A PRIMER ON UNIT ROOT TESTING

BY

PETER C.B. PHILLIPS
AND
ZHIJIE XIAO

COWLES FOUNDATION PAPER NO. 972

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281
1999

A PRIMER ON UNIT ROOT TESTING

Peter C. B. Phillips
Yale University

Zhijie Xiao
University of Illinois at Urbana-Champaign

Abstract. The immense literature and diversity of unit root tests can at times be confusing even to the specialist and presents a truly daunting prospect to the uninitiated. In consequence, much empirical work still makes use of the simplest testing procedures because it is unclear from the literature and from recent reviews which tests if any are superior. This paper presents a survey of unit root theory with an emphasis on testing principles and recent developments. The general framework adopted makes it possible to consider tests of stochastic trends against trend stationarity and trend breaks of a general type. The main tests are listed, and asymptotic distributions are given in a simple form that emphasizes commonalities in the theory. Some simulation results are reported, and an extensive list of references and all annotated bibliography are provided.

Keywords. Autoregressive unit root; Brownian motion; Functional central limit theorem; Integrated process; LM principle; Model selection; Moving average unit root; Nonstationarity; Quasi-differencing; Stationarity; Stochastic trend.

1. Introduction

At a casual level, many observed time series seem to display nonstationary characteristics. For economic time series nonstationary behavior is often the most dominant characteristic. Some series grow in a secular way over long periods of time, others appear to wander around as if they have no fixed population mean. Growth characteristics are especially evident in time series that represent aggregate economic behavior like gross domestic product and industrial production. Random wandering behavior is evident in many financial time series like interest rates and asset prices. Similar phenomena arise in data from other social sciences like communications and political science, one example being opinion poll data on presidential popularity. Any attempt to explain or forecast series of this type requires that a mechanism be introduced to capture the nonstationary elements in the series, or that the series be transformed in some way to achieve stationarity. Yet this is often much easier to say than it is to do in a satisfactory way. The problem is particularly delicate in the multivariate case, where several time series may have nonstationary characteristics and the interrelationships of these variables are the main object of study.

Before 1970, a very popular way of modeling nonstationarity was to use

0950-0804/98/05 0423-47 JOURNAL OF ECONOMIC SURVEYS Vol. 12, No. 5

deterministic trending functions like time polynomials to capture the secular movements in the series. Regression methods were commonly used to extract this trend and the residuals were then analyzed as a stationary time series. A model of the form

$$y_t = h_t + y_t^s,\ h_t = \gamma' x_t,\ t = 1, \ldots, n, \tag{1}$$

where y_t^s is a *stationary time series* and x_t is a k-vector of deterministic trends, is known as a *trend-stationary* time series. The trend function h_t may be more complex than a simple time polynomial. For example, time polynomials with sinusoidal factors and piecewise time polynomials may be used. The latter corresponds to a class of models with structural breaks in the deterministic trend. The outline of the theory of unit root tests that is given here will allow for these possibilities.

A major limitation of models like (1) is that the trending mechanism is non-stochastic. One way of introducing a stochastic element into the trend is to allow the process y_t^s to be generated as follows:

$$y_t^s = \alpha y_{t-1}^s + u_t,\ t = 1, \ldots n, \text{ with } \alpha = 1,\ u_t = C(L)\varepsilon_t, \tag{2}$$

and

$$C(L) = \sum_{j=0}^{\infty} c_j L^j,\ \sum_{j=0}^{\infty} j^{1/2} |c_j| < \infty,\ C(1) \neq 0, \tag{3}$$

where L is the lag operator for which $Ly_t = y_{t-1}$. The initial condition in (2) is set at $t = 0$, and y_0^s may be a constant or a random variable. In the latter case, we can even allow for distant initial conditions (see Phillips and Lee, 1996, and Canjels and Watson, 1997), so that y_0^s has a comparable stochastic order to the terminal data point y_n^s, viz. $O_p(\sqrt{n})$. If $\varepsilon_t \equiv \mathrm{iid}(0, \sigma^2)$, then u_t in (2) is a *linear process*, and y_t^s has an autoregressive unit root. It is common to make the more general assumption that ε_t is a stationary martingale difference sequence with respect to the natural filtration $(\mathcal{F}_t)$ with $E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = \sigma^2$, a.s. The second and third conditions of (3) ensure that u_t is *covariance stationary* and has positive *spectral density* at the origin, thereby ensuring that the unit root in y_t does not cancel (as it would if u_t had a *moving average unit root*). The $\frac{1}{2}$-summability condition in (3) is useful in validating the following expansion of the operator $C(L)$

$$C(L) = C(1) + \tilde{C}(L)(L - 1), \tag{4}$$

where $\tilde{C}(L) = \sum_{j=0}^{\infty} \tilde{c}_j L^j$ and $\tilde{c}_j = \sum_{j+1}^{\infty} c_s$. This expansion gives rise to an explicit martingale difference decomposition of u_t

$$u_t = C(1)\varepsilon_t + \tilde{\varepsilon}_{t-1} - \tilde{\varepsilon}_t, \quad \text{with } \tilde{\varepsilon}_t = \tilde{C}(L)\varepsilon_t, \tag{5}$$

