Testing for a Unit Root in Time Series Regression

Peter C.B. Phillips, Pierre Perron

Institutions: Cowles Foundation

Published on: 01 Jun 1988 - Biometrika (Oxford University Press)

Topics: KPSS test, Unit root test, Unit root, Dickey–Fuller test and Trend stationary

Related papers:

- Distribution of the Estimators for Autoregressive Time Series with a Unit Root
- Co-integration and Error Correction: Representation, Estimation and Testing
- Likelihood ratio statistics for autoregressive time series with a unit root
- Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?
- Maximum likelihood estimation and inference on cointegration — with applications to the demand for money

Share this paper:  

View more about this paper here: https://typeset.io/papers/testing-for-a-unit-root-in-time-series-regression-u3iv5t9ka
Note: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than acknowledgment that a writer had access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

TESTING FOR A UNIT ROOT IN TIME SERIES REGRESSION

by

Peter C.B. Phillips & Pierre Perron

September 1987
TESTING FOR A UNIT ROOT IN
TIME SERIES REGRESSION

by

Peter C. B. Phillips
Cowles Foundation for Research in Economics
Yale University

and

Pierre Perron
Université de Montréal

Corresponding author's address:
Prof. P. C. B. Phillips
Cowles Foundation
Box 2125, Yale Station
New Haven, CT 06520

0. SUMMARY

This paper proposes some new tests for detecting the presence of a unit root in quite general time series models. Our approach is nonparametric with respect to nuisance parameters and thereby allows for a very wide class of weakly dependent and possibly heterogeneously distributed data. The tests accommodate models with a fitted drift and a time trend so that they may be used to discriminate between unit root nonstationarity and stationarity about a deterministic trend. The limiting distributions of the statistics are obtained under both the unit root null and a sequence of local alternatives. The latter noncentral distribution theory yields local asymptotic power functions for the tests and facilitates comparisons with alternative procedures due to Dickey and Fuller. Some simulations are reported which provide evidence on the performance of the new tests in finite samples.

Some key words: Brownian motion, Noncentral distributions, Time series, Unit root, Weak convergence.
1. **INTRODUCTION**

Methods for detecting the presence of a unit root in parametric time series models have lately attracted a good deal of interest in both statistical theory and application. Recent articles by Fuller (1984) and Dickey et al. (1986) review much of the literature in the field. The latter article provides a helpful practical guide to the use of some of the formal tests that have been developed.

One major field of application where the hypothesis of a unit root has important implications is economics. This is explained by the fact that the presence of a unit root is often a theoretical implication of models which postulate the rational use of information that is available to economic agents. Examples from economics include various financial market variables such as futures contracts (Samuelson (1965)), stock prices (Samuelson (1973)), dividends (Kleidon (1986)), spot and forward exchange rates (Meese and Singleton (1983)) and even aggregate variables like real consumption (Hall (1978)). Formal statistical tests of the unit root hypothesis are of additional interest to economists because they can help to evaluate the nature of the nonstationarity that most macroeconomic data exhibit. In particular, they help in determining whether the trend is stochastic, through the presence of a unit root, or deterministic, through the presence of a polynomial time trend. A recent examination of historical economic time series by Nelson and Plosser (1982), for example, found strong evidence in favor of unit root nonstationarity using the Dickey-Fuller (1979) testing procedure.

Recently, Said and Dickey (1984) have shown that the Dickey-Fuller
procedure, which was originally developed for autoregressive representations of known order, remains valid asymptotically for a general ARIMA \((p, l, q)\) process in which \(p\) and \(q\) are of unknown orders. More specifically, Said and Dickey (1984) show that the Dickey-Fuller regression t-test for a unit root may still be used in an ARIMA \((p, l, q)\) model provided the lag length in the autoregression increases with the sample size, \(T\), at a controlled rate less than \(T^{1/3}\).

An alternative procedure for testing the presence of a unit root in a general time series setting has recently been proposed by Phillips (1987a). This approach is nonparametric with respect to nuisance parameters and thereby allows for a very wide class of time series models in which there is a unit root. This includes ARIMA models with heterogeneously as well as identically distributed innovations. The method seems to have significant advantages when there are moving average components in the time series and, at least in this respect, offers a promising alternative to the Dickey-Fuller and Said-Dickey procedure.

The present paper extends the study of Phillips (1987a) to the cases where (a) a drift and (b) a drift and a linear trend are included in the specification. These extensions are important for practical applications, where the presence of a non zero drift is very common. Moreover, in many cases and, particularly, with economic time series the main competing alternative to the presence of a unit root is a deterministic linear time trend. It is therefore important that regression tests for unit roots allow for this possibility.

The methods of the paper are asymptotic and rely on the theory of functional weak convergence. The limit distributions of the new test statistics
developed here are expressed as functionals of standard Brownian motion and are the same as those tabulated in Fuller (1976). This means that our tests may be used with existing tabulations even though they allow for much more general time series specifications. The asymptotic local power properties of our tests are studied using the theory of near-integrated processes developed in Phillips (1987b). Some simulation evidence on the finite sample performance of the new tests is also provided. Proofs of our main results are outlined in the appendix.

2. PRELIMINARIES

The models we shall consider are driven by a sequence of innovations which we denote by \((u_t)\). Throughout the paper we assume that \((u_t)\) satisfies the following general conditions:

(A) \(E(u_t) = 0\) all \(t\);

(B) \(\sup_t E|u_t|^{\beta + \epsilon} < \infty\) for some \(\beta > 2\) and \(\epsilon > 0\);

(C) \(\sigma^2 = \lim_{T \to \infty} E(T^{-1} s_T^2)\) exists and \(\sigma^2 > 0\) where

\[
s_t = \sum_{j=1}^{t} u_j;
\]

(D) \((u_t)\) is strong mixing with mixing coefficients \(\alpha_m\) that satisfy:

\[
\sum_{m=1}^{\infty} \alpha_m^{1-2/\beta} < \infty.
\]

These conditions allow for many weakly dependent and heterogeneously distributed time series. They include a wide variety of possible data
generating mechanisms such as finite order ARMA models under very general conditions on the underlying errors; see Withers (1981). Condition (B) controls the allowable heterogeneity of the process, whereas (D) controls the extent of permissible temporal dependence in relation to the probability of outlier occurrences; see Phillips (1987a) for more discussion of these conditions and Hall and Heyde (1980) for the definition of strong mixing and the mixing coefficients. If \( \{u_t\} \) is weakly stationary with spectral density \( f_u(\lambda) \) then condition (C) is a consequence of (B) and (D). In this case we have:

\[
\sigma^2 = E(u_1^2) + 2 \sum_{k=2}^{\infty} E(u_1 u_k) = 2\pi f_u(0).
\]

From the sequence of partial sums \( \{S_n\} \) we construct the random element

\[
X_T(r) = T^{-1/2} \sigma^{-1} S_{[Tr]} = T^{-1/2} \sigma^{-1} S_{j-1} ; \quad (j-1)/T \leq r < j/T \quad (j=1, \ldots, T)
\]

where \([Tr]\) denotes the integral part of \( Tr \). \( X_T(r) \) lies in \( D = D[0,1] \), the space of real valued functions on the interval \([0,1]\) that are right continuous and have finite left limits. Under very general conditions the random element \( X_T(r) \) obeys a central limit theory on the function space \( D \). In particular, under (A)-(D) above we have the following result, which is due to Herrndorf (1984):

\[
X_T(r) \Rightarrow W(r), \text{ as } T \to \infty.
\]

The notation " \( \Rightarrow \) " that is used here and elsewhere in the paper signifies weak convergence of the associated probability measures. In this case, the
probability measure of \( X_t(r) \) converges weakly to the probability measure of the standard Brownian motion \( \dot{W}(r) \). The reader is referred to Billingsley (1968) and Pollard (1984) for further discussion.

