Testing for a unit root in time series regression

BY PETER C. B. PHILLIPS

*Cowles Foundation for Research in Economics, Yale University, Yale Station,
New Haven, Connecticut 06520, U.S.A.*

AND PIERRE PERRON

*Département des Sciences Economiques, Université de Montréal, Montréal, Québec,
Canada H3C 3J7*

SUMMARY

This paper proposes 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 & Fuller. Simulations are reported on the performance of the new tests in finite samples.

Some key words: Brownian motion; Noncentral distribution; 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. Fuller (1984) and Dickey, Bell & Miller (1986) review much of the literature. The latter article provides a helpful practical guide to the use of some of the formal tests.

One major field of application where the hypothesis of a unit root has important implications is economics. This is because a unit root is often a theoretical implication of models which postulate the rational use of information that is available to economic agents. Examples include various financial market variables such as futures contracts (Samuelson, 1965), stock prices (Samuelson, 1973), dividends (Kleidon, 1986), spot and forward exchange rates (Meese & 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 & Plosser (1982), for example, found strong evidence in favour of unit root nonstationarity using the testing procedure of Dickey & Fuller (1979).

Recently, Said & 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 $(p, 1, q)$ process in which p and q are of unknown orders, that is for an autoregressive integrated moving average process of the indicated order. More specifically, Said & Dickey (1984) show that the Dickey-Fuller regression t test for a unit root may still be used in an ARIMA $(p, 1, q)$ 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 procedures.

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 nonzero 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 by 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 (Phillips, 1987b). Some simulation evidence on the finite sample performance of the new tests is also provided.

2. PRELIMINARIES

The models we consider are driven by a sequence of innovations denoted by $\{u_t\}$. Throughout we assume that $\{u_t\}$ satisfies the following general conditions:

   (i) $E(u_t) = 0$ for all t;
   (ii) $\sup_t E|u_t|^{\beta+\varepsilon} < \infty$ for some $\beta > 2$ and $\varepsilon > 0$;
   (iii) as $T \to \infty$, $\sigma^2 = \lim E(T^{-1}S_T^2)$ exists and $\sigma^2 > 0$, where $S_t = u_1 + \ldots + u_t$;
   (iv) $\{u_t\}$ is strong mixing with mixing coefficients α_m that satisfy $\Sigma\, \alpha_m^{1-2/\beta} < \infty$, where the sum is over $m = 1, \ldots, \infty$.

The conditions allow 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 (Withers, 1981). Condition (ii) controls the allowable heterogeneity of the process, whereas (iv) controls the extent of permissable temporal dependence in relation to the probability of outlier occurrences; see Phillips (1987a) for more discussion of these conditions, and Hall & Heyde (1980, p. 132) for the definition of strong mixing and the mixing coefficients. If $\{u_t\}$ is weakly stationary with spectral density $f_u(\lambda)$ then condition (iii) is a consequence

of (ii) and (iv). 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 $\{S_t\}$ we construct the random element

$$X_T(r) = T^{-\frac{1}{2}} \sigma^{-1} S_{[Tr]} = T^{\frac{1}{2}} \sigma^{-1} S_{j-1}$$

for $(j-1)/T \le r < j/T$ $(j = 1, \ldots, T)$, where $[Tr]$ denotes the integral part of Tr. Then $X_T(r)$ 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 (i)-(iv) above we have (Herrndorf, 1984) that, as $T \to \infty$,

$$X_T(r) \Rightarrow W(r). \tag{1}$$

The symbol '$\Rightarrow$' 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 $W(r)$; see Billingsley (1968, § 16) and Pollard (1984, Ch. 5) for further discussion.

