \ldots, q$. Assume it is desired to test $\zeta_1 = -1$. The derivatives of the function with respect to the autoregressive moving average parameters are given in (8.4.9) of Section 8.4. If we set $\zeta_1 = -1$ in the original model, we have

$$X_t + \sum_{j=1}^{p} \alpha_j X_{t-j} = \kappa + \sum_{i=2}^{q} \zeta_i e_{t-i+1} + e_t$$

for $t = p + 1, p + 2, \ldots$, where $X_t = \Sigma_{j=1}^{t} Y_j$, and

$$\kappa = -e_0 - \sum_{i=2}^{q} \zeta_i e_{1-i} \,.$$

Therefore, treating κ as fixed, X_t is a stationary process (or is converging to a stationary process) with mean κ. If the original model contains a mean μ, then the model for X_t contains the mean function $\kappa_0 + \kappa_1 t$. The testing procedure consists of the following steps:

1. Create the variable $X_t = \Sigma_{j=1}^{t} Y_j$. If the null model is of the form (10.4.16), estimate $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \ldots, \alpha_p, \zeta_2, \ldots, \zeta_q)$ and κ by any of the procedures of Section 8.4. If the null model contains an unknown mean, estimate $\boldsymbol{\theta}$ and (κ_0, κ_1) by any of the procedures of Section 8.4. Let $\tilde{e}_t = e_t(Y_{tj}; \tilde{\boldsymbol{\theta}}', \tilde{\kappa}, \tilde{\zeta})$ be the residuals from this fitting operation for (10.4.16).
2. For the model (10.4.16), evaluate the derivatives at $(\boldsymbol{\theta}', \kappa, \zeta_1) = (\tilde{\boldsymbol{\theta}}', \tilde{\kappa}, -1)$ and carry out one step of the Gauss–Newton procedure with $\tilde{e}_t$ as the dependent variable. The test of $\zeta_1 = -1$ is a test for a unit root of the moving average part of the process. The critical values for the test are the $\overset{*}{\tau}$ of Table 10.A.9. If the null model contains an unknown mean, $(\tilde{\kappa}_0, \tilde{\kappa}_1)$ is a part of the model and changes in these two parameters are estimated as part of the Gauss–Newton step. Critical values for the test of the unknown-mean null model based on ζ_1 are the values given as $\overset{*}{\tau}_\mu$ in Table 10.A.9.

Tanaka (1990), Tsay (1993), Saikkonen and Luukkonen (1993), and Breitung (1994) have discussed alternative testing procedures.

Example 10.4.1. To illustrate testing for a moving average unit root we use computer generated data. The data are given in Table 10.B.3. The model for the data is

$$\begin{aligned} Y_t - \mu + \alpha_1(Y_{t-1} - \mu) + \alpha_2(Y_{t-2} - \mu) &= e_t + \beta_1 e_{t-1} + \beta_2 e_{t-2} \\ &= e_t + \zeta_1 e_{t-1} + \zeta_2(e_{t-1} - e_{t-2}), \end{aligned} \tag{10.4.17}$$

where the e_t are $\mathrm{NI}(0, \sigma^2)$ random variables, $\zeta_1 = \beta_1 + \beta_2$, and $\zeta_2 = -\beta_2$. The maximum likelihood estimates of the parameters are

$$\begin{array}{llllll} (\hat{\mu}, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1, \hat{\beta}_2) = (- 0.021, & - 1.375, & 0.601, & - 0.471, & - 0.401). \\ \qquad\qquad (0.060) & (0.100) & (0.099) & (0.118) & (0.117) \end{array}$$

The roots of the moving average polynomial are 0.911 and −0.440. To test the hypothesis that one of the roots of the moving average polynomial is one, we create the variable $X_t = \Sigma_{j=1}^{t} Y_j$ and estimate the parameters of the model

$$\begin{aligned} X_t &= \kappa_0 + \kappa_1 t + u_t, \\ u_t + \alpha_1 u_{t-1} + \alpha_2 u_{t-2} &= e_t + \zeta_2 e_{t-1}, \end{aligned} \tag{10.4.18}$$

where $\kappa_0 = -e_0 - \zeta_2 e_{-1}$ and $\kappa_1 = \mu(1 + \alpha_1 + \alpha_2)$, by maximum likelihood. The estimates are

$$\begin{array}{llllll} (\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\zeta}_2, \tilde{\kappa}_0, \tilde{\kappa}_1) = (- 1.385, & 0.573, & 0.481, & 0.532, & - 0.009). \\ \qquad\qquad (0.100) & (0.098) & (0.110) & (1.427) & (0.024) \end{array}$$

We construct a test for the null model of the form (10.4.17). The dependent variable for the Gauss–Newton step is the residual $\tilde{e}_t$ obtained from the fit of model (10.4.18). We write the regression equation as

$$\begin{aligned} \tilde{e}_t = \Delta\kappa_0 &+ (t-2)\,\Delta\kappa_1 + \Delta\alpha_1\, Q_{\alpha_1,t} + \Delta\alpha_2\, Q_{\alpha_2,t} \\ &+ \Delta\zeta_1\, Q_{\zeta_1,t} + \Delta\zeta_2\, Q_{\zeta_2,t} + e_t. \end{aligned} \tag{10.4.19}$$

The derivatives defining the explanatory variables are given by the recursive equations

$$\begin{aligned} Q_{\alpha_1,t} &= -Y_{t-1} + Q_{\alpha_1,t-1} - \tilde{\zeta}_2(Q_{\alpha_1,t-1} - Q_{\alpha_1,t-2}) \\ &= -Y_1, \qquad t = 2, \\ Q_{\alpha_2,t} &= -Y_{t-2} + Q_{\alpha_2,t-1} - \tilde{\zeta}_2(Q_{\alpha_2,t-1} - Q_{\alpha_2,t-2}), \\ Q_{\zeta_1,t} &= \tilde{e}_{t-1} + Q_{\zeta_1,t-1} - \tilde{\zeta}_2(Q_{\zeta_1,t-1} - Q_{\zeta_1,t-2}), \\ Q_{\zeta_2,t} &= \tilde{e}_{t-1} - \tilde{e}_{t-2} + Q_{\zeta_2,t-1} - \tilde{\zeta}_2(Q_{\zeta_2,t-1} - Q_{\zeta_2,t-2}), \end{aligned}$$

where the equations hold for $t \geq 3$, and it is understood that the values are zero for $t = 1$ and $t = 2$ unless otherwise specified.

Regressing $\tilde{e}_t$ on the derivatives in a regression with 98 observations, we obtain the estimated regression equation,

$$\tilde{e}_t = \underset{(0.32)}{0.23} - \underset{(0.004)}{0.003}(t-2) - \underset{(0.102)}{0.011}Q_{\alpha_1,t} - \underset{(0.101)}{0.016}Q_{\alpha_2,t} - \underset{(0.051)}{0.081}Q_{\zeta_1,t} + \underset{(0.114)}{0.058}Q_{\zeta_2,t}.$$

The test that the moving average contains a unit root is

$$\overset{*}{\tau}_\mu = -(0.051)^{-1}(0.081) = -1.59\,.$$

The hypothesis of a unit root is just accepted at the 5% level, because -1.59 is just less than the tabular value of -1.563. The hypothesis is rejected at the 10% level because the calculated value is greater than the 10% tabular value of -1.711.

If the null model is (10.4.17) with $(\zeta_1, \mu) = (-1, 0)$, and the alternative is the model (10.4.7), we estimate the model in X_t that contains only κ_0. The maximum likelihood estimates are

$$(\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\zeta}_2, \tilde{\kappa}_0) = (\underset{(0.099)}{-1.385}, \underset{(0.098)}{0.572}, \underset{(0.109)}{0.483}, \underset{(0.724)}{0.074})\,. \qquad (10.4.20)$$

The derivatives for $(\alpha_1, \alpha_2, \zeta_1, \zeta_2)$ are found in the same manner as for the alternative model with a mean. One step of the Gauss–Newton procedure is computed with the intercept but no time variable. The result of the Gauss–Newton step is

$$\tilde{e}_t = \underset{(0.26)}{-0.11} - \underset{(0.102)}{0.002}\,Q_{\alpha_1,t} - \underset{(0.101)}{0.016}\,Q_{\alpha_2,t} - \underset{(0.042)}{0.059}\,Q_{\zeta_1,t} + \underset{(0.112)}{0.051}\,Q_{\zeta_2,t}$$

and the test statistic is $\overset{*}{\tau} = -1.40$. In this case, the hypothesis of a moving average unit root is accepted at the 10% level, because the calculated value is less than the tabular value of -1.282.