Using (1) it is a simple matter to deduce the asymptotic behavior of the sample moments of the process \( \{S_t\} \) and the innovations \( \{u_t\} \). The approach is fully developed in Phillips (1987a) and we state below those results which are most useful in our subsequent theory. Unless otherwise indicated the summations are over \( t = 1, \ldots, T \). As \( T \to \infty \):

\[
T^{-3/2} \sum_t S_t = \sigma \int_0^1 \dot{W}(r) dr ;
\]

\[
T^{-2} \sum_t^2 = \sigma^2 \int_0^1 \dot{W}(r)^2 dr ;
\]

\[
T^{-3/2} \sum_t u_t = \sigma \int_0^1 r \dot{W}(r) - \sigma \left[ \dot{W}(1) - \int_0^1 \dot{W}(r) dr \right] ;
\]

\[
T^{-5/2} \sum_t S_t = \sigma \int_0^1 r^2 \dot{W}(r) dr ;
\]

\[
T^{-1} \sum_t S_t S_{t-1} = (1/2) \left( \sigma^2 \dot{W}(1)^2 - \sigma_u^2 \right) ;
\]

where

\[
\sigma_u^2 = \lim_{T \to \infty} \frac{1}{T} \sum_t u_t^2.
\]

Joint weak convergence for the sample moments given above to their respective limits is also easily established and will be utilized in what follows.
3. THE MODELS AND ESTIMATORS

Let \( y_t \) be a time series which is generated by:

\[
y_t = \alpha y_{t-1} + u_t ; \quad t = 1, 2, \ldots
\]

\[
\alpha = 1 ;
\]

(2) \hspace{1cm} (3)

Initial conditions for (2) are set at \( t = 0 \) and \( y_0 \) may be any random variable, including a constant, whose distribution is fixed and independent of the sample size \( T \). The innovation sequence \( \{u_t\} \) satisfies conditions (A)-(D).

We shall consider the two least squares regression equations:

\[
y_t = \hat{\mu} + \hat{\alpha} y_{t-1} + \hat{u}_t ;
\]

(4)

\[
y_t = \tilde{\mu} + \tilde{\alpha} (t - T/2) + \bar{\alpha} y_{t-1} + \bar{u}_t ;
\]

(5)

where \( (\hat{\mu}, \hat{\alpha}) \) and \( (\tilde{\mu}, \tilde{\alpha}, \bar{\alpha}) \) are the conventional least squares regression coefficients. We use \( X \) to denote the \( T \times 3 \) matrix of observations of the regressions in (5). We also define the following regression t-statistics:

\[
t_{\alpha} = (\hat{\alpha} - \alpha) \left( \frac{\sum (y_{t-1} - \bar{y}_{t-1})^2}{\hat{s}} \right)^{1/2}
\]

\[
t_{\mu} = (\hat{\mu} - \mu) \left( \frac{\sum (y_{t-1} - \bar{y}_{t-1})^2 / \sigma^2_{y_{t-1}}}{\hat{s}} \right)^{1/2}
\]

\[
t_{\mu} = (\tilde{\mu} - \mu) / (\sigma_{\mu}^2 c_1)^{1/2}
\]

\[
t_{\beta} = (\tilde{\beta} - \beta) / (\sigma_{\beta}^2 c_2)^{1/2}
\]
\[ t^* = \left( \bar{\alpha} - \alpha \right) / \left( \bar{s}^2 c_3 \right)^{1/2} \]

where \( \bar{s} \) and \( \bar{s} \) are the standard errors of regressions (4) and (5) and \( c_i \) is the \( i \)th diagonal element of the matrix \((X'X)^{-1}\).

As in Dickey and Fuller (1979), we shall be concerned with the limiting distributions of the regression coefficients in (4) and (5) and their t-statistics under the maintained hypothesis that the data are generated by (2) and (3). Thus, the null values of the coefficients in the above tests become \( \alpha = 1, \mu = \beta = 0 \). However, it is important to note that the coefficient \( \bar{\alpha} \) and its t-statistic in (5) are invariant with respect to the introduction of a non-zero drift \( \mu \neq 0 \) in the process generating the time series \( \{y_t\} \). Thus, we may replace (2) with the generating mechanism

\[ y_t = \mu + \alpha y_{t-1} + u_t ; \quad t = 1, 2, \ldots \]  

(2)'

and the distributions and asymptotic distributions of the above mentioned statistics are unchanged. This means that there is no loss in generality by assuming \( \mu = 0 \) and the limiting distributions we shall give for these statistics remain valid for \( \mu \neq 0 \).

Under (2) and (3) \( y_t = S_t + y_0 \) and the asymptotic behavior of sample moments of \( y_t \) and \( u_t \) follows directly from that of \( S_t \) and \( u_t \) given earlier. Thus, as \( T \rightarrow \infty \)

\[ T^{-3/2} \Sigma y_t = \sigma \int_0^1 W(r)dr ; \]

\[ T^{-2} \Sigma y^2_t = \sigma^2 \int_0^1 W(r)^2dr ; \]
\[ T^{-5/2} \sum_{t=1}^{T} y_{t-1} = \sigma_u^2 \int_0^1 \hat{w}(r) dr; \]
\[ T^{-1} \sum_{t=1}^{T} u_t = (1/2)(\sigma_u^2 \hat{w}(1)^2 - \sigma_u^2). \]

Again, joint weak convergence to the stated limits applies.

4. **LIMITING DISTRIBUTIONS OF THE STATISTICS**

In this section we characterize the limiting distributions of the standardized coefficient estimators \( T(\hat{\alpha} - 1) \), \( T(\tilde{\alpha} - 1) \) and the t-statistics \( t_{\hat{\alpha}} \) and \( t_{\tilde{\alpha}} \) (\( \alpha = 1 \)), \( t_{\hat{\mu}} \) and \( t_{\tilde{\mu}} \) (\( \mu = 0 \)) and \( t_{\tilde{\beta}} \) (\( \beta = 0 \)) under the maintained hypothesis that the time series \( (y_t) \) is generated by (2) and (3).