Using (1) it is simple to deduce the asymptotic behaviour of the sample moments of the process $\{S_t\}$ and the innovations $\{u_t\}$. For a full development, see Phillips (1987a). We state below those results which are most useful subsequently. Unless otherwise indicated sums are over $t = 1, \ldots, T$. The limit distributions are expressed as functions of standard Brownian motion $W(r)$. To simplify formulae all integrals are understood to be taken over the interval $[0, 1]$, integrals such as $\int W, \int W^2, \int rW$ are understood to be taken with respect to Lebesgue measure and we write $W_1 = W(1)$. Then, as $T \to \infty$,

$$T^{-3/2} \sum S_t \Rightarrow \sigma \int W, \quad T^{-2} \sum S_t^2 \Rightarrow \sigma^2 \int W^2, \quad T^{-3/2} \sum tu_t \Rightarrow \sigma \int r \, dW = \sigma \left(W_1 - \int W \right),$$

$$T^{-5/2} \sum tS_t \Rightarrow \sigma \int rW, \quad T^{-1} \sum u_t S_{t-1} \Rightarrow \sigma^2 \int W \, dW + \lambda = \frac{1}{2}(\sigma^2 W_1^2 - \sigma_u^2),$$

where

$$\lambda = \frac{1}{2}(\sigma^2 - \sigma_u^2), \quad \sigma_u^2 = \lim_{T \to \infty} \frac{1}{T} \sum E(u_t^2).$$

Joint weak convergence for the sample moments given above to their respective limits is also easily established and will be used below.

3. THE MODELS AND ESTIMATORS

Let $\{y_t\}$ be a time series generated by

$$y_t = \alpha y_{t-1} + u_t \quad (t = 1, 2, \ldots), \tag{2}$$

$$\alpha = 1. \tag{3}$$

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 (i)-(iv).

We consider the two least-squares regression equations

$$y_t = \hat{\mu} + \hat{\alpha} y_{t-1} + \hat{u}_t, \tag{4}$$

$$y_t = \tilde{\mu} + \tilde{\beta}(t - \tfrac{1}{2}T) + \tilde{\alpha} y_{t-1} + \tilde{u}_t, \tag{5}$$

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

$$t_{\hat{\alpha}} = (\hat{\alpha} - \alpha)\{\sum (y_{t-1} - \bar{y}_{-1})^2\}^{\frac{1}{2}}/\hat{s}, \quad t_{\hat{\mu}} = (\hat{\mu} - \mu)\{\sum (y_{t-1} - \bar{y}_{-1})^2/\sum y_{t-1}^2\}^{\frac{1}{2}}/\hat{s},$$

$$t_{\tilde{\mu}} = (\tilde{\mu} - \mu)/(\tilde{s}^2 c_1)^{\frac{1}{2}}, \quad t_{\tilde{\beta}} = (\tilde{\beta} - \beta)/(\tilde{s}^2 c_2)^{\frac{1}{2}}, \quad t_{\tilde{\alpha}} = (\tilde{\alpha} - \alpha)/(\tilde{s}^2 c_3)^{\frac{1}{2}},$$

where $\hat{s}$ and $\tilde{s}$ are the standard errors of regressions (4) and (5), c_i is the ith diagonal element of the matrix $(X'X)^{-1}$, and $\bar{y}_{-1} = T^{-1} \sum y_{t-1}$.

Following Dickey & Fuller (1979), we shall be concerned with the limiting distributions of the regression coefficients in (4) and (5) and their t statistics under the 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, the coefficient $\tilde{\alpha}$ and its t statistic in (5) are invariant with respect to the introduction of a nonzero drift $\mu \neq 0$ in the generating process. Thus we may replace (2) with the generating mechanism

$$y_t = \mu + \alpha y_{t-1} + u_t, \quad (t = 1, 2, \ldots), \tag{2'}$$

and the distributions and asymptotic distributions of the above-mentioned statistics are unchanged. Thus there is no loss in generality by assuming $\mu = 0$.

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

$$T^{-3/2} \sum y_t \Rightarrow \sigma \int W, \quad T^{-2} \sum y_t^2 \Rightarrow \sigma^2 \int W^2,$$

$$T^{-5/2} \sum t y_{t-1} \Rightarrow \sigma \int rW, \quad T^{-1} \sum y_{t-1} u_t \Rightarrow \sigma^2 \int W \, dW + \lambda.$$