If the null is (10.4.17) with $(\zeta_1, \mu) = (-1, 0)$ and the alternative is the model (10.4.17), we use the residuals associated with the estimates (10.4.20) but include time in the Gauss–Newton step. The result is

$$\begin{aligned}\tilde{e}_t = {} & \underset{(0.36)}{0.43} - \underset{(0.004)}{0.006}\,(t-2) - \underset{(0.102)}{0.009}\,Q_{\alpha_1,t} \\ & - \underset{(0.100)}{0.018}\,Q_{\alpha_2,t} - \underset{(0.053)}{0.102}\,Q_{\zeta_1,t} + \underset{(0.113)}{0.070}\,Q_{\zeta_2,t}\,,\end{aligned}$$

and the test statistic is $\hat{\tau}_\tau = -1.91$. Because the 10% critical value of the test obtained from Table 10.A.2 is -1.22, the null hypothesis is accepted at the 10% level. ▲▲

REFERENCES

Section 10.1. Bhargava (1986), Dickey (1976), Dickey and Fuller (1979, 1981), Dickey, Hasza, and Fuller (1984), Dickey and Said (1982), Elliott, Rothenberg, and Stock (1992), Evans and Savin (1981a,b), Gonzalez–Farias and Dickey (1992), Lai and Wei (1985a,b), Pantula (1982, 1988a), Pantula, Gonzalez–Farias, and Fuller (1994), Phillips (1987a–c), Phillips and Perron (1988), Rao (1961), Reeves (1972), Sargan and Bhargava (1983), Stigum (1974), White (1958).

Section 10.2. Anderson (1959), Basawa (1987), Fuller and Hasza (1980, 1981), Hasza (1977), Rao (1961, 1978a,b), Rubin (1950), Stigum (1974), Venkataraman (1967, 1973), White (1958, 1959).

Section 10.3. Ahn and Reinsel (1990), Chan and Wei (1988), Fountis and Dickey (1989), Fuller, Hasza, and Goebel (1981), Johansen (1988), Johansen and Juselius (1990), Lai and Wei (1985a,b), Nagaraj and Fuller (1991), Phillips (1987a–c, 1988, 1991), Phillips and Durlauf (1986), Stock and Watson (1988).

Section 10.4. Arellano (1992), Arellano and Pantula (1995), Chang (1989), Tsay (1993).

EXERCISES

1. Prove

$$\operatorname*{plim}_{n\to\infty}\left[\sum_{t=1}^{n}\left(\sum_{j=1}^{t}e_j\right)^2\right]^{-1}\sum_{t=1}^{n}\left(X_0+\sum_{j=1}^{t}e_j\right)^2=1$$

for fixed X_0, where the e_t are iid$(0, \sigma^2)$.

2. Let the model (10.1.1) hold with $\rho = 1$.
   (a) Show that

$$n\left[\left(\sum_{t=1}^{n}Y_t^2\right)^{-1}\sum_{t=2}^{n}Y_{t-1}Y_t-1\right]\xrightarrow{\mathcal{L}}-(2G)^{-1}(T^2+1)\,.$$

   (b) Show that

$$n\left\{\left[\sum_{t=2}^{n}Y_{t-1}^2\sum_{t=2}^{n}Y_t^2\right]^{-1/2}\sum_{t=2}^{n}Y_{t-1}Y_t-1\right\}\xrightarrow{\mathcal{L}}-(2G)^{-1}\,.$$

3. Let Y_t be the pth order autoregressive process with a unit root considered in Theorem 10.1.2. Let $\hat{m}_1$ be the largest root of the characteristic equation associated with the ordinary least squares estimator. Show that

$$n\left\{\left[\left(\sum_{t=2}^{n}Y_t^2\right)^{-1}\sum_{t=2}^{n}Y_{t-1}^2\right]^{1/2}\hat{m}_1-1\right\}\xrightarrow{\mathcal{L}}-(2G)^{-1}\,,$$

where G is defined in Theorem 10.1.1.

4. Assume that $Y_t = Y_{t-1} + e_t$, $t = 1, 2, \ldots$, where $Y_0 = 0$, and $e_t \sim \mathrm{NI}(0, \sigma^2)$. Show that

$$\mathrm{Var}\{\bar{y}_{(-1)}\} = [6(n-1)]^{-1} n(2n-1)\sigma^2 ,$$

$$\mathrm{Cov}\{\bar{y}_{(-1)}, \bar{e}_{(0)}\} = 0.5(n-1)^{-1}(n-2)\sigma^2 ,$$

$$E\left\{\sum_{t=2}^{n} (Y_{t-1} - \bar{y}_{(-1)})e_t\right\} = -0.5(n-2)\sigma^2 ,$$

$$\mathrm{Var}\left\{\sum_{t=2}^{n} (Y_{t-1} - \bar{y}_{(-1)})e_t\right\} = (12)^{-1}(n+6)(n-2)\sigma^4 ,$$

$$E\left\{\sum_{t=2}^{n} (Y_{t-1} - \bar{y}_{(-1)})^2\right\} = 6^{-1} n(n-2)\sigma^2 ,$$

$$\mathrm{Var}\left\{\sum_{t=2}^{n} (Y_{t-1} - \bar{y}_{(-1)})^2\right\} = (90)^{-1} n(n-2)(2n^2 - 4n + 9)\sigma^4 ,$$

$$\mathrm{Cov}\left\{\sum_{t=2}^{n} (e_t - \bar{e}_{(0)})(Y_{t-1} - \bar{y}_{(-1)}), \sum_{t=2}^{n} (Y_{t-1} - \bar{y}_{(-1)})^2\right\}$$
$$= -(12)^{-1} n(n-2)(2n+1)\sigma^4 ,$$

$$\mathrm{Cov}\left\{\sum_{t=2}^{n} (Y_{t-1} - \bar{y}_{(-1)})(Y_t - \bar{y}_{(0)}), \sum_{t=2}^{n} (Y_{t-1} - \bar{y}_{(-1)})^2\right\} = 6^{-1} n(n-2)\sigma^4 .$$

5. Assume that the time series Y_t satisfies the equation

$$Y_t = \rho Y_{t-1} + e_t , \qquad t = 1, 2, \ldots ,$$

where $Y_0 = 0$ and $e_t \sim \mathrm{NI}(0, 1)$. Let $\overset{*}{\rho} = \hat{\rho}_l$ if $\hat{\rho}_l \le 1$ and $\overset{*}{\rho} = 1$ if $\hat{\rho}_l > 1$, where

$$\hat{\rho}_l = \left(\sum_{t=2}^{n-1} Y_t^2\right)^{-1} \sum_{t=2}^{n} Y_{t-1}Y_t .$$

Let $\hat{\rho}_m$ be the value of ρ that maximizes (10.1.47). Show that the limiting distribution of $n(\overset{*}{\rho} - 1)$ has a smaller second moment about zero than the limiting distribution of $n(\hat{\rho}_m - 1)$ when $\rho^0 = 1$.

6. Show that, under normality, the likelihood ratio statistic for testing $H_0\colon \alpha_1 = \alpha_1^0$ against $H_A\colon \alpha_1 \neq \alpha_1^0$ for the model (10.2.6) is a function of $\hat{\tau}$ defined in Corollary 10.2.1.2.

7. Plot the residuals from the model of Example 10.1.1 against time. Do you believe the original errors are identically distributed over time? Do you believe the original errors are normally distributed? Fit autoregressive models of orders 2, 3, and 4 to the square roots of the original observations. Plot the residuals from the second order model.

8. Let the assumptions of Theorem 10.3.1 hold for the model

$$Y_t = \beta_0 + \boldsymbol{\beta}'\mathbf{X}_{t-1} + \theta_1 Y_{t-1} + e_t .$$

Show that

$$t_{1\mu} \xrightarrow{\mathscr{L}} \rho_{ue}\hat{\tau}_{\mu l} + (1-\rho_{ue}^2)^{1/2} d ,$$

where

$$t_{1\mu} = [\hat{V}\{\hat{\theta}_1\}]^{-1/2}(\hat{\theta}_1 - 1) ,$$

$\hat{V}\{\hat{\theta}_1\}$ is the ordinary least squares estimator of the variance of the ordinary least squares estimator $\hat{\theta}_1$, $\hat{\tau}_\mu$ has the distribution of Table 10.A.2, $\hat{\tau}_\mu \xrightarrow{\mathscr{L}} \hat{\tau}_{\mu l}$, and d is a $N(0, 1)$ random variable independent of $\hat{\tau}_{\mu l}$.

9. Use the facts that $\hat{\mathbf{K}}'\hat{\mathbf{\Lambda}}\hat{\mathbf{K}} = \overset{*}{\mathbf{S}}_{hh}$ and $\hat{\mathbf{K}}'\mathbf{\Sigma}_{ee}\hat{\mathbf{K}} = \mathbf{I}$, where $\hat{\mathbf{\Lambda}} = \operatorname{diag}(\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_k)$, to show that the expression (10.3.47) for $\hat{\mathbf{B}}_{0,12}'$ is equivalent to that in (10.3.37) when $g = 1$.

10. Let the following model hold:

$$\begin{aligned} Y_{1t} &= \pi_1 Y_{2,t-1} + e_{1t} , \\ Y_{2t} &= Y_{2,t-1} + e_{2t} \end{aligned}$$

for $t = 2, 3, \ldots,$ where $Y_{20} = 0$ and $(e_{1t}, e_{2t})' \sim \mathrm{NI}(\mathbf{0}, \mathbf{\Sigma})$. Show that the maximum likelihood estimator of π_1 is given by the coefficient of $Y_{2,t-1}$ in the regression of Y_{1t} on $(Y_{2,t-1}, Y_{2t} - Y_{2,t-1})$. What is the coefficient of $Y_{2t} - Y_{2,t-1}$ estimating? Show that the limiting distribution of the usual regression t-statistic for the coefficient of $Y_{2,t-1}$ in the multiple regression is that of a $N(0, 1)$ random variable. See Phillips (1991).