**THEOREM 1.** For the regression model (4), as \( T \to \infty \):

(a) \( T(\hat{\alpha} - 1) \to (\int_0^1 \hat{w}(r)^2 dr - (\int_0^1 \hat{w}(r) dr)^2)^{-1} \left\{ (1/2)(\hat{w}(1)^2 - \sigma_u^2/\sigma^2) \cdot \hat{w}(1) \int_0^1 \hat{w}(r) dr \right\}; \)

(b) \( t_{\hat{\alpha}} \to (\sigma/\sigma_u) \left\{ \int_0^1 \hat{w}(r)^2 dr - (\int_0^1 \hat{w}(r) dr)^2 \right\}^{-1/2} \)
\[ \cdot \left\{ (1/2)(\hat{w}(1)^2 - \sigma_u^2/\sigma^2) \cdot \hat{w}(1) \int_0^1 \hat{w}(r) dr \right\}; \]

(c) \( t_{\hat{\mu}} \to (\sigma/\sigma_u) \left\{ \int_0^1 \hat{w}(r)^2 dr - (\int_0^1 \hat{w}(r) dr)^2 \right\}^{-1/2} \)
\[ \left\{ \hat{w}(1) \int_0^1 \hat{w}(r)^2 dr - (1/2)(\hat{w}(1)^2 - \sigma_u^2/\sigma^2) \cdot \int_0^1 \hat{w}(r) dr \right\}. \]

For the regression model (5), as \( T \to \infty \):

(d) \( T(\tilde{\alpha} - 1) \to ((1/2)(\tilde{w}(1)^2 - \sigma_u^2/\sigma^2) + A_1)/D \)

(e) \( t_{\tilde{\alpha}} \to (\sigma/\sigma_u) ((1/2)(\tilde{w}(1)^2 - \sigma_u^2/\sigma^2) + A_1)/D^{1/2} \)

(f) \( t_{\tilde{\mu}} \to (\sigma/\sigma_u) (A_2 - (1/2)(\tilde{w}(1)^2 - \sigma_u^2/\sigma^2) \cdot \int_0^1 \tilde{w}(r) dr)/D^{1/2} \cdot A_3 \)

(g) \( t_{\tilde{\beta}} \to (\sigma/\sigma_u) [A_4 + \sqrt{3}(\tilde{w}(1)^2 - \sigma_u^2/\sigma^2)((1/2)\int_0^1 \tilde{w}(r) dr - \int_0^1 \tilde{w}(r) dr)]/D^{1/2} \cdot A_5 \)
where

\[ D = \int_0^1 W(r)^2 \, dr - 12 \left( \int_0^1 r W(r) \, dr \right)^2 + 12 \int_0^1 W(r) \, dr \int_0^1 r W(r) \, dr - \int_0^1 W(r) \, dr \left( \int_0^1 W(r) \, dr - \frac{1}{2} W(1) \right) - 4 \left( \int_0^1 W(r) \, dr \right)^2 \]

\[ A_1 = 12 \left( \int_0^1 r W(r) \, dr - \frac{1}{2} \int_0^1 W(r) \, dr \right) \left( \int_0^1 W(r) \, dr - \left( \frac{1}{2} W(1) \right) \right) - \int_0^1 W(r) \, dr \]

\[ A_2 = W(1) \left( \int_0^1 W(r)^2 \, dr - 12 \left( \int_0^1 r W(r) \, dr \right)^2 + 18 \int_0^1 W(r) \, dr \int_0^1 r W(r) \, dr - 6 \left( \int_0^1 W(r) \, dr \right)^2 \right) \]

\[ + 6 \int_0^1 W(r) \, dr \left( \left( \int_0^1 W(r) \, dr \right)^2 - 2 \int_0^1 r W(r) \, dr \int_0^1 W(r) \, dr \right) \]

\[ A_3 = \left[ D + \left( \int_0^1 W(r) \, dr \right)^2 \right]^{1/2} \]

\[ A_4 = 2 \sqrt{3} W(1) \left( \frac{1}{2} \int_0^1 W(r)^2 \, dr + \int_0^1 W(r) \, dr \int_0^1 r W(r) \, dr - \left( \int_0^1 W(r) \, dr \right)^2 \right) \]

\[ + 2 \sqrt{3} \int_0^1 W(r) \, dr \left( \left( \int_0^1 W(r) \, dr \right)^2 - \int_0^1 W(r)^2 \, dr \right) \]

\[ A_5 = \left( \int_0^1 W(r)^2 \, dr - \left( \int_0^1 W(r) \, dr \right)^2 \right)^{1/2} . \]

When the innovation sequence \((u_t)\) is independent and identically distributed, we have \(\sigma_u^2 = \sigma^2\). In this case, the limiting distributions of the statistics given in Theorem 1 are independent of nuisance parameters, as is readily seen by inspection; and percentage points of the asymptotic distributions have been calculated by Monte Carlo methods by Dickey and Fuller. Specifically, critical values of \(T(\hat{\alpha} - 1), T(\tilde{\alpha} - 1), t_\hat{\alpha}\) and \(t_{\tilde{\alpha}}\) are tabulated in Fuller (1976, tables 8.5.1 and 8.5.2); and tabulated critical
values of $t_{\hat{\mu}}$, $t_{-\mu}$, and $t_{\beta}$ can also be found in Dickey and Fuller (1981, tables I-III).

Theorem 1 extends the results of Dickey and Fuller to the general case of weakly dependent and heterogeneously distributed data. Interestingly, our result shows that the limiting distributions of these statistics have the same general form for a very wide class of innovation processes $(u_t)$. This feature enables us to derive transformations of the statistics that opens the way to hypothesis testing in the general case. This is the approach that we pursue in the next section.

We remark that independent and identically distributed innovations $(u_t)$ are not necessary for the equivalence $\sigma^2 = \sigma_u^2$. The equivalence also holds for innovations that are martingale differences under mild additional moment conditions. Thus, unmodified versions of the Dickey-Fuller tests are valid asymptotically in the presence of some heterogeneity in the innovation sequence provided the innovations are martingale differences and (1) holds. However, when the innovations are non orthogonal and $\sigma^2 \neq \sigma_u^2$, the Dickey-Fuller tests do not have the correct asymptotic size.

5. **STATISTICAL INFERENCE IN THE PRESENCE OF A UNIT ROOT**

The limiting distributions of the regression coefficients and associated $t$-statistics given in Section 4 all depend upon the nuisance parameters $\sigma^2$ and $\sigma_u^2$. This presents an obstacle to conventional procedures of inference in the general case where $\sigma^2 \neq \sigma_u^2$. However, since $\sigma^2$ and $\sigma_u^2$ may be consistently estimated there exist simple transformations of the test statistics which eliminate the nuisance parameters asymptotically. This
idea was first developed by Phillips (1987a) in the context of tests for a
unit root. Here we show how the procedure may be extended to apply quite
generally to statistical tests in regressions with a fitted drift and time
trend.

Consistent estimates of \( \sigma^2 \) are provided by \( \bar{s}^2 \), \( \tilde{s}^2 \) and
\[ s^2 = T^{-1} \sum (y_t - y_{t-1})^2 \]
for data generated by (2) and (3). When there is a
non zero drift in the model, as in (2)'s, both \( \bar{s}^2 \) and \( \tilde{s}^2 \) are consistent.
Since we often wish to allow for a non zero drift in regressions such as (5)
we shall use \( \tilde{s}^2 \) as our preferred estimator of \( \sigma^2 \) in this regression.