This decomposition is sometimes called the martingale decomposition in the probability literature (see Hall and Heyde, 1980) because the first term of (5) is a martingale difference and the partial sums $\sum_{s=1}^{t} u_s$ correspondingly have the leading martingale term $C(1) \sum_{s=1}^{t} \varepsilon_s$. The expansion (4) was obtained, but not validated, in the work of Beveridge and Nelson (1981) on decomposing aggregated economic data into long run and short run components and is significant in this context. Thus, if y_t^s is generated by (2) then (5) reveals that

$$y_t^s = C(1)\sum_{s=1}^{t}\varepsilon_s + \tilde{\varepsilon}_0 - \tilde{\varepsilon}_t + y_0^s = Y_t^s + \eta_t,$$

where $Y_t^s = C(1)\sum_{s=1}^{t}\varepsilon_s$ and $\eta_t = \tilde{\varepsilon}_0 - \tilde{\varepsilon}_t + y_0^s$ are the long run and short run components of y_t^s respectively. The decomposition (5) was justified in a simple algebraic way using (3) by Phillips and Solo (1992), who showed how to use it to prove strong laws, central limit theorems, functional laws, and laws of iterated logarithms for time series. It is now commonly used in this way in the development of an asymptotic theory for nonstationary time series.

The output of (2) can be written as the accumulated process $y_t^s = \sum_{j=1}^{t} u_j + y_0^s$, and is called a *stochastic trend* by virtue of the fact that it is of stochastic order $O_p(t^{1/2})$. The process y_t^s is *difference stationary* in the sense that $\Delta y_t^s = u_t$ is a stationary process where $\Delta = 1 - L$ is the differencing operator. The terminology *integrated process* of order one (written as $I(1)$) is in common use because of the above representation. In consequence, we call a stationary time series an $I(0)$ process (integrated of order zero). The assumptions given above are sufficient to ensure that y_t^s satisfies a *functional central limit theorem* (see Phillips and Solo, 1992, for a demonstration), which is an important element in the development of the asymptotic theory of all unit root tests. As a result, $n^{-1/2}y_{[nr]}^s \Rightarrow B(r)$, a Brownian motion with variance $\omega^2 = \sigma^2 C(1)^2$, where $[nr]$ signifies the integer part of nr and $r \in [0,1]$ represents some fraction of the sample data. The parameter $\omega^2 = 2\pi f_u(0)$ is called the *long-run variance* of u_t.

In contrast to stationary or trend stationary time series, models with a stochastic trend have time dependent variances that go to infinity with time, are persistent in the sense that shocks have permanent effects on the values of the process, and have infinite spectrum at the origin. These properties of stochastic trends have considerable relevance in economic applications. For instance, under the real business cycle hypothesis, policy actions are required to bring real GNP back to its original path due to the persistent effects of innovations. However, less policy intervention is needed in trend stationary models because shocks only have a transitory effect.

Testing for the presence of a stochastic trend in the model (1) is equivalent to testing the null hypothesis that the autoregressive parameter $\alpha = 1$ in (2), and is known as a *unit root test*. The alternative hypothesis that $|\alpha| < 1$ corresponds to the version of the model in which y_t is trend stationary. The test can be interpreted as a test of difference stationarity versus trend-stationarity in the time series y_t. There are now a wide variety of such tests, based on parametric, semi-parametric and non-parametric methods and employing both classical and Bayesian principles of statistical testing. The literature is immense. This paper seeks to cover the main principles of testing, the most commonly used tests in practical work, a comparison of the finite sample properties among these tests, and recent developments.