11. Let model (10.1.25) hold with $\theta_1 = 1$ and assume $e_t \sim \mathrm{NI}(0, \sigma^2)$. Show that $\Delta\bar{Y}$ of (10.1.26) is uncorrelated with $\hat{\theta}_1 - 1$ but that $n^{1/2}\,\Delta\bar{Y}$ is not independent of $n(\hat{\theta}_1 - 1)$ in the limit.

12. Assume that (Z_t, Y_t) of Table 10.B.4 satisfy the model

$$\begin{aligned} Z_t &= \beta_{30} + \beta_{31} Y_t + e_{3t} , \\ Y_t &= \pi_{20} + \pi_{21} Y_{t-1} + \pi_{22}\,\Delta Y_{t-1} + e_{2t} , \end{aligned}$$

where $(e_{3t}, e_{2t})' \sim \mathrm{NI}(\mathbf{0}, \mathbf{\Sigma}_{ee})$, $\pi_{21} \in (-1, 1]$, and $\pi_{20} = 0$ if $\pi_{21} = 1$. Estimate the parameters of the model by ordinary least squares and by the system method of Example 10.3.1. Also estimate the parameters subject to the restriction that $(\pi_{20}, \pi_{21}) = (0, 1)$.

APPENDIX A

Appendix 10.A. Percentiles for Unit Root Distributions

Table 10.A.1. Empirical Cumulative Distribution of $n(\hat{\rho}-1)$ for $\rho = 1$

Sample Size n	Probability of a Smaller Value								
	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
					$\hat{\rho}$				
25	−11.8	−9.3	−7.3	−5.3	−0.82	1.01	1.41	1.78	2.28
50	−12.8	−9.9	−7.7	−5.5	−0.84	0.97	1.34	1.69	2.16
100	−13.3	−10.2	−7.9	−5.6	−0.85	0.95	1.31	1.65	2.09
250	−13.6	−10.4	−8.0	−5.7	−0.86	0.94	1.29	1.62	2.05
500	−13.7	−10.4	−8.0	−5.7	−0.86	0.93	1.29	1.61	2.04
∞	−13.7	−10.5	−8.1	−5.7	−0.86	0.93	1.28	1.60	2.03
					$\hat{\rho}_\mu$				
25	−17.2	−14.6	−12.5	−10.2	−4.22	−0.76	0.00	0.64	1.39
50	−18.9	−15.7	−13.3	−10.7	−4.29	−0.81	−0.07	0.53	1.22
100	−19.8	−16.3	−13.7	−11.0	−4.32	−0.83	−0.11	0.47	1.13
250	−20.3	−16.7	−13.9	−11.1	−4.34	−0.84	−0.13	0.44	1.08
500	−20.5	−16.8	−14.0	−11.2	−4.35	−0.85	−0.14	0.42	1.07
∞	−20.6	−16.9	−14.1	−11.3	−4.36	−0.85	−0.14	0.41	1.05
					$\hat{\rho}_\tau$				
25	−22.5	−20.0	−17.9	−15.6	−8.49	−3.65	−2.51	−1.53	−0.46
50	−25.8	−22.4	−19.7	−16.8	−8.80	−3.71	−2.60	−1.67	−0.67
100	−27.4	−23.7	−20.6	−17.5	−8.96	−3.74	−2.63	−1.74	−0.76
250	−28.5	−24.4	−21.3	−17.9	−9.05	−3.76	−2.65	−1.79	−0.83
500	−28.9	−24.7	−21.5	−18.1	−9.08	−3.76	−2.66	−1.80	−0.86
∞	−29.4	−25.0	−21.7	−18.3	−9.11	−3.77	−2.67	−1.81	−0.88

NOTE. This table was constructed by David A. Dickey using the Monte Carlo method. Details are given in Dickey (1976). Standard errors of the estimates vary, but most are less than 0.10 for entries in the left half of the table and less than 0.02 for entries in the right half of the table. Some entries differ slightly from those of the first edition.

Table 10.A.2. Empirical Cumulative Distribution of $\hat{\tau}$ for $\rho = 1$

Sample Size n	Probability of a Smaller Value								
	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
					$\hat{\tau}$				
25	−2.65	−2.26	−1.95	−1.60	−0.47	0.92	1.33	1.70	2.15
50	−2.62	−2.25	−1.95	−1.61	−0.49	0.91	1.31	1.66	2.08
100	−2.60	−2.24	−1.95	−1.61	−0.50	0.90	1.29	1.64	2.04
250	−2.58	−2.24	−1.95	−1.62	−0.50	0.89	1.28	1.63	2.02
500	−2.58	−2.23	−1.95	−1.62	−0.50	0.89	1.28	1.62	2.01
∞	−2.58	−2.23	−1.95	−1.62	−0.51	0.89	1.28	1.62	2.01
					$\hat{\tau}_\mu$				
25	−3.75	−3.33	−2.99	−2.64	−1.53	−0.37	0.00	0.34	0.71
50	−3.59	−3.23	−2.93	−2.60	−1.55	−0.41	−0.04	0.28	0.66
100	−3.50	−3.17	−2.90	−2.59	−1.56	−0.42	−0.06	0.26	0.63
250	−3.45	−3.14	−2.88	−2.58	−1.56	−0.42	−0.07	0.24	0.62
500	−3.44	−3.13	−2.87	−2.57	−1.57	−0.44	−0.07	0.24	0.61
∞	−3.42	−3.12	−2.86	−2.57	−1.57	−0.44	−0.08	0.23	0.60
					$\hat{\tau}_\tau$				
25	−4.38	−3.95	−3.60	−3.24	−2.14	−1.14	−0.81	−0.50	−0.15
50	−4.16	−3.80	−3.50	−3.18	−2.16	−1.19	−0.87	−0.58	−0.24
100	−4.05	−3.73	−3.45	−3.15	−2.17	−1.22	−0.90	−0.62	−0.28
250	−3.98	−3.69	−3.42	−3.13	−2.18	−1.23	−0.92	−0.64	−0.31
500	−3.97	−3.67	−3.42	−3.13	−2.18	−1.24	−0.93	−0.65	−0.32
∞	−3.96	−3.67	−3.41	−3.13	−2.18	−1.25	−0.94	−0.66	−0.32

NOTE. This table was constructed by David A. Dickey using the Monte Carlo method. Details are given in Dickey (1976). Standard errors of the estimates vary, but most are less than 0.014. Some entries differ slightly from those of the first edition.

Table 10.A.3. Cumulative Distribution of Simple Symmetric $\hat{\tau}_s$ for $\rho = 1$

Sample Size	Probability of a Smaller Value								
n	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
					No Adjustment				
25	−2.72	−2.35	−2.05	−1.74	−0.87	−0.43	−0.37	−0.33	−0.29
50	−2.71	−2.36	−2.08	−1.78	−0.90	−0.44	−0.38	−0.33	−0.29
100	−2.70	−2.37	−2.09	−1.79	−0.91	−0.45	−0.38	−0.34	−0.30
250	−2.70	−2.37	−2.10	−1.80	−0.92	−0.46	−0.39	−0.34	−0.30
500	−2.70	−2.37	−2.10	−1.80	−0.92	−0.46	−0.39	−0.34	−0.30
∞	−2.69	−2.37	−2.10	−1.81	−0.93	−0.46	−0.39	−0.34	−0.30
					Mean Removed				
25	−3.40	−3.02	−2.71	−2.37	−1.42	−0.83	−0.73	−0.65	−0.59
50	−3.28	−2.94	−2.66	−2.35	−1.44	−0.84	−0.73	−0.65	−0.58
100	−3.23	−2.90	−2.64	−2.34	−1.44	−0.84	−0.73	−0.65	−0.58
250	−3.20	−2.88	−2.62	−2.34	−1.45	−0.85	−0.73	−0.66	−0.58
500	−3.19	−2.88	−2.62	−2.33	−1.45	−0.85	−0.73	−0.66	−0.58
∞	−3.17	−2.87	−2.62	−2.33	−1.45	−0.85	−0.73	−0.66	−0.58
					Linear Trend Removed				
25	−4.19	−3.76	−3.45	−3.09	−2.10	−1.42	−1.28	−1.18	−1.07
50	−3.99	−3.36	−3.36	−3.04	−2.11	−1.44	−1.29	−1.18	−1.07
100	−3.89	−3.57	−3.31	−3.02	−2.12	−1.44	−1.30	−1.19	−1.07
250	−3.84	−3.54	−3.29	−3.01	−2.12	−1.45	−1.30	−1.19	−1.07
500	−3.82	−3.52	−3.28	−3.00	−2.12	−1.45	−1.30	−1.19	−1.07
∞	−3.80	−3.51	−3.27	−2.99	−2.12	−1.45	−1.30	−1.19	−1.07
					Quadratic Trend Removed				
25	−4.75	−4.30	−3.97	−3.61	−2.57	−1.84	−1.69	−1.57	−1.45
50	−4.50	−4.14	−3.85	−3.54	−2.58	−1.86	−1.70	−1.58	−1.45
100	−4.38	−4.05	−3.79	−3.50	−2.58	−1.87	−1.70	−1.58	−1.45
250	−4.30	−4.00	−3.76	−3.47	−2.58	−1.87	−1.70	−1.58	−1.44
500	−4.28	−3.99	−3.74	−3.47	−2.58	−1.87	−1.71	−1.58	−1.44
∞	−4.26	−3.97	−3.73	−3.46	−2.58	−1.87	−1.71	−1.58	−1.44

NOTE. This table was constructed by David A. Dickey.