Consistent estimation of \( \sigma^2 \) is discussed in Phillips (1987a). When
\( (u_t) \) is weakly stationary with spectral density \( f_u(\lambda) \) we have
\[ \sigma^2 = 2\pi f_u(0) \]. In this case, estimation of \( \sigma^2 \) is equivalent to estimating
the spectral density of \( (u_t) \) at the origin. Many consistent estimates are
available. Consider, for example, the simple estimate based on truncated
sample autocovariances, viz

\[ s^2_{T\ell} = T^{-1} \sum^T_{t=1} u_t^2 + 2T^{-1} \sum^T_{s=1} \sum^T_{t=s+1} u_t u_{t-s} \quad (6) \]

where \( u_t = y_t - y_{t-1} \). Conditions for the consistency of \( s^2_{T\ell} \) are explori-
ed in Phillips (1987a). It is shown there that \( s^2_{T\ell} \to \sigma^2 \) in probability as
\( T \to \infty \) provided the moment condition

\[ (B') \quad \sup_t E|u_t|^{2\beta} < \infty, \text{ for some } \beta > 2 \]

holds in place of (B) of Section 2 and provided

\[ (E) \quad \ell \to \infty \text{ as } T \to \infty \text{ and } \ell^2/T \to 0 \].
According to this result, if we allow the number of estimated autocovariances in (6) to increase as $T \to \infty$ but control the rate of increase so that $\ell^4/T \to 0$ then $s_{T \ell}^2$ yields a consistent estimator of $\sigma^2$. Of course, faster rates of increase in $\ell$ are allowable if we make stronger assumptions on $(u_t)$ as, for example, in the case of spectral estimation for weakly stationary processes. In what follows we shall assume that conditions (B') and (E) hold, in addition to (A), (0) and (D) given earlier.

Rather than using first differences $u_t = y_t - y_{t-1}$ in the construction of $s_{T \ell}^2$ we could use the residuals from the regression equations (4) and (5). Since the coefficients in these regressions are consistent it is easy to show that these modifications to $s_{T \ell}^2$ which we denote by $\tilde{s}_{T \ell}^2$ and $\frac{\Sigma}{\sigma_{T \ell}^2}$ respectively, are also consistent estimates of $\sigma^2$ under the same conditions. Once again $\tilde{s}_{T \ell}^2$ will be the preferred estimator when we wish to allow for a non zero drift as in (2)'.

Note that (6) is not constrained to be non negative as it is presently defined and it can take on negative values when there are large negative sample serial covariances. Simple modifications to (6) overcome this difficulty. For example, the weighted variance estimators

$$\sigma_{T \ell}^2 = T^{-1} \sum_{t=1}^{T} \frac{\Sigma}{\lambda^2} \sum_{s=1}^{\ell} w_{s \ell} \frac{\Sigma}{s} \sum_{t=1}^{T} \hat{u}_t \hat{u}_{t-s}$$

and

$$\tilde{s}_{T \ell}^2 = T^{-1} \sum_{t=1}^{T} \frac{\Sigma}{\lambda^2} \sum_{s=1}^{\ell} w_{s \ell} \frac{\Sigma}{s} \sum_{t=1}^{T} \hat{u}_t \hat{u}_{t-s}$$

where

$$w_{s \ell} = 1 - s/(\ell+1)$$

(7)
are non negative and, for stationary \((u_t)\), are simply \(2\pi\) times the corresponding Bartlett estimates of the spectrum \(f_u(\lambda)\) at the origin \(\lambda = 0\).

Other choices of lag window besides the triangular window (9) are possible.

We use the Parzen window in the simulations reported in Section 7 below.

The estimator (7) was recently suggested in the context of variance estimates by Newey and West (1987).

We now define some simple transformations of conventional test statistics from the regressions (4) and (5) which eliminate the nuisance parameter dependencies asymptotically. Specifically, we define:

\[
Z(\hat{\alpha}) = T(\hat{\alpha} - 1) - (1/2) \left( T^{-2} \Sigma (y_{t-1} - \bar{y}_{t-1})^2 \right)^{-1} (\hat{\sigma}_{Tl}^2 - \hat{s}^2);
\]

\[
Z(t_{\alpha}^2) = (\hat{s}/\hat{\sigma}_{Tl}) t_{\alpha} - (1/2) \hat{\sigma}_{Tl} \left( T^{-2} \Sigma (y_{t-1} - \bar{y}_{t-1})^2 \right)^{-1/2} (\hat{\sigma}_{Tl}^2 - \hat{s}^2);
\]

\[
Z(t_{\mu}^2) = (\hat{s}/\hat{\sigma}_{Tl}) t_{\mu} + (1/2\hat{\sigma}_{Tl}^2) (\hat{\sigma}_{Tl}^2 - \hat{s}^2) \left( T^{-2} \Sigma (y_{t-1} - \bar{y}_{t-1})^2 \right)^{-1/2}
\]

\[
\cdot \left( T^{-2} \Sigma y_{t-1}^2 \right)^{-1/2} \left( T^{-3/2} \Sigma y_{t-1} \right);
\]

\[
Z(\tilde{\alpha}) = T(\tilde{\alpha} - 1) - (T^6/24D_X)(\hat{\sigma}_{Tl}^2 - \hat{s}^2);
\]

\[
Z(t_{\alpha}^\prime) = (\hat{s}/\hat{\sigma}_{Tl}) t_{\alpha}^\prime - (T^3/4\sqrt{3} D_X^{1/2}) (\hat{\sigma}_{Tl} - \hat{s}^2);
\]

\[
Z(t_{\mu}^\prime) = (\hat{s}/\hat{\sigma}_{Tl}) t_{\mu}^\prime + (T^3/24D_X^{1/2} \Sigma X \hat{\sigma}_{Tl}) (\hat{\sigma}_{Tl}^2 - \hat{s}^2) (T^{-3/2} \Sigma y_{t-1});
\]
$$Z(t_β) = \left( \frac{\bar{s}}{\bar{σ}^2_{Tt}} \right)^{t_β} - \left( \frac{T^3}{2s_X} \bar{σ}^2_{Tt} \right) \left\{ T^{-2} \Sigma(y_{t-1} - \bar{y}_{t-1})^2 \right\}^{-1/2}$$

$$\cdot \left( \bar{σ}^2_{Tt} - \bar{s}^2 \right) \left( (1/2) T^{-3/2} \Sigma y_{t-1} - T^{-5/2} \Sigma(t-1)y_{t-1} \right)$$

where

$$D_X = \text{det}(X'X) = (T^2(T^2 - 1)/12) \Sigma y_{t-1}^2 - T(\Sigma y_{t-1})^2$$

$$+ T(T+1)(\Sigma y_{t-1}) (\Sigma y_{t-1}) - (T(T+1)(2T+1)/6)(\Sigma y_{t-1})^2$$

and

$$E_X = \left\{ T^{-6} D_X + (1/12) \left\{ T^{-3/2} \Sigma y_{t-1} \right\}^2 \right\}^{1/2}.$$

These $Z$ statistics extend those developed in Phillips (1987a) for the case of an autoregression with no fitted constant or time trend. The idea behind their construction is to correct the conventional regression statistics so that they allow for the effects of serially correlated and heterogeneously distributed innovations. Thus, the standard errors of regression $\hat{s}$ and $\bar{s}$ which measure scale effects in the conventional t-ratios are now replaced by the general standard error estimates $\hat{σ}^2_{Tt}$ and $\bar{σ}^2_{Tt}$ which allow for serial covariation as well as variance. Each $Z$ statistic also involves an additive correction term whose magnitude depends on the difference between the corresponding variance estimates $\hat{σ}^2_{Tt} - \bar{s}^2$ or $\bar{σ}^2_{Tt} - \bar{s}^2$. Once again these differences capture the effects of serial correlation and the transformations are designed to remove these effects asymptotically. The limiting distributions of the $Z$ statistics are given in our next main result.
THEOREM 2

(a) For the regression model (4) the statistics \( Z(\hat{\alpha}) \), \( Z(t_{\alpha}^\cdot) \) and \( Z(t_{\cdot \cdot}^\cdot) \) have limit distributions given by those of \( T(\hat{\alpha} - 1) \), \( t_{\alpha}^\cdot \) and \( t_{\cdot \cdot}^\cdot \), respectively in Theorem 1 with \( \sigma^2 = \sigma_u^2 \).