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) $T(\hat{\alpha} - 1) \Rightarrow (\int W_*^2)^{-1}(\int W_* \, dW + \lambda'),$
  (b) $t_{\hat{\alpha}} \Rightarrow (\sigma/\sigma_u)(\int W_*^2)^{-\frac{1}{2}}(\int W_* \, dW + \lambda'),$
  (c) $t_{\hat{\mu}} \Rightarrow (\sigma/\sigma_u)(\int W^2 \int W_*^2)^{-\frac{1}{2}}\{W_1 \int W^2 - (\int W \, dW + \lambda') \int W\}.$
For the regression model (5), *as* $T \to \infty$,
  (d) $T(\tilde{\alpha} - 1) \Rightarrow (\int W \, dW + \lambda' + A_1)/D,$
  (e) $t_{\tilde{\alpha}} \Rightarrow (\sigma/\sigma_u)(\int W \, dW + \lambda' + A_1)/D^{\frac{1}{2}},$
  (f) $t_{\tilde{\mu}} \Rightarrow (\sigma/\sigma_u)\{A_2 - (\int W \, dW + \lambda') \int W\}/D^{\frac{1}{2}}A_3,$
  (g) $t_{\tilde{\beta}} \Rightarrow (\sigma/\sigma_u)\{A_4 + (\int W \, dW + \lambda')(\tfrac{1}{2}\int W - \int rW)\}/(D/12)^{\frac{1}{2}}(\int W_*^2)^{\frac{1}{2}},$

where

$$D = \int W^2 - 12\left(\int rW\right)^2 + 12 \int W \int rW - 4\left(\int W\right)^2,$$

$$A_1 = 12\left(\int rW - \frac{1}{2}\int W\right)\left(\int W - \frac{1}{2} W_1\right) - W_1 \int W,$$

$$A_2 = W_1\left\{\int W^2 - 12\left(\int rW\right)^2 + 18 \int W \int rW - 6\left(\int W\right)^2\right\}$$

$$+ 6 \int W\left\{\left(\int W\right)^2 - 2 \int rW \int W\right\},$$

$$A_3 = \left\{D + \left(\int W\right)^2\right\}^{\frac{1}{2}}, \quad A_4 = W_1\left\{\frac{1}{2}\int W^2 + \int W \int rW - \left(\int W\right)^2\right\} - \int W \int W_*^2,$$

$$W_*(r) = W(r) - \int W, \quad \lambda' = \lambda/\sigma^2.$$

Note that $W_*(r)$ in the above formulae can be called a Brownian motion. The stated results follow in a direct way from the asymptotic behaviour of the sample moments given earlier. In some cases the calculations are lengthy and full details will be supplied on request.

When the innovation sequence $\{u_t\}$ is independent and identically distributed, we have $\sigma_u^2 = \sigma^2$ and $\lambda = \lambda' = 0$. 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 & Fuller. Specifically, critical values of $T(\hat{\alpha} - 1)$, $T(\tilde{\alpha} - 1)$, $t_{\hat{\alpha}}$ and $t_{\tilde{\alpha}}$ are tabulated by Fuller (1976, Tables 8.5.1, 8.5.2); and tabulated critical values of $t_{\hat{\mu}}$, $t_{\tilde{\mu}}$ and $t_{\hat{\beta}}$ are also given by Dickey & Fuller (1981, Tables I-III).

Theorem 1 extends the results of Dickey & 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 § 5.

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 nonorthogonal 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 § 4 all depend upon the nuisance parameters σ^2 and σ_u^2. This presents an obstacle to conventional procedures of inference in the general case where $\sigma^2 \neq \sigma_u^2$. However, since σ^2 and σ_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 σ_u^2 are provided by $\hat{s}^2$, $\tilde{s}^2$ and $s^2 = T^{-1}\Sigma(y_t - y_{t-1})^2$ for data generated by (2) and (3). When there is a nonzero drift in the model, as in (2'), both $\hat{s}^2$ and $\tilde{s}^2$ are consistent. Since we often wish to allow for a nonzero drift in regressions such as (5) we use $\tilde{s}^2$ as our preferred estimator of σ_u^2 in this regression.

Consistent estimation of σ^2 is discussed by 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 σ^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, namely

$$s_{Tl}^2 = T^{-1} \sum_{t=1}^{T} u_t^2 + 2T^{-1} \sum_{s=1}^{l} \sum_{t=s+1}^{T} u_t u_{t-s}, \tag{6}$$

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

   (ii') $\sup_t E|u_t|^{2\beta} < \infty$ for some $\beta > 2$

holds in place of (ii) of § 2, and provided

   (v) $l \to \infty$ as $T \to \infty$ and $l^4/T \to 0$.