Table 10.A.4. Cumulative Distribution of Weighted Symmetric $\hat{\tau}_w$ for $\rho = 1$

Sample Size n	Probability of a Smaller Value								
	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
					No Adjustment				
25	−2.73	−2.37	−2.09	−1.80	−1.05	−0.05	0.24	0.48	0.80
50	−2.74	−2.40	−2.13	−1.85	−1.09	−0.06	0.24	0.51	0.83
100	−2.74	−2.42	−2.16	−1.88	−1.10	−0.06	0.25	0.53	0.85
250	−2.75	−2.43	−2.17	−1.89	−1.12	−0.06	0.25	0.53	0.86
500	−2.75	−2.43	−2.18	−1.90	−1.12	−0.06	0.25	0.54	0.86
∞	−2.76	−2.44	−2.18	−1.90	−1.12	−0.07	0.26	0.54	0.86
					Mean removed				
25	−3.33	−2.92	−2.60	−2.26	−1.19	−0.07	0.25	0.51	0.84
50	−3.21	−2.85	−2.57	−2.25	−1.19	−0.04	0.30	0.58	0.93
100	−3.16	−2.82	−2.55	−2.24	−1.20	−0.02	0.32	0.62	0.98
250	−3.12	−2.80	−2.54	−2.23	−1.20	−0.01	0.34	0.64	1.01
500	−3.11	−2.80	−2.53	−2.23	−1.20	−0.00	0.34	0.65	1.01
∞	−3.10	−2.79	−2.52	−2.22	−1.20	−0.00	0.35	0.65	1.01
					Linear Trend Removed				
25	−4.11	−3.70	−3.37	−3.02	−1.98	−1.07	−0.82	−0.60	−0.35
50	−3.93	−3.57	−3.28	−2.96	−1.96	−1.01	−0.72	−0.48	−0.19
100	−3.84	−3.51	−3.24	−2.94	−1.96	−0.97	−0.68	−0.42	−0.11
250	−3.78	−3.47	−3.21	−2.92	−1.95	−0.95	−0.65	−0.38	−0.06
500	−3.76	−3.45	−3.20	−2.91	−1.95	−0.95	−0.64	−0.37	−0.05
∞	−3.75	−3.45	−3.19	−2.91	−1.94	−0.94	−0.63	−0.36	−0.03

NOTE. This table was constructed by Heon Jin Park.

Table 10.A.5. Cumulative Distribution of Maximum Likelihood $\hat{\tau}_m$ for $\rho = 1$

Sample Size n	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
					No Adjustment				
25	−2.80	−2.44	−2.16	−1.86	−0.99	−0.40	−0.32	−0.27	−0.23
50	−2.78	−2.44	−2.17	−1.87	−1.00	−0.40	−0.32	−0.27	−0.22
100	−2.77	−2.44	−2.17	−1.88	−1.00	−0.39	−0.32	−0.27	−0.22
250	−2.76	−2.44	−2.17	−1.88	−1.00	−0.39	−0.32	−0.27	−0.22
500	−2.75	−2.44	−2.17	−1.88	−1.00	−0.39	−0.31	−0.27	−0.22
∞	−2.75	−2.44	−2.17	−1.88	−1.00	−0.39	−0.31	−0.27	−0.22
					Mean Removed				
25	−3.52	−3.11	−2.78	−2.44	−1.44	−0.65	−0.51	−0.42	−0.35
50	−3.33	−2.98	−2.69	−2.38	−1.42	−0.64	−0.51	−0.42	−0.34
100	−3.24	−2.92	−2.64	−2.34	−1.41	−0.64	−0.50	−0.42	−0.34
250	−3.19	−2.88	−2.61	−2.32	−1.40	−0.64	−0.50	−0.41	−0.34
500	−3.17	−2.86	−2.60	−2.31	−1.39	−0.64	−0.50	−0.41	−0.34
∞	−3.16	−2.85	−2.59	−2.30	−1.38	−0.63	−0.50	−0.41	−0.34
					Linear Trend Removed				
25	−4.40	−3.97	−3.62	−3.25	−2.17	−1.35	−1.15	−1.00	−0.87
50	−4.07	−3.72	−3.43	−3.11	−2.12	−1.34	−1.14	−1.00	−0.86
100	−3.92	−3.60	−3.34	−3.04	−2.10	−1.33	−1.14	−0.99	−0.86
250	−3.84	−3.54	−3.28	−3.00	−2.08	−1.32	−1.13	−0.99	−0.85
500	−3.82	−3.52	−3.27	−2.98	−2.08	−1.32	−1.13	−0.99	−0.85
∞	−3.79	−3.50	−3.25	−2.97	−2.07	−1.31	−1.13	−0.99	−0.85

NOTE. This table was constructed by Heon Jin Park.

Table 10.A.6. Vector Ordinary Least Squares: No Adjustment

			Probability of a Smaller Value								
D.f.	H_0	H_1	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
23	1	0	0.00	0.00	0.01	0.02	0.58	2.70	3.68	4.62	5.80
22	2	0	1.18	1.49	1.84	2.31	4.75	8.35	9.53	10.60	11.89
	2	1	1.01	1.28	1.58	1.99	4.16	7.42	8.48	9.47	10.59
21	3	0	5.38	6.11	6.80	7.67	11.34	15.76	17.09	18.34	19.83
	3	1	5.11	5.83	6.50	7.32	10.84	15.06	16.34	17.49	18.86
	3	2	3.34	3.82	4.30	4.90	7.48	10.63	11.58	12.37	13.32
20	4	0	12.06	13.11	14.04	15.19	19.62	24.63	26.10	27.47	29.00
	4	1	11.78	12.82	13.72	14.84	19.15	24.03	25.46	26.75	28.29
	4	2	9.78	10.67	11.46	12.44	16.17	20.27	21.43	22.51	23.70
	4	3	5.81	6.40	6.92	7.55	10.05	12.77	13.54	14.19	14.91
48	1	0	0.00	0.00	0.01	0.02	0.59	2.83	3.90	4.96	6.36
47	2	0	1.21	1.56	1.93	2.43	5.12	9.35	10.84	12.23	13.90
	2	1	1.04	1.34	1.65	2.10	4.50	8.39	9.79	11.02	12.67
46	3	0	5.80	6.64	7.43	8.44	12.84	18.53	20.38	22.06	24.13
	3	1	5.53	6.35	7.10	8.06	12.31	17.84	19.62	21.24	23.19
	3	2	3.61	4.17	4.73	5.42	8.62	12.95	14.38	15.70	17.23
45	4	0	13.55	14.85	16.04	17.47	23.31	30.26	32.42	34.36	36.67
	4	1	13.26	14.53	15.68	17.10	22.81	29.65	31.80	33.67	35.88
	4	2	11.08	12.15	13.19	14.43	19.49	25.56	27.40	29.07	31.01
	4	3	6.62	7.34	8.02	8.87	12.41	16.82	18.19	19.38	20.83
98	1	0	0.00	0.00	0.01	0.02	0.59	2.91	4.03	5.16	6.69
97	2	0	1.23	1.59	1.97	2.49	5.31	9.91	11.56	13.15	15.08
	2	1	1.06	1.37	1.69	2.15	4.67	8.91	10.49	11.98	13.84
96	3	0	6.00	6.89	7.74	8.83	13.62	20.08	22.23	24.21	26.64
	3	1	5.73	6.59	7.39	8.44	13.09	19.37	21.46	23.40	25.74
	3	2	3.74	4.34	4.94	5.69	9.23	14.29	16.00	17.59	19.56
95	4	0	14.33	15.76	17.06	18.66	25.32	33.50	36.10	38.44	41.23
	4	1	14.01	15.41	16.68	18.26	24.81	32.86	35.44	37.74	40.48
	4	2	11.72	12.93	14.07	15.46	21.32	28.61	30.94	33.02	35.46
	4	3	7.05	7.84	8.60	9.56	13.71	19.19	20.97	22.61	24.61