(b) For the regression model (5) the statistics \( Z(\bar{\alpha}) \), \( Z(t_{\alpha}^-) \), \( Z(t_{\mu}^-) \) and \( Z(t_{\bar{\beta}}^-) \) have limit distributions given by those of \( T(\bar{\alpha} - 1) \), \( t_{\alpha}^- \), \( t_{\mu}^- \) and \( t_{\bar{\beta}}^- \), respectively, in Theorem 1 with \( \sigma^2 = \sigma_u^2 \). The stated results for \( Z(\bar{\alpha}) \) and \( Z(t_{\alpha}^-) \) remain valid if the generating mechanism of \( (y_t) \) is (2)', rather than (2).

Theorem 2 shows that the limiting distributions of the \( Z \) statistics are invariant within a wide class of weakly dependent and possibly heterogeneously distributed innovations \( (u_t) \). Furthermore, the limiting distributions of the \( Z \) statistics are identical to the limiting distributions of the original untransformed statistics considered in Section 4, when \( \sigma_u^2 = \sigma^2 \). Thus, the critical values derived in the studies of Dickey and Fuller under the assumption of independent and identically distributed errors \( (u_t) \) may be used with the new tests proposed here, which are valid under much more general conditions.

6. POWER FUNCTIONS FOR UNIT ROOT TESTS

We may develop asymptotic power functions for unit root tests by considering the sequence of local alternatives to (3) given by:

\[
\alpha = e^{c/T} - 1 + c/T.
\]

(10)

When \( c = 0 \) (10) reduces to the null hypothesis (3); \( c > 0 \) gives local
explosive alternatives to (3); and \( c < 0 \) corresponds to local stationary alternatives. The idea of developing a noncentral asymptotic distribution theory using the specification (10) was explored in Phillips (1987b). Time series generated by models such as (2) or (2)', with a coefficient \( \alpha \) of the form (10) were called near-integrated in that paper. The asymptotic theory developed there showed that the sample moments of a near integrated time series converge weakly to corresponding functionals of a diffusion process rather than standard Brownian motion. Specifically, we have, as \( T \to \infty \):

\[ T^{-3/2} \Sigma \gamma T = \sigma \int_0^1 J_c(r) dr ; \]

\[ T^{-1/2} \Sigma \gamma^2 T = \sigma^2 \int_0^1 J_c(r)^2 dr ; \]

\[ T^{-5/2} \Sigma \gamma T = \sigma \int_0^1 r J_c(r) dr ; \]

\[ T^{-1} \Sigma \gamma T^{-1} u_T = \sigma^2 \int_0^1 J_c(r) dW(r) + (1/2)(\sigma^2 - \sigma^2_u) ; \]

where

\[ J_c(r) = \int_0^r e^{(r-s)c} dW(s) \]

is the Ornstein-Uhlenback process. It is generated in continuous time by the stochastic differential equation

\[ dJ_c(r) = cJ_c(r) dr + dW(r) \]

with initial condition \( J_c(0) = 0 \). As before, joint weak convergence of these sample moments to their respective limits also applies. Note that \( \int_0^1 J_c dW \) is interpreted as a stochastic integral in the above formulae.
Using these results for sample moments we may now develop an asymptotic theory for the regression coefficients and t-statistics in (4) and (5).

Moreover, it is a simple matter to find the noncentral asymptotic distributions of the new unit root test statistics developed in Section 6. The main results of interest are contained in the following theorem which concentrates on estimates of the autoregressive coefficient $\alpha$ and its associated t-ratio.

**THEOREM 3.** If \( \{y_t\} \) is a near-integrated time series generated by (2) and (10), then as \( T \to \infty \):

(a) \( Z(\hat{\alpha}) \to c + (\int_0^1 \beta_c(r)^2 \, dr - (\int_0^1 \beta_c(r) \, dr)^2)^{-1} \left( \int_0^1 \beta_c(r) \, dW(r) - \bar{W}(1) \int_0^1 \beta_c(r) \, dr \right) \)

(b) \( Z(t_{\alpha}) \to c \left( \int_0^1 \beta_c(r)^2 \, dr - (\int_0^1 \beta_c(r) \, dr)^2 \right)^{1/2} + \left( \int_0^1 \beta_c(r)^2 \, dr - (\int_0^1 \beta_c(r) \, dr)^2 \right)^{-1/2} \left( \int_0^1 \beta_c(r) \, dW(r) - \bar{W}(1) \int_0^1 \beta_c(r) \, dr \right) \).

(c) \( Z(\tilde{\alpha}) \to c + (\int_0^1 \beta_c(r) \, dW(r) + A_{1c})/D_c \)

(d) \( Z(t_{\tilde{\alpha}}) \to cD_c^{1/2} + (\int_0^1 \beta_c(r) \, dW(r) + A_{1c})/D_c^{1/2} \).

where

\[
D_c = \int_0^1 \beta_c(r)^2 \, dr - 12 \left( \int_0^1 \beta_c(r) \, dr \right)^2 + 12 \int_0^1 \beta_c(r) \, dr \int_0^1 \beta_c(r) \, dr - 4 \left( \int_0^1 \beta_c(r) \, dr \right)^2,
\]

and

\[
A_{1c} = 12 \left[ (1/2) \int_0^1 \beta_c(r) \, dW(1) - (1/3) \int_0^1 \beta_c(r) \, dr \right] - \int_0^1 \beta_c(r) \int_0^1 \beta_c(r) \, dr - (1/2) \int_0^1 \beta_c(r) \, dr \]

Results (c) and (d) remain valid if the generating mechanism of \( \{y_t\} \) is (2)' rather than (2).
Theorem 3 gives the noncentral limiting distributions of the $Z$ statistics for testing $\alpha = 1$ under the sequence of local alternatives (10) to the unit root hypothesis (3). It therefore delivers asymptotic local power functions for the new unit root tests.

It is interesting to compare these asymptotic local power functions with those of the conventional Dickey-Fuller tests. The latter are based on the statistics $T(\hat{\alpha}-1)$ and $t_{\hat{\alpha}}$ for regression (4) and $T(\tilde{\alpha}-1)$ and $t_{\tilde{\alpha}}$ for regression (5). When the innovation sequence $\{u_t\}$ is independent and identically distributed these statistics have identical limiting distributions under the local alternative hypothesis (10) as the $Z$ statistics given above. We deduce that the new tests based on $Z(\hat{\alpha})$, $Z(t_{\hat{\alpha}})$, $Z(\tilde{\alpha})$ and $Z(t_{\tilde{\alpha}})$ have the same asymptotic local power properties for a wide class of possible time series innovations $\{u_t\}$ as the regression based tests $(T(\hat{\alpha}-1), t_{\hat{\alpha}})$ and $(T(\tilde{\alpha}-1), t_{\tilde{\alpha}})$ do in the case of independent and identically distributed errors. Thus, there is no loss in asymptotic power in the use of the new tests over the Dickey-Fuller procedure in spite of the fact that they allow for a more general class of error processes.