According to this, if we allow the number of estimated autocovariances in (6) to increase as $T \to \infty$ but control the rate of increase so that $l^4/T \to 0$, then s_{Tl}^2 yields a consistent estimator of σ^2. Of course, faster rates of increase in l 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 (ii') and (v) hold, in addition to (i), (iii) and (iv) given earlier.

Rather than using first differences $u_t = y_t - y_{t-1}$ in the construction of s_{Tl}^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_{Tl}^2, which we denote by $\hat{s}_{Tl}^2$ and $\tilde{s}_{Tl}^2$ respectively, are also consistent estimates of σ^2 under the same conditions. Once again $\tilde{s}_{Tl}^2$ will be the preferred estimator when we wish to allow for a nonzero drift as in (2').

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

$$\hat{\sigma}_{Tl}^2 = T^{-1} \sum_{t=1}^{T} \hat{u}_t^2 + 2T^{-1} \sum_{s=1}^{l} w_{sl} \sum_{t=s+1}^{T} \hat{u}_t \hat{u}_{t-s}, \tag{7}$$

$$\tilde{\sigma}_{Tl}^2 = T^{-1} \sum_{t=1}^{T} \tilde{u}_t^2 + 2T^{-1} \sum_{s=1}^{l} w_{sl} \sum_{t=s+1}^{T} \tilde{u}_t \tilde{u}_{t-s}, \tag{8}$$

where

$$w_{sl} = 1 - s/(l+1), \tag{9}$$

are nonnegative and, for stationary $\{u_t\}$, are simply 2π times the corresponding Bartlett estimates of $f_u(0)$. Other choices of lag window besides the triangular window (9) are possible. We use the Parzen window in the simulations reported in § 7. The estimator (7) was recently suggested in the context of variance estimates by Newey & 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

$$Z(\hat{\alpha}) = T(\hat{\alpha} - 1) - \hat{\lambda}/\bar{m}_{yy}, \quad Z(t_{\hat{\alpha}}) = (\hat{s}/\hat{\sigma}_{Tl})t_{\hat{\alpha}} - \hat{\lambda}'\hat{\sigma}_{Tl}/\bar{m}_{yy}^{\frac{1}{2}},$$

$$Z(t_{\hat{\mu}}) = (\hat{s}/\hat{\sigma}_{Tl})t_{\hat{\mu}} + \hat{\lambda}'\hat{\sigma}_{Tl}m_y/\bar{m}_{yy}^{\frac{1}{2}}m_{yy}^{\frac{1}{2}}, \quad Z(\tilde{\alpha}) = T(\tilde{\alpha} - 1) - \tilde{\lambda}/M,$$

$$Z(t_{\tilde{\alpha}}) = (\tilde{s}/\tilde{\sigma}_{Tl})t_{\tilde{\alpha}} - \tilde{\lambda}'\tilde{\sigma}_{Tl}/M^{\frac{1}{2}}, \quad Z(t_{\tilde{\mu}}) = (\tilde{s}/\tilde{\sigma}_{Tl})t_{\tilde{\mu}} - \tilde{\lambda}'\tilde{\sigma}_{Tl}m_y/M^{\frac{1}{2}}(M + m_y^2)^{\frac{1}{2}},$$

$$Z(t_{\tilde{\beta}}) = (\tilde{s}/\tilde{\sigma}_{Tl})t_{\tilde{\beta}} - \tilde{\lambda}'\tilde{\sigma}_{Tl}(\tfrac{1}{2}m_y - m_{ty})/(M/12)^{\frac{1}{2}}\bar{m}_{yy}^{\frac{1}{2}},$$

where

$$m_{yy} = T^{-2}\sum y_t^2, \quad \bar{m}_{yy} = T^{-2}\sum (y_t - \bar{y})^2, \quad m_y = T^{-3/2}\sum y_t, \quad m_{ty} = T^{-5/2}\sum ty_t,$$

$$M = (1 - T^{-2})m_{yy} - 12m_{ty}^2 + 12(1 + T^{-1})m_{ty}m_y - (4 + 6T^{-1} + 2T^{-2})m_y^2,$$

$$\hat{\lambda} = \tfrac{1}{2}(\hat{\sigma}_{Tl}^2 - \hat{s}^2), \quad \hat{\lambda}' = \hat{\lambda}/\hat{\sigma}_{Tl}^2, \quad \tilde{\lambda} = \tfrac{1}{2}(\tilde{\sigma}_{Tl}^2 - \tilde{s}^2), \quad \tilde{\lambda}' = \tilde{\lambda}/\tilde{\sigma}_{Tl}^2.$$