Table 10.A.6.—*Continued*

			Probability of a Smaller Value								
D.f.	H_0	H_1	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
248	1	0	0.00	0.00	0.01	0.02	0.60	2.95	4.10	5.27	6.87
247	2	0	1.25	1.61	1.99	2.53	5.42	10.26	12.01	13.73	15.87
	2	1	1.07	1.38	1.71	2.18	4.77	9.25	10.93	12.58	14.59
246	3	0	6.12	7.03	7.92	9.05	14.11	21.07	23.42	25.60	28.28
	3	1	5.83	6.71	7.56	8.65	13.57	20.35	22.65	24.79	27.41
	3	2	3.82	4.44	5.06	5.85	9.61	15.15	17.05	18.83	21.11
245	4	0	14.80	16.32	17.69	19.39	26.57	35.59	38.49	41.15	44.27
	4	1	14.46	15.94	17.30	18.98	26.06	34.95	37.81	40.45	43.55
	4	2	12.11	13.42	14.62	16.10	22.47	30.59	33.25	35.66	38.48
	4	3	7.31	8.15	8.96	9.99	14.54	20.74	22.83	24.79	27.21
498	1	0	0.00	0.00	0.01	0.02	0.60	2.96	4.12	5.30	6.91
497	2	0	1.25	1.61	2.00	2.54	5.46	10.37	12.16	13.92	16.14
	2	1	1.08	1.39	1.72	2.19	4.80	9.36	11.08	12.75	14.86
496	3	0	6.16	7.07	7.98	9.12	14.28	21.42	23.84	26.07	28.85
	3	1	5.86	6.75	7.61	8.72	13.73	20.69	23.06	25.27	27.98
	3	2	3.84	4.47	5.09	5.90	9.74	15.43	17.42	19.28	21.66
495	4	0	14.95	16.50	17.90	19.65	26.99	36.31	39.32	42.10	45.34
	4	1	14.62	16.12	17.51	19.23	26.48	35.67	38.64	41.42	44.62
	4	2	12.23	13.57	14.80	16.33	22.86	31.27	34.04	36.58	39.56
	4	3	7.40	8.27	9.09	10.14	14.82	21.28	23.49	25.55	28.13
∞	1	0	0.00	0.00	0.01	0.02	0.60	2.97	4.13	5.31	6.94
	2	0	1.25	1.62	2.01	2.55	5.49	10.48	12.31	14.10	16.42
	2	1	1.08	1.39	1.72	2.20	4.83	9.48	11.23	12.88	15.14
	3	0	6.19	7.11	8.03	9.18	14.46	21.78	24.26	26.54	29.41
	3	1	5.89	6.78	7.66	8.77	13.91	21.06	23.47	25.75	28.56
	3	2	3.87	4.49	5.13	5.95	9.88	15.71	17.82	19.77	22.22
	4	0	15.10	16.67	18.12	19.91	27.41	37.04	40.17	43.08	46.47
	4	1	14.78	16.29	17.73	19.50	26.89	36.40	39.51	42.44	45.69
	4	2	12.36	13.72	15.00	16.56	23.25	31.96	34.81	37.54	40.71
	4	3	7.50	8.39	9.21	10.31	15.11	21.83	24.18	26.31	29.07

NOTE. This table was constructed by Heon Jin Park. See equation (10.3.48) for the definition of the statistics.

Table 10.A.7. Vector Ordinary Least Squares: Mean Removed

			Probability of a Smaller Value								
D.f.	H_0	H_1	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
22	1	0	0.00	0.02	0.07	0.24	2.13	5.20	6.25	7.23	8.44
21	2	0	2.51	2.99	3.51	4.18	7.32	11.27	12.57	13.68	15.03
	2	1	2.07	2.46	2.85	3.38	5.79	8.92	9.89	10.75	11.75
20	3	0	7.90	8.74	9.50	10.46	14.38	18.93	20.30	21.53	23.03
	3	1	7.33	8.11	8.80	9.67	13.18	17.17	18.36	19.43	20.65
	3	2	4.53	5.05	5.53	6.13	8.59	11.41	12.23	12.93	13.73
19	4	0	15.22	16.27	17.23	18.37	22.74	27.65	29.10	30.41	32.04
	4	1	14.62	15.63	16.53	17.60	21.69	26.18	27.50	28.66	29.99
	4	2	11.78	12.62	13.37	14.28	17.69	21.34	22.41	23.30	24.32
	4	3	6.79	7.33	7.82	8.39	10.64	13.02	13.65	14.19	14.81
47	1	0	0.00	0.02	0.07	0.25	2.30	5.88	7.20	8.43	9.99
46	2	0	2.69	3.28	3.85	4.63	8.33	13.42	15.07	16.57	18.40
	2	1	2.23	2.68	3.13	3.74	6.65	10.80	12.25	13.51	15.10
45	3	0	8.92	9.93	10.91	12.13	17.22	23.46	25.41	27.19	29.30
	3	1	8.28	9.23	10.11	11.24	15.87	21.60	23.36	24.98	26.90
	3	2	5.15	5.81	6.43	7.20	10.54	14.88	16.23	17.44	18.92
44	4	0	18.02	19.46	20.75	22.33	28.58	35.83	38.06	40.07	42.40
	4	1	17.33	18.74	19.95	21.45	27.42	34.28	36.36	38.28	40.49
	4	2	14.05	15.23	16.28	17.57	22.72	28.65	30.48	32.09	33.99
	4	3	8.19	8.96	9.65	10.50	14.01	18.23	19.54	20.65	22.10
97	1	0	0.00	0.02	0.07	0.26	2.38	6.26	7.72	9.10	10.82
96	2	0	2.77	3.40	4.03	4.86	8.89	14.60	16.53	18.29	20.50
	2	1	2.31	2.79	3.27	3.93	7.12	11.90	13.59	15.11	17.07
95	3	0	9.44	10.58	11.66	13.03	18.79	26.10	28.45	30.61	33.26
	3	1	8.78	9.84	10.82	12.08	17.38	24.18	26.36	28.37	30.83
	3	2	5.48	6.21	6.91	7.77	11.65	16.93	18.67	20.30	22.26
94	4	0	19.47	21.12	22.61	24.47	31.87	40.67	43.45	45.94	48.96
	4	1	18.75	20.34	21.76	23.53	30.65	39.07	41.72	44.14	47.05
	4	2	15.24	16.59	17.83	19.35	25.59	33.08	35.44	37.58	40.13
	4	3	8.93	9.81	10.64	11.66	15.99	21.50	23.29	24.91	26.97

Table 10.A.7.—*Continued*

D.f.	H_0	H_1	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
			Probability of a Smaller Value								
247	1	0	0.00	0.02	0.07	0.26	2.43	6.47	8.01	9.48	11.34
246	2	0	2.82	3.47	4.13	5.00	9.23	15.36	17.47	19.42	21.93
	2	1	2.35	2.86	3.36	4.04	7.41	12.60	14.44	16.17	18.38
245	3	0	9.76	10.99	12.13	13.58	19.77	27.83	30.45	32.88	35.92
	3	1	9.08	10.21	11.26	12.60	18.33	25.85	28.33	30.63	33.49
	3	2	5.69	6.46	7.20	8.13	12.35	18.29	20.32	22.24	24.57
244	4	0	20.38	22.15	23.78	25.81	33.97	43.88	47.05	49.89	53.48
	4	1	19.64	21.35	22.90	24.85	32.71	42.25	45.32	48.09	51.53
	4	2	15.99	17.45	18.81	20.50	27.44	36.03	38.79	41.32	44.39
	4	3	9.41	10.36	11.27	12.40	17.28	23.73	25.91	27.93	30.44
497	1	0	0.00	0.02	0.07	0.26	2.45	6.52	8.10	9.59	11.53
496	2	0	2.84	3.49	4.15	5.04	9.33	15.62	17.79	19.82	22.43
	2	1	2.36	2.88	3.39	4.07	7.50	12.83	14.72	16.54	18.85
495	3	0	9.86	11.13	12.28	13.75	20.10	28.43	31.15	33.68	36.84
	3	1	9.19	10.34	11.41	12.76	18.65	26.43	29.02	31.41	34.42
	3	2	5.76	6.54	7.30	8.25	12.59	18.77	20.91	22.92	25.38
494	4	0	20.69	22.51	24.19	26.27	34.70	45.00	48.32	51.30	55.09
	4	1	19.96	21.69	23.30	25.30	33.42	43.37	46.59	49.48	53.13
	4	2	16.25	17.75	19.15	20.90	28.08	37.07	39.98	42.65	45.93
	4	3	9.58	10.56	11.49	12.65	17.72	24.53	26.85	29.02	31.70
∞	1	0	0.00	0.02	0.07	0.26	2.46	6.56	8.16	9.68	11.74
	2	0	2.86	3.51	4.17	5.07	9.41	15.89	18.11	20.21	22.94
	2	1	2.36	2.88	3.41	4.10	7.59	13.05	14.99	16.93	19.36
	3	0	9.97	11.25	12.45	13.91	20.42	29.03	31.85	34.49	37.73
	3	1	9.29	10.46	11.56	12.92	18.96	27.02	29.70	32.18	35.32
	3	2	5.83	6.62	7.39	8.37	12.83	19.27	21.54	23.59	26.19
	4	0	21.04	22.89	24.63	26.74	35.44	46.17	49.62	52.77	56.73
	4	1	20.31	22.07	23.73	25.76	34.14	44.55	47.90	50.91	54.79
	4	2	16.54	18.09	19.52	21.33	28.72	38.12	41.20	44.00	47.55
	4	3	9.78	10.78	11.73	12.92	18.17	25.33	27.85	30.17	33.01

NOTE. This table was constructed by Heon Jin Park. See equation (10.3.48) for the definition of the statistics.