7. EXPERIMENTAL EVIDENCE

Simulations were run to assess the adequacy of the new tests and to evaluate their performance in comparison with the procedure suggested by Said and Dickey (1984). As explained earlier, Said and Dickey recommend the use of the Dickey-Fuller regression $t$-test for a unit root in the autoregression

$$\Delta y_t = \hat{\mu} + \hat{\alpha} y_{t-1} + \sum_{i=1}^{l} \phi_i \Delta y_{t-i} + \hat{\nu}_t$$

(11)
We denote this test statistic \( t(\hat{\alpha}_x) \). Said and Dickey show that when the lag length \( l \to \infty \) in (11) as \( T \to \infty \) then \( t(\hat{\alpha}_x) \) has the same limit distribution as the conventional Dickey-Fuller \( t \)-test. This corresponds with our statistic \( Z(t_\alpha) \) given above. Note that Said and Dickey do not suggest a statistic based on the coefficient \( \hat{\alpha}_x \) in (11), since the limit distribution of \( T\hat{\alpha}_x \) depends on nuisance parameters. Thus, there is no analogue of our \( Z(\hat{\alpha}) \) test in Said and Dickey (1984).

Data were generated by the model (2) with moving average errors

\[
u_t = e_t + \theta e_{t-1}
\]  

(12)

and the \( e_t \) independent and identically distributed \( N(0,1) \). We set \( \gamma_0 = 0 \) and used various lag lengths \( l \) in (11) and lag truncations \( \ell \) in (7) to evaluate the effects of these choices on test performance. A Parzen window and fitted residuals \( \hat{u}_t \) from (4) were used in the construction of the variance estimate \( \hat{\sigma}^2_T \). The simulations reported in Table 1 are based on 2,000 replications and give results for one-sided tests under the null hypothesis \( \alpha = 1 \) and for the alternative \( \alpha = 0.85 \).

The results show size and power computations for six different values of \( \theta \) in (12). When \( \theta = 0 \) there is no need to employ the transformations leading to \( Z(\hat{\alpha}) \) and \( Z(t_\alpha) \) or the long autoregression (11). However, we gather from Table 1 that there is little loss in accuracy with respect to the size of the \( Z(\hat{\alpha}) \) and \( Z(t_\alpha) \) tests. In fact, the \( Z(\hat{\alpha}) \) test is conservative and at the same time has greater power than either \( Z(t_\alpha) \) or \( t_{\alpha_x} \) for all choices of \( \ell \). The \( Z(t_\alpha) \) and \( t_{\alpha_x} \) tests are both liberal in terms of size at \( T = 100 \) and the size distortions of \( t_{\alpha_x} \) increase appreciably with the length of the autoregression. At the same time, the
### TABLE 1

(T = 100, nominal size 5%)

(a) **Said-Dickey $t(\hat{\gamma}_*)$ test**

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>12</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta = 0.0$</td>
<td>0.068</td>
<td>0.064</td>
<td>0.078</td>
<td>0.086</td>
<td>0.106</td>
<td>0.557</td>
<td>0.472</td>
<td>0.406</td>
<td>0.354</td>
<td>0.303</td>
</tr>
<tr>
<td>$\beta = 0.5$</td>
<td>0.052</td>
<td>0.064</td>
<td>0.071</td>
<td>0.085</td>
<td>0.102</td>
<td>0.378</td>
<td>0.411</td>
<td>0.377</td>
<td>0.348</td>
<td>0.308</td>
</tr>
<tr>
<td>$\beta = 0.8$</td>
<td>0.037</td>
<td>0.051</td>
<td>0.072</td>
<td>0.082</td>
<td>0.109</td>
<td>0.267</td>
<td>0.302</td>
<td>0.313</td>
<td>0.308</td>
<td>0.273</td>
</tr>
<tr>
<td>$\beta = -0.2$</td>
<td>0.063</td>
<td>0.065</td>
<td>0.084</td>
<td>0.092</td>
<td>0.111</td>
<td>0.591</td>
<td>0.490</td>
<td>0.421</td>
<td>0.358</td>
<td>0.304</td>
</tr>
<tr>
<td>$\beta = -0.5$</td>
<td>0.116</td>
<td>0.076</td>
<td>0.073</td>
<td>0.086</td>
<td>0.105</td>
<td>0.868</td>
<td>0.626</td>
<td>0.512</td>
<td>0.434</td>
<td>0.350</td>
</tr>
<tr>
<td>$\beta = -0.8$</td>
<td>0.677</td>
<td>0.343</td>
<td>0.201</td>
<td>0.142</td>
<td>0.120</td>
<td>1.00</td>
<td>0.988</td>
<td>0.900</td>
<td>0.772</td>
<td>0.523</td>
</tr>
</tbody>
</table>

(b) **$Z(\hat{\alpha})$ test**

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>12</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta = 0.0$</td>
<td>0.044</td>
<td>0.049</td>
<td>0.049</td>
<td>0.049</td>
<td>0.055</td>
<td>0.772</td>
<td>0.776</td>
<td>0.784</td>
<td>0.794</td>
<td>0.805</td>
</tr>
<tr>
<td>$\beta = 0.5$</td>
<td>0.010</td>
<td>0.021</td>
<td>0.022</td>
<td>0.021</td>
<td>0.014</td>
<td>0.301</td>
<td>0.524</td>
<td>0.534</td>
<td>0.500</td>
<td>0.378</td>
</tr>
<tr>
<td>$\beta = 0.8$</td>
<td>0.003</td>
<td>0.020</td>
<td>0.024</td>
<td>0.022</td>
<td>0.015</td>
<td>0.247</td>
<td>0.513</td>
<td>0.526</td>
<td>0.474</td>
<td>0.330</td>
</tr>
<tr>
<td>$\beta = -0.2$</td>
<td>0.134</td>
<td>0.105</td>
<td>0.110</td>
<td>0.115</td>
<td>0.133</td>
<td>0.960</td>
<td>0.940</td>
<td>0.946</td>
<td>0.952</td>
<td>0.967</td>
</tr>
<tr>
<td>$\beta = -0.5$</td>
<td>0.537</td>
<td>0.417</td>
<td>0.428</td>
<td>0.451</td>
<td>0.516</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>$\beta = -0.8$</td>
<td>0.997</td>
<td>0.988</td>
<td>0.988</td>
<td>0.991</td>
<td>0.995</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

(c) **$Z(t^\alpha)$ test**

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>12</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta = 0.0$</td>
<td>0.063</td>
<td>0.062</td>
<td>0.066</td>
<td>0.069</td>
<td>0.069</td>
<td>0.669</td>
<td>0.688</td>
<td>0.696</td>
<td>0.705</td>
<td>0.705</td>
</tr>
<tr>
<td>$\beta = 0.5$</td>
<td>0.028</td>
<td>0.036</td>
<td>0.038</td>
<td>0.035</td>
<td>0.030</td>
<td>0.191</td>
<td>0.367</td>
<td>0.371</td>
<td>0.333</td>
<td>0.214</td>
</tr>
<tr>
<td>$\beta = 0.8$</td>
<td>0.026</td>
<td>0.034</td>
<td>0.035</td>
<td>0.031</td>
<td>0.026</td>
<td>0.144</td>
<td>0.353</td>
<td>0.357</td>
<td>0.305</td>
<td>0.167</td>
</tr>
<tr>
<td>$\beta = -0.2$</td>
<td>0.135</td>
<td>0.115</td>
<td>0.119</td>
<td>0.127</td>
<td>0.140</td>
<td>0.933</td>
<td>0.910</td>
<td>0.918</td>
<td>0.930</td>
<td>0.945</td>
</tr>
<tr>
<td>$\beta = -0.5$</td>
<td>0.534</td>
<td>0.438</td>
<td>0.445</td>
<td>0.460</td>
<td>0.518</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>$\beta = -0.8$</td>
<td>0.995</td>
<td>0.988</td>
<td>0.990</td>
<td>0.993</td>
<td>0.995</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
power of the $t_{\hat{\alpha}}$ tests decreases as $l$ increases. Thus, the cost of using $t_{\hat{\alpha}}$ when $\theta = 0$ is appreciably greater than that of using $Z(t_{\hat{\alpha}})$ and $Z(\hat{\alpha})$ is the preferred test.