These Z statistics extend those of 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 $\tilde{s}$ which measure scale effects in the conventional t ratios are now replaced by the general standard error estimates $\hat{\sigma}_{Tl}$ and $\tilde{\sigma}_{Tl}$ which allow for serial convariation as well as variance. Each Z statistic also involves an additive correction term whose magnitude depends on the difference between the corresponding variance estimates $\hat{\sigma}_{Tl}^2 - \hat{s}^2$ or $\tilde{\sigma}_{Tl}^2 - \tilde{s}^2$. 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, which follows simply from Theorem 1.

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

(b) *For the regression model* (5) *the statistics* $Z(\tilde{\alpha})$, $Z(t_{\tilde{\alpha}})$, $Z(t_{\tilde{\mu}})$ *and* $Z(t_{\tilde{\beta}})$ *have limit distributions given by those of* $T(\tilde{\alpha} - 1)$, $t_{\tilde{\alpha}}$, $t_{\tilde{\mu}}$ *and* $t_{\tilde{\beta}}$, *respectively, in Theorem* 1 *with* $\sigma^2 = \sigma_u^2$. *The stated results for* $Z(\tilde{\alpha})$ *and* $Z(t_{\tilde{\alpha}})$ *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 § 4, when $\sigma_u^2 = \sigma^2$. Thus, the critical values derived in the studies of Dickey & 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

$$\alpha = e^{c/T} \sim 1 + c/T. \tag{10}$$

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 by Phillips (1987b). Time series generated by models such as (2) or (2') with a coefficient α 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 $T \to \infty$,

$$T^{-3/2} \sum y_t \Rightarrow \sigma \int J_c, \quad T^{-2} \sum y_t^2 \Rightarrow \sigma^2 \int J_c^2,$$

$$T^{-5/2} \sum t y_t \Rightarrow \sigma \int r J_c, \quad T^{-1} \sum y_{t-1} u_t \Rightarrow \sigma^2 \int J_c \, dW + \lambda,$$

where

$$J_c(r) = \int_0^r e^{(r-s)c} \, dW(s)$$

is the Ornstein–Uhlenbeck process 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 $\int J_c \, dW$ 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 § 6. The main results of interest are contained in the following theorem which concentrates on estimates of the autoregressive coefficient α and its associated t-ratio. The derivations follow those of Phillips (1987b) and will be supplied on request.

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

   (a) $Z(\hat\alpha) \Rightarrow c + (\int J_c^{*2})^{-1} \int J_c^* \, dW,$
   (b) $Z(t_{\hat\alpha}) \Rightarrow c(\int J_c^{*2})^{\frac{1}{2}} + (\int J_c^{*2})^{-\frac{1}{2}} \int J_c^* \, dW,$
   (c) $Z(\tilde\alpha) \Rightarrow c + (\int J_c \, dW + A_{1c})/D_c,$
   (d) $Z(t_{\tilde\alpha}) \Rightarrow cD_c^{\frac{1}{2}} + (\int J_c \, dW + A_{1c})/D_c^{\frac{1}{2}},$

where

$$D_c = \int J_c^2 - 12 \left(\int r J_c \right)^2 + 12 \int r J_c \int J_c - 4 \left(\int J_c \right)^2,$$

$$A_{1c} = 6 W_1 \int r J_c - 4 W_1 \int J_c - 12 \int r \, dW \int r J_c + 6 \int r \, dW \int J_c, \quad J_c^*(r) = J_c(r) - \int J_c.$$

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 & Dickey (1984). As explained earlier, Said & Dickey recommend the use of the Dickey–Fuller regression t test for a unit root in the autoregression

$$\Delta y_t = \hat{\mu} + \hat{\alpha}_* y_{t-1} + \sum_{i=1}^{l} \hat{\phi}_i \Delta y_{t-i} + \hat{v}_t. \tag{11}$$