Table 10.A.8. Vector Ordinary Least Squares: Linear Trend Removed

			Probability of a Smaller Value								
D.f.	H_0	H_1	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
21	1	0	0.08	0.33	0.71	1.32	3.89	7.19	8.24	9.22	10.32
20	2	0	4.20	4.93	5.62	6.49	10.09	14.22	15.44	16.51	17.78
	2	1	3.24	3.73	4.20	4.79	7.31	10.30	11.18	11.96	12.83
19	3	0	10.65	11.62	12.49	13.56	17.72	22.26	23.62	24.78	26.18
	3	1	9.56	10.38	11.14	12.04	15.50	19.26	20.33	21.27	22.39
	3	2	5.73	6.26	6.75	7.35	9.70	12.22	12.95	13.54	14.21
18	4	0	18.56	19.68	20.64	21.84	26.30	31.04	32.44	33.67	35.12
	4	1	17.52	18.51	19.40	20.47	24.41	28.53	29.74	30.79	31.96
	4	2	13.77	14.55	15.28	16.17	19.37	22.62	23.56	24.34	25.22
	4	3	7.79	8.28	8.74	9.30	11.34	13.39	13.93	14.38	14.87
46	1	0	0.11	0.40	0.80	1.43	4.32	8.44	9.86	11.13	12.76
45	2	0	4.71	5.56	6.41	7.46	11.95	17.50	19.31	20.90	22.79
	2	1	3.57	4.16	4.75	5.48	8.74	13.06	14.46	15.75	17.29
44	3	0	12.52	13.79	14.96	16.38	22.08	28.75	30.81	32.67	34.83
	3	1	11.18	12.30	13.30	14.51	19.47	25.33	27.13	28.77	30.70
	3	2	6.79	7.49	8.16	8.98	12.46	16.71	18.07	19.31	20.78
43	4	0	22.95	24.58	25.98	27.73	34.47	42.06	44.25	46.18	48.65
	4	1	21.60	23.05	24.39	26.00	32.18	39.12	41.20	43.01	45.21
	4	2	17.06	18.31	19.46	20.82	26.03	31.92	33.69	35.21	37.14
	4	3	9.77	10.59	11.31	12.17	15.69	19.81	21.05	22.17	23.45
96	1	0	0.13	0.43	0.86	1.50	4.54	9.09	10.72	12.23	14.13
95	2	0	4.96	5.89	6.79	7.94	12.95	19.36	21.46	23.41	25.75
	2	1	3.76	4.39	5.02	5.84	9.51	14.62	16.36	17.98	20.02
94	3	0	13.48	14.88	16.22	17.85	24.49	32.53	35.06	37.37	40.12
	3	1	12.02	13.26	14.40	15.82	21.69	28.90	31.22	33.31	35.84
	3	2	7.32	8.12	8.89	9.85	14.04	19.47	21.28	22.90	24.95
93	4	0	25.21	27.12	28.81	30.92	39.08	48.57	51.44	54.05	57.18
	4	1	23.71	25.44	27.06	29.01	36.60	45.50	48.22	50.68	53.57
	4	2	18.80	20.32	21.70	23.35	29.88	37.63	40.07	42.16	44.76
	4	3	10.85	11.83	12.71	13.78	18.26	23.89	25.69	27.33	29.28

Table 10.A.8.—*Continued*

			Probability of a Smaller Value								
D.f.	H_0	H_1	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
246	1	0	0.14	0.44	0.88	1.54	4.68	9.49	11.26	12.93	15.03
245	2	0	5.10	6.08	7.02	8.22	13.56	20.57	22.87	25.06	27.74
	2	1	3.86	4.53	5.18	6.05	9.99	15.66	17.63	19.48	21.85
244	3	0	14.05	15.55	16.98	18.75	26.02	35.01	37.89	40.52	43.72
	3	1	12.53	13.84	15.08	16.63	23.10	31.26	33.95	36.37	39.36
	3	2	7.63	8.51	9.34	10.40	15.05	21.33	23.47	25.39	27.84
243	4	0	26.61	28.71	30.58	32.92	42.05	52.92	56.31	59.42	63.11
	4	1	25.01	26.95	28.73	30.89	39.47	49.76	52.99	55.92	59.43
	4	2	19.91	21.59	23.11	24.94	32.38	41.49	44.42	47.00	50.16
	4	3	11.54	12.61	13.59	14.81	19.96	26.71	28.96	31.02	33.54
496	1	0	0.14	0.45	0.89	1.55	4.72	9.62	11.44	13.17	15.34
495	2	0	5.15	6.13	7.09	8.32	13.76	20.99	23.39	25.66	28.46
	2	1	3.89	4.57	5.24	6.11	10.15	16.03	18.09	20.03	22.49
494	3	0	14.23	15.77	17.23	19.05	26.55	35.87	38.88	41.64	45.02
	3	1	12.70	14.04	15.30	16.91	23.59	32.10	34.91	37.45	40.63
	3	2	7.72	8.64	9.49	10.58	15.40	21.98	24.25	26.28	28.89
493	4	0	27.09	29.26	31.18	33.61	43.09	54.47	58.05	61.33	65.28
	4	1	25.46	27.48	29.30	31.53	40.46	51.26	54.69	57.80	61.59
	4	2	20.31	22.03	23.58	25.49	33.26	42.87	45.95	48.74	52.18
	4	3	11.79	12.88	13.89	15.16	20.56	27.73	30.15	32.38	35.13
∞	1	0	0.14	0.45	0.89	1.55	4.77	9.77	11.63	13.41	15.67
	2	0	5.19	6.16	7.16	8.41	13.95	21.44	24.00	26.32	29.24
	2	1	3.91	4.60	5.29	6.17	10.30	16.43	18.59	20.64	23.15
	3	0	14.40	16.00	17.47	19.36	27.08	36.76	39.92	42.83	46.41
	3	1	12.85	14.23	15.54	17.19	24.08	32.97	35.92	38.57	41.97
	3	2	7.82	8.77	9.64	10.78	15.75	22.64	25.05	27.25	30.00
	4	0	27.60	29.86	31.81	34.31	44.16	56.11	59.88	63.27	67.64
	4	1	25.94	28.06	29.89	32.18	41.48	52.80	56.47	59.74	63.97
	4	2	20.75	22.50	24.06	26.05	34.15	44.30	47.50	50.57	54.44
	4	3	12.06	13.16	14.19	15.51	21.19	28.78	31.42	33.81	36.88

NOTE. This table was constructed by Heon Jin Park. See equation (10.3.48) for the definition of the statistics.

Table 10.A.9. Percentiles of Test Statistics for Test of Unit Root of Moving Average

Sample Size n	Probability of a Smaller Value 0.00	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
					$\overset{*}{\tau}$ of (10.4.8)				
25	−3.87	−3.47	−3.16	−2.83	−1.88	−1.28	−1.16	−1.08	−1.00
50	−3.71	−3.37	−3.09	−2.79	−1.89	−1.28	−1.16	−1.07	−0.98
100	−3.64	−3.32	−3.06	−2.77	−1.89	−1.28	−1.16	−1.06	−0.97
250	−3.59	−3.29	−3.04	−2.76	−1.90	−1.28	−1.16	−1.06	−0.97
500	−3.58	−3.28	−3.03	−2.75	−1.90	−1.28	−1.16	−1.06	−0.97
750	−3.57	−3.28	−3.02	−2.75	−1.90	−1.28	−1.16	−1.06	−0.96
∞	−3.56	−3.27	−3.02	−2.75	−1.90	−1.28	−1.16	−1.06	−0.96
					$\overset{*}{\tau}_{\mu}$ of (10.4.15)				
25	−4.51	−4.09	−3.75	−3.40	−2.39	−1.69	−1.55	−1.45	−1.34
50	−4.28	−3.93	−3.64	−3.33	−2.39	−1.70	−1.56	−1.45	−1.34
100	−4.17	−3.85	−3.58	−3.29	−2.39	−1.71	−1.56	−1.45	−1.33
250	−4.10	−3.80	−3.55	−3.27	−2.40	−1.72	−1.56	−1.45	−1.33
500	−4.08	−3.78	−3.54	−3.27	−2.40	−1.72	−1.56	−1.45	−1.33
750	−4.07	−3.78	−3.53	−3.26	−2.40	−1.72	−1.57	−1.45	−1.33
∞	−4.06	−3.77	−3.53	−3.26	−2.40	−1.72	−1.57	−1.45	−1.33

NOTE. This table was constructed by David A. Dickey.