When $\theta > 0$ similar results apply. Since $T = 100$ moderate choices of $l$ around $l = 8$ seem appropriate. We observe that the $Z(\hat{\alpha})$ test is again very conservative and has higher power than the $Z(t_{\hat{\alpha}})$ and $t_{\hat{\alpha}}$ tests for all choices of $l \geq 4$. The $Z(t_{\hat{\alpha}})$ test is also conservative and has similar power to $t_{\hat{\alpha}}$ for $l = 4, 6, 8$. The $t_{\hat{\alpha}}$ test is liberal for $l \geq 4$ and the size distortions increase with the length of the autoregression. These results again suggest that $Z(\hat{\alpha})$ is the preferred test.

When $\theta < 0$ the results are very different. Both $Z(\hat{\alpha})$ and $Z(t_{\hat{\alpha}})$ have significant size distortions and are too liberal to be useful for $\theta = -0.5, -0.8$. The $t_{\hat{\alpha}}$ test also suffers size distortions but these are attenuated as the lag length in the autoregression (11) increases. However, as $l$ increases the $t_{\hat{\alpha}}$ test also suffers appreciable loss in power. None of the tests therefore has accurate size and good power properties when $\theta < 0$. But, of the three tests, $t_{\hat{\alpha}}$ seems preferable in this case.

Asymptotic expansions recently obtained in Phillips (1987c) may be used to shed light on this simulation evidence. In particular, formula (34) of that paper may be extended to apply to the regression (4) with fitted intercept giving:

$$T(\hat{\alpha}-1) = (\int_0^1 \overline{w}(r)^2 dr)^{-1} \left( \int_0^1 \overline{w}(r) d\overline{w}(r) - k_0 - (2T)^{-1/2} \overline{k_1} \eta \right) + o_p(T^{-1})$$

where the symbol $" = "$ signifies equivalence in distribution,

$$\overline{w}(r) = \overline{w}(r) - \int_0^1 \overline{w}(r) dr ,$$
\[ k_0 = \frac{1}{2} \frac{\sigma_u^2}{\sigma^2} = \frac{1}{2} \frac{\frac{1+\theta^2}{(1+\theta)^2}} \]

\[ k_1 = \frac{(1 + 4\theta^2 + \theta^4)^{1/2}}{(1+\theta)^2} \]

and \( \eta \) is standard \( N(0,1) \) and independent of \( W(r) \). The correction term on the asymptotic depends on \( k_1 \). For \( \theta \geq 0 \), we have \( 0 < k_1 \leq 1 \).

When \( \theta < 0 \), \( k_1 \) is unbounded and rapidly becomes large as \( \theta \to -1 \).

These results show clearly how the quality of the asymptotic approximation given in Theorem 1(a) depends on the size and the magnitude of \( \theta \). When \( \theta > 0 \) we have \( k_1 < 1 \) and the asymptotic distribution may be expected to deliver at least as good an approximation to the finite sample distribution of \( T(\hat{\theta} - 1) \) as it does for the case where \( \theta = 0 \) and \( k_1 = 1 \). When \( \theta < 0 \) this is not the case and the asymptotic distribution may be expected to be poor, particularly as \( \theta \to -1 \). Similar behavior may be inferred for the test statistics \( Z(\hat{\theta}) \) and \( Z(t_{\hat{\theta}}) \), which are both based on \( T(\hat{\theta} - 1) \). This helps to explain the size distortions in these tests that were evident in the simulations when \( \theta < 0 \).

8. CONCLUDING COMMENTS

The approach presented here gives a simple way to test for a unit root in univariate time series against stationary and trend alternatives. One needs only to estimate a first order autoregression with a constant and possibly a time trend and to calculate the appropriate transformed \( Z \) statistic. The distribution theory underlying this procedure is asymptotic and
critical values already provided by Fuller (1976) may be used in mounting the tests.

As we have seen in Section 6, there is no loss in asymptotic local power in the use of the Z tests for a unit root. But the simulations reported in Section 7 indicate that test performance can differ substantially in finite samples among asymptotically equivalent tests. For models with positive moving average errors the Z(â) test is conservative and has better power properties than the other tests. For models with independent and identically distributed errors where the transformations that lead to the Z tests are not strictly needed, Z(â) again seems to be the preferred test. For models with moving average errors and negative serial correlation the Z tests suffer appreciably size distortions and are not recommended. In such cases the Said-Dickey procedure of using a long auto-regression seems preferable.

ACKNOWLEDGMENTS

The authors thank the referees, Joon Park and Gene Savin for helpful comments on an earlier draft. Our thanks also go to Glena Ames for her skill and effort in typing the manuscript of this paper, to the NSF for research support at Yale University and to the Université de Montréal for computing facilities.
APPENDIX: PROOFS OF THEOREMS 1-3

Proof of Theorem 1. To prove part (a) we simply note that

\[ T(\hat{\alpha} - 1) = \left( T^{-2} \Sigma(y_{t-1} - \bar{y}_{t-1}) \right)^{-1} \left( T^{-1} \Sigma u_t (y_{t-1} - \bar{y}_{t-1}) \right) \]

and the stated result follows by direct application of the continuous mapping theorem and the joint convergence of the requisite sample moments. To prove (b) we note from (a) that \( \hat{\alpha} \to 1 \) in probability and, by the strong law for weakly dependent sequences such as \( \{u_t\} \), \( \hat{\sigma}^2 \to \sigma_u^2 \) almost surely as \( T \to \infty \). Once again the stated result follows from the continuous mapping theorem and the earlier results on sample moments. To prove (c) we observe that

\[ T^{1/2} \hat{\mu} = T^{-1/2} s_T - (T^{-3/2} \Sigma y_{t-1}) \hat{\alpha} \]

\[ \Rightarrow \sigma \hat{W}(1) - \sigma \int_0^1 u(r) dr \left[ \int_0^1 u(r)^2 dr - \left( \int_0^1 u(r) dr \right)^2 \right]^{-1} \]

\[ \cdot (1/2) (W(1)^2 - \sigma_u^2) - \hat{W}(1) \int_0^1 u(r) dr \]

and the stated result follows directly from the definition of \( t_\mu \), with \( \mu = 0 \), and the fact that \( \hat{\sigma} \to \sigma_u \) in probability.