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

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

$$u_t = e_t + \theta e_{t-1} \tag{12}$$

and the e_t independent and identically distributed $N(0, 1)$. We set $y_0 = 0$, and used various lag lengths l in (11) and lag truncations l 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 $\hat{\sigma}_{Tl}^2$. The simulations reported in Table 1 are based on 2000 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 θ in (12). When $\theta = 0$ there is no need to employ the transformations leading to $Z(\hat{\alpha})$ and $Z(t_{\hat{\alpha}})$ 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 $Z(t_{\hat{\alpha}})$ tests. In fact, the $Z(\hat{\alpha})$ test is conservative and at the same time has greater power than either $Z(t_{\hat{\alpha}})$ or $t(\hat{\alpha}_*)$ for all choices of l. The $Z(t_{\hat{\alpha}})$ and $t(\hat{\alpha}_*)$ tests are both liberal in terms of size at $T = 100$ and the size distortions of $t(\hat{\alpha}_*)$ increase appreciably with the length of the autoregression. At the same time, the 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 $l \ge 4$. 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 $l \ge 4$ and the size distortions increase with the length of the autoregression. These results again suggest that $Z(\hat{\alpha})$ is the preferred test.

Table 1. *Simulations based on* 2000 *replications*; $T = 100$, *nominal size* 5%

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

| | Size | | | | | Power ($\alpha = 0.85$) | | | | |
θ	$l=2$	$l=4$	$l=6$	$l=8$	$l=12$	$l=2$	$l=4$	$l=6$	$l=8$	$l=12$
0·0	0·068	0·064	0·078	0·086	0·106	0·557	0·472	0·406	0·354	0·303
0·5	0·052	0·064	0·071	0·085	0·102	0·378	0·411	0·377	0·348	0·308
0·8	0·037	0·051	0·072	0·082	0·109	0·267	0·302	0·313	0·308	0·273
−0·2	0·063	0·065	0·084	0·092	0·111	0·591	0·490	0·421	0·358	0·304
−0·5	0·116	0·076	0·073	0·086	0·105	0·868	0·626	0·512	0·434	0·350
−0·8	0·677	0·343	0·201	0·142	0·120	1·00	0·988	0·900	0·772	0·523

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

| | Size | | | | | Power ($\alpha = 0.85$) | | | | |
θ	$l=2$	$l=4$	$l=6$	$l=8$	$l=12$	$l=2$	$l=4$	$l=6$	$l=8$	$l=12$
0·0	0·044	0·049	0·049	0·049	0·055	0·772	0·776	0·784	0·794	0·805
0·5	0·010	0·021	0·022	0·021	0·014	0·301	0·524	0·534	0·500	0·378
0·8	0·003	0·020	0·024	0·022	0·015	0·247	0·513	0·526	0·474	0·330
−0·2	0·134	0·105	0·110	0·115	0·133	0·960	0·940	0·946	0·952	0·967
−0·5	0·537	0·417	0·428	0·451	0·516	1·00	1·00	1·00	1·00	1·00
−0·8	0·997	0·988	0·988	0·991	0·995	1·00	1·00	1·00	1·00	1·00

(c) $Z(t_{\hat{\alpha}})$ *test*

| | Size | | | | | Power ($\alpha = 0.85$) | | | | |
θ	$l=2$	$l=4$	$l=6$	$l=8$	$l=12$	$l=2$	$l=4$	$l=6$	$l=8$	$l=12$
0·0	0·063	0·062	0·066	0·069	0·069	0·669	0·688	0·696	0·705	0·705
0·5	0·028	0·036	0·038	0·035	0·030	0·191	0·367	0·371	0·333	0·214
0·8	0·026	0·034	0·035	0·031	0·026	0·144	0·353	0·357	0·305	0·167
−0·2	0·135	0·115	0·119	0·127	0·140	0·933	0·910	0·918	0·930	0·945
−0·5	0·534	0·438	0·445	0·460	0·518	1·00	1·00	1·00	1·00	1·00
−0·8	0·995	0·988	0·990	0·993	0·995	1·00	1·00	1·00	1·00	1·00