APPENDIX B

Appendix 10.B. Data Used in Examples

Table 10.B.1. Seasonally Adjusted Interest Rates

Obs.	U1	U2	U3
1	4.12808	4.39345	5.02055
2	4.11977	4.10203	4.47400
3	4.02260	3.39690	3.82110
4	4.08650	3.23849	3.62040
5	3.89452	3.33345	3.65055
6	3.29188	2.52677	3.15180
7	3.13677	2.25655	2.95945
8	2.78538	2.15797	2.64600
9	2.41740	2.39310	2.71890
10	2.34835	2.29151	2.84960
11	2.35063	2.32655	2.79945
12	2.00812	2.18323	2.58820
13	1.58808	2.28345	2.70055
14	2.68977	2.56203	2.92400
15	2.20260	2.47690	2.90110
16	1.66650	2.29849	2.81040
17	2.02452	2.33345	2.79055
18	1.70188	2.39677	2.85180
19	1.06678	2.19655	2.71945
20	1.80538	2.24797	2.73600
21	1.69740	2.19310	2.73890
22	2.13835	2.29151	2.82960
23	2.53062	2.43655	2.82945
24	2.35812	2.53323	2.91820
25	2.27808	2.76345	3.26055
26	2.51977	2.87203	3.38400
27	2.88260	2.80690	3.12110
28	2.85650	2.73849	2.97040
29	2.33452	2.72345	2.98055

Table 10.B.1.—*Continued*

Obs.	U1	U2	U3
30	2.65188	2.79677	2.94180
31	2.61678	2.87655	3.09945
32	2.73538	2.67797	2.92600
33	2.71740	2.69310	2.84890
34	2.77835	2.73151	2.82960
35	2.85062	2.78655	2.86945
36	2.95812	2.80323	2.88820
37	3.04808	2.95345	3.07055
38	3.14977	3.06203	3.17399
39	3.16260	2.97690	3.11110
40	3.06651	3.90849	3.10040
41	3.04452	2.96345	3.13055
42	2.96188	3.05677	3.16180
43	2.92678	3.13656	3.32946
44	3.29538	3.17797	3.32601
45	3.29740	3.29310	3.42890
46	3.37835	3.44151	3.53960
47	3.39062	3.47655	3.59945
48	2.40811	3.45323	3.63820
49	3.61808	3.56344	3.75054
50	3.62977	3.67203	3.88399
51	3.61260	3.62690	3.92110
52	3.63651	3.47849	3.82040
53	3.54453	3.52345	3.78055
54	3.47188	3.54677	3.75180
55	3.32678	3.41656	3.56946
56	3.30538	3.35797	3.49601
57	3.26740	3.44309	3.58890
58	3.23835	3.56151	3.71960
59	3.43062	3.59655	3.78945
60	3.87811	3.77323	3.90820
61	4.03808	3.85344	3.98054
62	4.12977	4.07203	4.17399
63	4.22260	4.01691	4.16110
64	4.25651	3.93849	4.07040
65	4.14453	3.93345	4.03055
66	4.01188	3.86677	3.94180
67	3.99678	3.79656	3.81946
68	3.92538	3.69797	3.78601
69	3.82740	3.83309	3.94890
70	3.95835	4.02151	4.08960
71	4.01062	4.04655	4.15945
72	4.34811	4.31323	4.50820
73	4.55808	4.63344	4.76054
74	4.74977	4.79203	4.98399

Table 10.B.1.—*Continued*

Obs.	U1	U2	U3
75	4.83260	4.67691	4.95111
76	4.83651	4.62849	4.81040
77	4.94453	4.68345	4.92055
78	5.14188	4.56677	4.83180
79	5.20678	4.75656	4.86946
80	5.33538	4.81797	5.16601
81	5.21740	5.28309	5.65889
82	5.40835	5.34151	5.44960
83	5.68062	5.27655	5.41945
84	5.42811	4.89323	4.94819
85	5.07808	4.76344	4.68054
86	5.14977	4.70203	4.74399
87	4.71260	4.34691	4.32111
88	4.21651	3.84849	3.97040
89	3.98453	3.64345	3.95055
90	3.95188	3.60677	4.21181
91	3.69678	4.16656	4.82946
92	3.69538	4.12797	4.87601
93	3.81739	4.33309	4.95889
94	3.75835	4.55151	5.14960
95	4.03062	4.68656	5.31945
96	4.53811	4.90323	5.53819
97	4.73808	5.04344	5.37054
98	4.86977	5.12203	5.41399
99	5.23260	5.25691	5.56111
100	5.92651	5.38849	5.54040
101	6.16453	5.70344	5.91055
102	6.04188	5.58677	5.73181
103	5.92678	5.26656	5.31946
104	5.83538	4.94797	4.97601
105	5.59739	5.10309	5.03889
106	5.79835	5.34151	5.25960
107	5.72062	5.40656	5.43945
108	6.04811	5.89323	5.91819
109	6.43807	6.18344	6.14054
110	6.78977	6.26203	6.38399
111	6.97260	6.10691	6.34111
112	7.57651	6.11849	6.11040
113	8.71453	6.08344	6.18055
114	8.87188	6.50677	6.93181
115	8.51678	6.95656	7.11946
116	8.99538	6.83797	7.10601
117	8.96739	7.00309	7.20889
118	8.87835	6.99151	7.13960
119	8.76062	7.19656	7.28945

Table 10.B.1.—*Continued*

Obs.	U1	U2	U3
120	8.99810	7.75323	7.56819
121	9.11807	7.91344	7.58054
122	9.12977	7.27203	7.22399
123	7.94260	6.71691	6.64111
124	8.26651	6.51849	6.60040
125	7.98453	6.88344	7.19055
126	7.57188	6.74677	7.12181
127	7.11678	6.40657	6.54947
128	6.41538	6.26797	6.37601
129	6.10739	6.04309	6.24889
130	6.07835	5.90151	6.15960
131	5.51062	5.23656	5.31945
132	4.92810	4.80322	4.78819
133	4.27807	4.48343	4.47053
134	3.86977	3.84203	4.01399
135	3.89260	3.46691	3.74111
136	4.31651	3.86849	4.16040
137	4.67453	4.18344	4.71055
138	4.88188	4.81678	5.38181
139	5.21679	5.35657	5.66947
140	5.37538	4.79798	5.34601
141	5.36739	4.60309	5.04889
142	5.07835	4.45151	4.67960
143	4.82062	4.17656	4.41945
144	4.16810	3.94322	4.34819
145	3.63807	3.42343	3.89053
146	3.43977	3.34202	4.23399
147	4.0126	3.81691	4.57111
148	4.3365	3.71849	4.72040
149	4.3145	3.73344	4.53055
150	4.4319	3.97678	4.76181
151	4.4568	3.93657	4.82947
152	4.6054	3.87798	4.72601
153	4.6874	4.57309	5.29889
154	4.9183	4.73151	5.31960
155	4.9706	4.73656	5.12945
156	5.3581	5.00322	5.22819
157	6.0781	5.45343	5.65053
158	6.7298	5.74202	6.10399
159	7.2726	6.17691	6.67111
160	7.2865	6.26849	6.58040
161	7.8845	6.40344	6.70055
162	8.4619	7.25678	7.10181
163	10.3068	7.96657	7.89947
164	10.3054	8.52798	8.14601

Table 10.B.1.—*Continued*

Obs.	U1	U2	U3
165	10.5974	8.20309	7.92889
166	9.8883	7.21151	7.09960
167	9.9406	7.78656	7.32945
168	9.9781	7.38322	6.95819
169	9.7881	7.81343	7.08053
170	9.1198	7.26202	6.68399
171	9.5326	8.04692	7.48111
172	10.6765	8.33849	8.15040
173	11.3545	8.27344	8.28055
174	11.9019	7.96678	8.21181
175	12.8268	7.50657	7.96947
176	11.8154	8.81798	8.70601
177	11.1574	7.97308	8.37889
178	9.9383	7.45151	7.51960
179	9.3606	7.42656	7.21945
180	8.5581	7.08322	6.73819
181	7.2681	6.30343	6.34053
182	6.3898	5.64202	5.73399
183	5.7226	5.57692	5.84111
184	5.6565	5.61849	6.47040
185	5.2645	5.27344	5.98055
186	5.5219	5.40678	5.91181
187	6.0068	6.08657	6.56947
188	5.9454	6.29798	6.98601
189	6.0574	6.33308	7.05889
190	5.6983	5.95151	6.40960
191	5.1306	5.43656	5.99945
192	5.2281	5.37322	6.10819
193	5.0081	4.91343	5.15053
194	4.9198	5.02202	5.70399
195	5.0226	5.08692	5.96111
196	4.9865	4.86849	5.61040
197	5.3345	5.24344	6.05055
198	5.4519	5.47678	6.17181
199	5.2168	5.18657	5.74947
200	5.0954	4.99798	5.46601
201	5.0674	4.99308	5.35889
202	4.9083	4.91151	5.11960
203	4.8606	4.70656	4.92945
204	4.6781	4.28322	4.58819
205	4.7481	4.66343	5.07052
206	4.8298	4.81202	5.33399
207	4.8726	4.68692	5.33111
208	4.8965	4.54849	5.17040
209	5.3945	5.00344	5.50055