2. Classical unit root tests

2.1. *The Dickey-Fuller tests*

Combining (1) and (2) gives the regression model

$$y_t = \beta' x_t + \alpha y_{t-1} + u_t. \tag{6}$$

In (6) the deterministic component is constructed so that $\Delta h_t = \gamma'\Delta x_t = \gamma' A x_t = \beta' x_t$, for some matrix A and vector β. This usually involves raising the degree of the deterministic trends to ensure that the maximum trend degrees in (6) and (1) are the same. Then, at least one element of the parameter vector β is zero and, consequently, there are surplus trend variables in the regression equation (6). It will subsequently be useful to make this redundancy explicit, and this can be done by rewriting the trend component in (6) as $\Delta h_t = \beta' x_t = \breve{\beta}'\breve{x}_t$, where $\breve{x}_t = Sx_t$ for some eliminator matrix S that eliminates redundant rows of x_t. The formulation (6) therefore results in some inefficiency in the regression because $\breve{x}_t$ is of smaller dimension than x_t. There is an alternative approach that avoids this problem of redundant variables and it will be discussed in Sections 3.1 and 3.2 below. The regression (6) does have the advantage that the detrended data is invariant to the parameters in the trend function in (1).

To develop an asymptotic theory it is assumed that there exists a scaling matrix D_n and a piecewise continuous function $X(r)$ such that $D_n^{-1} x_{[nr]} \to X(r)$ as $n \to \infty$ uniformly in $r \in [0, 1]$. For example, if h_t is a p-degree time polynomial, then $D_n = \mathrm{diag}(1, n, \ldots, n^p)$ and $X(r) = (1, r, \ldots, r^p)'$. Correspondingly, it is also assumed that there exists a matrix F_n for which $F_n^{-1}\breve{x}_{[nr]} \to \breve{X}(r)$ as $n \to \infty$, uniformly in $r \in [0, 1]$. In general, $\breve{X}(r) = SX(r)$.

The Dickey-Fuller tests (Dickey and Fuller, 1979, 1981) dealt with Gaussian random walks with independent residuals. Let $\hat{\alpha}$ be the OLS estimator of α in (6) and $t_{\hat{\alpha}}$ be the corresponding t-ratio statistics, under the null hypothesis $\alpha = 1$, the large sample theory for these quantities involves functionals of Brownian motion, some of which are stochastic integrals. The limit theory forms the basis of the unit root tests. If the shocks u_t are $iid(0, \sigma^2)$ random variates, the large sample theory for the coefficient estimator $\hat{\alpha}$ and its regression t-ratio statistic $t_{\hat{\alpha}}$ are given by the following functionals of Brownian motion,

$$n(\hat{\alpha} - 1) \Rightarrow \left[\int_0^1 W_X(r) dW(r)\right]\left[\int_0^1 W_X(r)^2\right]^{-1}, \tag{7}$$

$$t_{\hat{\alpha}} \Rightarrow \left[\int_0^1 W_X(r) dW(r)\right]\left[\int_0^1 W_X(r)^2\right]^{-1/2}, \tag{8}$$

where $\Rightarrow$ signifies weak convergence, W is standard Brownian motion, and $W_X(r) = W(r) - \int_0^1 WX'(\int_0^1 XX')^{-1}X(r)$ is the Hilbert projection in $L_2[0, 1]$ of W onto the space orthogonal to X. In the special case where there is no deterministic component x_t, these limit distributions reduce to the commonly known

Dickey-Fuller distributions given by the functionals $[\int_0^1 WdW][\int_0^1 W^2]^{-1}$ and $[\int_0^1 WdW][\int_0^1 W^2]^{-1/2}$. Dickey and Fuller (1979, 1981) did not themselves use these representations, but used equivalent $\mathbb{R}_\infty$ formulations in terms of linear combinations of functions of iid $N(0, 1)$ variates, rather than stochastic process representations on function spaces. The latter first appeared, but were unproved, in White (1958), and were later developed in progressive degrees of generality by Lai and Wei (1982), Solo (1984) and Phillips (1987). Multivariate regression cases were dealt with using these methods in Phillips (1986) and Phillips and Durlauf (1986).

In the more general case where the residual process u_t is stationary, the limit distributions of $\hat{a}$ and $t_{\hat{a}}$ have additional bias terms due to the presence of serial correlation. These were explored in Phillips (1987a). As a result, the limiting distributions of the two statistics in (7) and (8) become dependent on the nuisance parameters. Such a problem can be solved in a parametric or non-parametric way, leading to two major classes of unit root tests that are distinguished by their treatment of the autocorrelation in the stationary residual process u_t. One approach, proposed by Phillips (1987a), adjusts $\hat{a}$ and $t_{\hat{a}}$ based on nonparametric estimates of the nuisance parameters to account for the serial correlation. This approach is said to be semi-parametric since its treatment of the regression coefficient a is parametric but it deals with the stationary residual nonparametrically. The second approach, the augmented Dickey-Fuller (ADF) test, adds lags to the autocorrelation to eliminate the effect of serial correlation on the test statistics. Such a device relies on specifying the stationary part of the process in terms of a parametric model (commonly an autoregression) and is therefore fully parametric. This approach was explored by Said and Dickey (1984).