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 by 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

$$T(\hat{\alpha} - 1) \equiv \left(\int W_*^2 \right)^{-1} \left\{ \int W_* \, dW - k_0 - (2T)^{-\frac{1}{2}} k_1 \eta \right\} + O_p(T^{-1}),$$

where $\equiv$ signifies equivalence in distribution,

$$k_0 = \tfrac{1}{2}\sigma_u^2 / \sigma^2 = \tfrac{1}{2}(1 + \theta^2)/(1 + \theta)^2, \quad k_1 = (1 + 4\theta^2 + \theta^4)^{\frac{1}{2}}/(1 + \theta)^2,$$

and η 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 θ. 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 $T(\hat{a} - 1)$ 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 behaviour may be inferred for the test statistics $Z(\hat{a})$ and $Z(t_{\hat{a}})$, which are both based on $T_{(\hat{a}-1)}$. This helps to explain the size distortions in these tests that were evident in the simulations when $\theta < 0$.

8. CONCLUDING COMMENTS

The present approach gives a simple 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.

As we have seen in § 6, there is no loss in asymptotic local power in the use of the $Z(\hat{a})$ tests for a unit root. But the simulations reported in § 7 indicate that test performance can differ substantially in finite samples among asymptotically equivalent tests. For models with positive moving average errors the $Z(\hat{a})$ 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, $Z(\hat{a})$ again seems to be the preferred test. For models with moving average errors and negative serial correlation the Z tests suffer appreciable size distortions and are not recommended. In such cases the Said–Dickey procedure of using a long autoregression seems preferable.

ACKNOWLEDGMENTS

REFERENCES

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. New York: Wiley.

DICKEY, D. A., BELL, W. R. & MILLER, R. B. (1986). Unit roots series model: Tests and implications. *Am. Statistician* **40**, 12-26.

DICKEY, D. A. & FULLER, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Statist. Assoc.* **74**, 427-31.

DICKEY, D. A. & FULLER, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* **49**, 1057-72.

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

FULLER, W. A. (1984). Nonstationary autoregressive time series. In *Handbook of Statistics*, **5**, *Time Series in the Time Domain*, Ed. E. J. Hannan, P. R. Krishnaiah and M. M. Rao, Ch. 1, pp. 1-23. Amsterdam: North Holland.

HALL, P. & HEYDE, C. C. (1980). *Martingale Limit Theory and its Applications*. New York: Academic Press.

HALL, R. E. (1978). Stochastic implications of the life cycle-permanent income hypothesis: Theory and evidence. *J. Polit. Econ.* **86**, 971-87.

HERRNDORF, N. (1984). A functional central limit theorem for weakly dependent sequences of random variables. *Ann. Prob.* **12**, 141-53.

KLEIDON, A. W. (1986). Variance bounds tests and stock price valuation models. *J. Polit. Econ.* **94**, 953-1001.

MEESE, R. A. & SINGLETON, K. J. (1983). On unit roots and the empirical modeling of exchange rates. *Int. Econ. Rev.* **24**, 1029-35.

NELSON, C. R. & PLOSSER, C. I. (1982). Trends and random walks in macroeconomic time series: Some evidence and implications. *J. Monet. Econ.* **10**, 139-62.

NEWEY, W. K. & WEST, K. D. (1987). A simple, positive definite, heteroskedasticity and autocorrelation · consistent covariance matrix. *Econometrica* **55**, 703-8.

PHILLIPS, P. C. B. (1987a). Time series regression with a unit root. *Econometrica* **55**, 277-301.

PHILLIPS, P. C. B. (1987b). Towards a unified asymptotic theory of autoregression. *Biometrika* **74**, 535-48.

PHILLIPS, P. C. B. (1987c). Asymptotic expansions in nonstationary vector autoregressions. *Econometric Theory* **2**, 45-68.

POLLARD, D. (1984). *Convergence of Stochastic Processes.* New York: Springer-Verlag.

SAID, S. E. & DICKEY, D. A. (1984). Testing for unit roots in autoregressive-moving average of unknown order. *Biometrika* **71**, 599-607.

SAMULESON, P. A. (1965). Proof that properly anticipated prices fluctuate randomly. *Indust. Manag. Rev.* **6**, 41-50.

SAMUELSON, P. A. (1973). Proof that properly discounted present values of assets vibrate randomly. *Bell J. Econ. Manag. Sci.* **4**, 369-74.

WITHERS, C. S. (1981). Conditions for linear processes to be strong mixing. *Z. Wahr. verw. Geb.* **57**, 477-80.