To prove (d)-(g) we first note that elementary but tedious calculations yield the formulae:
\[ \tilde{\alpha} - 1 = D_X^{-1} \left[ \frac{(T(T+1)/2)\Sigma y_{t-1}\Sigma u_t}{t-1} - \frac{(T(T+1)(2T+1)/6)\Sigma y_{t-1}^2}{t-1} \right] \\
+ T\Sigma y_{t-1}^2 \Sigma u_t + (T(T+1)/2)\Sigma y_{t-1} \Sigma u_t + (T^2(T^2-1)/12)\Sigma y_{t-1} y_{t-1} \right] \]

\[ \tilde{\mu} = D_X^{-1} \left[ \frac{(T(T+1)(T+2)/12)\Sigma y_{t-1}^2}{t-1} \Sigma u_t - (\Sigma y_{t-1}^2)^2 \Sigma u_t \right] \\
+ (T/2)\Sigma y_{t-1} \Sigma y_{t-1} \Sigma u_t - (T/2)\Sigma y_{t-1}^2 \Sigma u_t \]

\[ + \Sigma y_{t-1} \Sigma y_{t-1} \Sigma u_t - (T/2)(\Sigma y_{t-1}^2)^2 \Sigma u_t \]

\[ + (T/2)\Sigma y_{t-1} \Sigma u_t y_{t-1} - (T(T+1)(T+2)/12)\Sigma y_{t-1} \Sigma u_t y_{t-1} \right] \]

\[ \tilde{\beta} = D_X^{-1} \left[ -\frac{(T(T+1)/2)\Sigma y_{t-1}^2}{t-1} \Sigma u_t + \Sigma y_{t-1} \Sigma y_{t-1} \Sigma u_t + T\Sigma y_{t-1}^2 \Sigma u_t - (\Sigma y_{t-1}^2)^2 \Sigma u_t \right] \\
- T\Sigma y_{t-1} \Sigma u_t y_{t-1} + (T(T+1)/2)\Sigma y_{t-1} \Sigma u_t y_{t-1} \right] \]

where

\[ D_X = \text{det}(X'X) = \frac{T^2(T^2-1)/12}{2} \Sigma y_{t-1}^2 - T(\Sigma y_{t-1}^2) + T(T+1)\Sigma y_{t-1} \Sigma y_{t-1} \\
- (T(T+1)(2T+1)/6) (\Sigma y_{t-1}^2)^2 . \]

It now follows that:

\[ T^{-6}D_X \Rightarrow (\sigma^2/12) \left[ \int_0^1 w(r) dr - 12 \left( \int_0^1 r w(r) dr \right)^2 \right] \\
+ 12 \int_0^1 w(r) dr \int_0^1 r w(r) dr - 4 \left( \int_0^1 r w(r) dr \right)^2 \]

\[ - (\sigma^2/12)D . \]

and thus
\[ T(\bar{a}-1) \Rightarrow (12/D) \left[ (1/2)W(1)\int rW(r)dr - (1/3)W(1)\int_0^1 W(r)dr \right. \\
\left. - (\int_0^1 rW(r)dr - (1/2)\int_0^1 W(r)dr)W(1) - \int_0^1 W(r)dr \right] + (1/24)(\bar{W}(1)^2 - \sigma_u^2/\sigma^2) \right] \]

\[ = (A_1 + (1/2)(\bar{W}(1)^2 - \sigma_u^2/\sigma^2))/D \]

as required for part (d). The proofs of (e), (f) and (g) follow in the same manner. We simply note here that:

\[ t_{\sigma} = D_X^{1/(2)}(\bar{a}-1)/\left\{ \sqrt{\left[ T^2(T^2 - 1)/12 \right]/s} \right\} \]

\[ \Rightarrow (\sigma/\sigma_u)(A_1 + (1/2)(\bar{W}(1)^2 - \sigma_u^2/\sigma^2))/D^{1/2} \]

as given in (e);

\[ t_{\mu} = \bar{\mu}_X^{1/2} \left[ (T(T^2+2)/12)\Sigma y_{t-1} - (\Sigma y_{t-1})^2 + T\Sigma y_{t-1}\Sigma y_{t-1} - (T^2/4)(\Sigma y_{t-1})^2 \right]^{-1/2}/s \]

\[ \Rightarrow (\sigma/\sigma_u)(A_2 - (1/2)(\bar{W}(1)^2 - \sigma_u^2/\sigma^2)/D^{1/2}A_3 \]

as given in (f); and

\[ t_{\beta} = \bar{\beta}X^{1/2} \left[ T\Sigma y_{t-1} - (\Sigma y_{t-1})^2 \right]^{-1/2}/s \]

\[ = T^{3/2}\bar{\tau}X^{1/2} \left[ T - (T^{-1/2}\Sigma y_{t-1})^2 \right]^{-1/2}/s \]

\[ \Rightarrow (\sigma/\sigma_u)(A_4 + \sqrt{3}(\bar{W}(1)^2 - \sigma_u^2/\sigma^2)(1/2)\int_0^1 W(r)dr - \int_0^1 rW(r)dr))/D^{1/2}A_5 \]

as given in (g).
Proof of Theorem 2. Note that

\[
Z(\hat{\alpha}) = T(\hat{\alpha} - \alpha) - (1/2) \left\{ T^{-2} \Sigma (y_{t-1} - \bar{y}_{t-1})^2 \right\}^{-1} \left( \hat{\sigma}_{T \ell}^2 - \hat{s}^2 \right) \\
\Rightarrow \left\{ \int_0^1 W(r)^2 dr - \left[ \int_0^1 W(r) dr \right]^2 \right\}^{-1} \left\{ (1/2)(W(1)^2 - \sigma_u^2/\sigma^2) - W(1) \int_0^1 W(r) dr \right\} \\
- (1/2)(1 - \sigma_u^2/\sigma^2) \\
= \left\{ \int_0^1 W(r)^2 dr - \left[ \int_0^1 W(r) dr \right]^2 \right\}^{-1} \left\{ (1/2)(W(1)^2 - 1) - W(1) \int_0^1 W(r) dr \right\}
\]

giving the result stated. The proofs for the other statistics follow the same lines.

Proof of Theorem 3. To prove part (a) we first note that

\[
T(\hat{\alpha} - \alpha) = \left\{ T^{-2} \Sigma (y_{t-1} - \bar{y}_{t-1})^2 \right\}^{-1} \left( T^{-1} \Sigma u_t (y_{t-1} - \bar{y}_{t-1}) \right) \\
\Rightarrow \left\{ \int_0^1 J_c(r)^2 dr - \left[ \int_0^1 J_c(r) dr \right]^2 \right\}^{-1} \left\{ \int_0^1 J_c(r) dW(r) \right\} \\
+ (1/2)(1 - \sigma_u^2/\sigma^2) - W(1) \int_0^1 J_c(r) dr)
\]

Additionally, \( \hat{s}^2 \to \sigma_u^2 \) and \( \hat{\sigma}_{T \ell}^2 \to \sigma^2 \) in probability. It follows from these results and the definition of \( Z(\hat{\alpha}) \) that

\[
Z(\hat{\alpha}) \Rightarrow c + \left\{ \int_0^1 J_c(r)^2 dr - \left[ \int_0^1 J_c(r) dr \right]^2 \right\}^{-1} \left\{ \int_0^1 J_c(r) dW(r) - W(1) \int_0^1 J_c(r) dr \right\}
\]

as required for (a). The proofs of (b) through (d) are entirely analogous.
REFERENCES