Table 10.B.1.—*Continued*

Obs.	U1	U2	U3
210	5.3619	5.08678	5.46181
211	5.3268	5.14658	5.49948
212	5.7054	5.34798	5.79601
213	5.9574	5.72308	5.98889
214	6.3483	6.15151	6.44960
215	6.4206	6.05656	6.44945
216	6.5881	6.00322	6.46819
217	6.8381	6.48342	6.87052
218	6.9298	6.59202	7.03399
219	6.9726	6.37692	6.96112
220	7.0565	6.29849	7.03040
221	7.4045	6.45344	7.35055
222	7.5719	6.79678	7.58181
223	7.7168	6.96658	7.71948
224	7.8454	6.93798	7.55601
225	8.2674	7.76308	7.86888
226	8.8383	7.98151	8.37960
227	9.6706	8.59656	9.12945
228	10.0581	9.01322	9.38819
229	10.2081	9.39342	9.61052
230	10.2098	9.46202	9.56399
231	10.2726	9.56692	9.52112
232	10.1765	9.46849	9.35040
233	10.2845	9.65344	9.34055
234	10.2619	9.12678	8.86181
235	10.3768	9.19658	8.79948
236	10.7454	9.37798	8.98601

SOURCE. *Banking and Monetary Statistics: 1941–1970*, Federal Reserve Systems, Washington, D.C., and *Annual Statistical Digest* (1970–1979), Federal Reserve Systems, Washington, D.C. These data differ from those in the original sources because they are seasonally adjusted in a different manner.

Table 10.B.2. Artificial Autoregressive Moving Average Data

Time	Observation	Time	Observation	Time	Observation
1	−0.589	35	−4.546	68	−5.698
2	−0.665	36	−5.403	69	−3.757
3	1.655	37	−4.411	70	−1.778
4	2.689	38	−3.709	71	−0.637
5	1.695	39	−0.834	72	1.473
6	−2.183	40	2.498	73	3.078
7	−6.983	41	3.103	74	3.157
8	−10.053	42	−0.698	75	3.902
9	−10.566	43	−6.606	76	5.557
10	−8.129	44	−10.295	77	6.295
11	−2.242	45	−8.954	78	4.099
12	4.792	46	−4.039	79	−0.518
13	9.154	47	2.415	80	−5.132
14	9.521	48	8.275	81	−7.221
15	6.793	49	12.414	82	−6.819
16	2.258	50	13.220	83	−5.023
17	−2.285	51	9.959	84	−3.389
18	−4.200	52	3.281	85	−1.546
19	−2.712	53	−3.934	86	1.469
20	−0.547	54	−8.842	87	3.959
21	0.389	55	−9.985	88	3.387
22	0.731	56	−7.533	89	0.777
23	−1.075	57	−3.458	90	−2.547
24	−2.228	58	2.159	91	−5.955
25	−2.305	59	8.960	92	−8.043
26	−0.519	60	14.563	93	−5.394
27	1.788	61	15.020	94	−0.869
28	4.267	62	9.890	95	2.738
29	5.913	63	3.059	96	3.708
30	5.483	64	−3.910	97	3.245
31	3.745	65	−7.551	98	−0.757
32	2.017	66	−8.689	99	−5.353
33	0.318	67	−7.357	100	−8.213
34	−1.804				

Table 10.B.3. Computer Generated Moving Average Data

Time	Observation	Time	Observation	Time	Observation
1	0.0824	35	−0.2092	68	−0.8659
2	−0.7421	36	1.1027	69	−0.3166
3	−2.5863	37	0.7200	70	1.2318
4	−1.0847	38	−0.1186	71	2.6148
5	0.0811	39	−0.7560	72	1.3507
6	2.1080	40	−1.1736	73	0.3786
7	1.2283	41	0.1670	74	0.0751
8	0.2222	42	0.4029	75	−0.1134
9	0.1139	43	1.4065	76	−0.9633
10	0.9652	44	−0.6458	77	−2.1476
11	1.7643	45	−1.8765	78	−1.6736
12	2.2358	46	−1.6670	79	0.2056
13	0.8803	47	−0.9132	80	2.3535
14	0.8975	48	0.9964	81	1.7980
15	0.8376	49	1.6263	82	−0.1434
16	−0.3025	50	0.5360	83	−1.4796
17	−2.5194	51	1.0236	84	−1.4369
18	−2.7883	52	0.7965	85	−0.7226
19	−3.5046	53	0.2285	86	0.4259
20	−4.1382	54	−0.2926	87	1.3133
21	−3.6002	55	−1.3110	88	1.8455
22	−1.5214	56	0.3846	89	0.5274
23	2.0135	57	1.9833	90	−1.8385
24	3.2380	58	−0.1662	91	−3.3115
25	1.5426	59	−0.0513	92	−1.7593
26	1.1661	60	3.0565	93	0.0670
27	0.6807	61	1.9968	94	0.4853
28	−0.8016	62	0.4235	95	3.0867
29	−1.5890	63	−1.3013	96	1.7763
30	−1.1519	64	−1.2351	97	1.5428
31	−1.3396	65	−1.7102	98	0.2979
32	0.0265	66	−1.2697	99	−1.4703
33	1.7851	67	−0.6817	100	−1.4294
34	0.1965				

Table 10.B.4. Computer Generated Data for Vector Process

Time	X_t	Y_t	Z_t
1	−5.59	−10.78	−6.64
2	−6.85	−12.11	−7.91
3	−7.25	−13.21	−8.12
4	−6.72	−12.48	−6.14
5	−9.56	−12.73	−9.75
6	−8.20	−13.50	−8.82
7	−7.93	−13.66	−8.06
8	−8.85	−13.72	−8.89
9	−9.55	−14.25	−9.98
10	−7.56	−12.61	−6.24
11	−6.47	−9.67	−4.13
12	−4.98	−7.05	−2.88
13	−3.14	−4.43	−1.05
14	−1.65	−2.79	−0.33
15	−0.78	−3.03	−0.98
16	−0.65	−3.47	−1.00
17	2.67	−0.30	5.21
18	5.28	5.08	9.59
19	5.96	8.73	8.88
20	8.97	12.29	11.83
21	13.18	15.86	16.03
22	14.49	19.61	17.49
23	17.45	21.46	18.93
24	19.15	22.98	20.37
25	20.99	23.74	21.60
26	20.56	23.45	20.33
27	19.87	22.11	18.80
28	19.88	21.98	19.77
29	18.84	21.47	18.44
30	18.25	20.14	17.19
31	17.46	18.61	16.24
32	16.03	17.23	14.93
33	14.89	16.32	14.17
34	14.69	14.99	13.62
35	15.46	15.41	15.80
36	14.56	15.60	14.71
37	14.45	14.62	13.66
38	14.27	14.44	14.13
39	14.63	14.98	15.06
40	14.12	16.56	15.38
41	14.89	16.39	14.75
42	15.09	16.30	15.02
43	14.16	15.56	13.57
44	13.46	13.34	11.68

Table 10.B.4.—*Continued*

Time	X_t	Y_t	Z_t
45	13.96	12.44	13.25
46	12.93	11.53	12.20
47	10.99	11.53	10.99
48	9.51	10.19	8.44
49	9.63	8.48	8.26
50	8.58	7.14	7.50
51	8.00	5.00	6.29
52	5.16	3.39	3.88
53	3.46	1.33	1.81
54	4.60	1.13	4.43
55	3.10	1.14	3.11
56	3.53	2.69	4.77
57	5.53	4.21	6.75
58	5.20	4.70	5.59
59	6.41	4.57	6.31
60	6.20	5.29	6.78
61	6.30	5.93	6.81
62	7.63	7.08	8.55
63	8.95	9.84	11.15
64	9.25	12.11	11.07
65	11.25	14.41	13.08
66	15.56	17.82	18.29
67	17.16	21.12	19.81
68	19.77	24.07	22.12
69	19.82	24.78	20.39
70	24.38	26.91	26.08
71	23.36	28.22	24.40
72	24.19	28.20	24.18
73	25.47	28.35	25.59
74	23.57	27.55	22.94
75	22.30	25.98	21.05
76	22.44	25.37	21.95
77	23.50	26.42	24.34
78	23.66	28.76	25.53
79	25.15	30.08	26.21
80	26.57	31.39	27.62
81	27.17	32.18	27.81
82	28.03	33.44	29.04
83	28.43	33.19	28.22
84	27.36	32.43	26.74
85	27.87	30.06	25.98
86	27.17	29.50	26.72
87	26.52	29.52	26.54
88	24.90	28.90	24.40

Table 10.B.4. —*Continued*

Time	X_t	Y_t	Z_t
89	24.56	27.58	23.50
90	24.04	26.97	23.56
91	21.84	26.34	21.33
92	23.68	26.11	23.50
93	22.50	24.44	21.16
94	20.36	24.37	20.30
95	20.53	23.21	19.60
96	20.59	22.00	19.63
97	19.75	20.66	18.67
98	17.93	18.60	16.28
99	15.65	16.22	13.75
100	15.12	14.62	13.83