2.2. *The semi-parametric Z_a and Z_t tests*

The semiparametric Z_a and Z_t tests were developed in Phillips (1987a) and extend the original unit root tests of Dickey and Fuller (1979, 1981), which were based on the statistic $n(\hat{a} - 1)$ and t_a in the Gaussian AR(1) model. Phillips and Perron (1988), Ouliaris *et al.* (1989), and Park and Sung (1994) give various extensions of these semiparametric tests. Following Phillips (1987a), when the residual process u_t in (4) is a general stationary time series, the asymptotic distributions are given as follows:

$$n(\hat{a} - 1) \Rightarrow \left[\int_0^1 B_X(r)dB(r) + \lambda\right]\left[\int_0^1 B_X^2(r)dr\right]^{-1}, \tag{9}$$

and

$$t_{\hat{a}} \Rightarrow \sigma_u^{-1}\left[\int_0^1 B_X(r)dB(r) + \lambda\right]\left[\int_0^1 B_X^2(r)dr\right]^{-1/2}, \tag{10}$$

where $\sigma_u^2 = \mathrm{var}(u_t)$, $B(r)$ is Brownian motion with variance $\omega^2 = \sigma^2 C(1)^2$, $\lambda = \sum_{j=1}^{\infty} E(u_0 u_j)$ and $B_X(r)$ is detrended Brownian motion defined by the $L_2[0, 1]$

Hilbert space projection of $B(r)$ onto the space orthogonal to the span of $X(r)$, viz., $B_X(r) = B(r) - (\int_0^1 BX')(\int_0^1 XX')^{-1}X(r)$. In (9) and (10), ω^2 and λ are nuisance parameters and may be consistently estimated by nonparametric *kernel* techniques, analogous to those that are used in the estimation of the spectral density (e.g., see Andrews, 1991). Let $\hat{\omega}^2$ and $\hat{\lambda}$ be such estimates. Using the limit theory in (9) and (10) and these nonparametric estimates of the nuisance parameters, the following statistics are formed to test the unit root hypothesis:

$$Z_a = n(\hat{a} - 1) - \hat{\lambda}\left(n^{-2}\sum_{t=2}^{n} y_{X,t-1}^2\right)^{-1} \Rightarrow \left[\int_0^1 W_X\, dW\right]\left[\int_0^1 W_X^2\right]^{-1}, \tag{11}$$

$$Z_t = \hat{\sigma}_u \hat{\omega}^{-1} t_a - \hat{\lambda}\left\{\hat{\omega}\left(n^{-2}\sum_{t=2}^{n} y_{X,t-1}^2\right)^{1/2}\right\}^{-1} \Rightarrow \left[\int_0^1 W_X\, dW\right]\left[\int_0^1 W_X^2\right]^{-1/2}, \tag{12}$$

where $y_{X,t}$ is the residual from a regression of y_t on x_t. The limit variates shown in (11) and (12) involve standard Brownian motion $W(r) = (1/\omega)B(r)$ and the standardized process $W_X(r) = (1/\omega)B_X(r)$, so they are free of nuisance parameters and produce similar tests for a unit root.

The limit variates (11) and (12) simplify to those of the original Dickey-Fuller tests in the case of a fitted intercept or linear trend and can be used to construct critical values for the tests. This is typically done by large scale simulations, since the limit distributions are non-standard. Fuller (1976/1996) gives some numerical tabulations for the intercept and linear trend cases. Computerized tabulations are given in Ouliaris and Phillips (1994) for the case of polynomial trends. These limit distributions are asymmetric and have long left tails. In the case of the Z_a test, for instance, we reject the null hypothesis of a unit root at the 5% level if $Z_a < cv(Z_a; 5\%)$, the 5% critical value of the test. Both the Z_a and Z_t tests are one-sided. They measure the support in the data from a unit root against the alternative that the data are stationary about the deterministic trend x_t. When there is no deterministic trend in the regression model, the alternative hypothesis is just stationarity. In this case, the limit variates involve only the standard Brownian motion W, and $W_X = W$ in (11) and (12).

2.3. *The parametric ADF tests*

The most common parametric unit root test is the augmented Dickey-Fuller (ADF) test. This test was originally proposed by Dickey and Fuller (1979, 1981) for the case where u_t in (6) is an AR(p) process. The unit root hypothesis in (6) corresponds to the hypothesis $a = 0$ in the following regression:

$$\Delta y_t = a y_{t-1} + \sum_{j=1}^{k-1} \varphi_j \Delta y_{t-j} + \beta' x_t + \varepsilon_t. \tag{13}$$

This hypothesis can be tested by means of the regression coefficient $\hat{a}$ or its t-ratio statistic t_a, which have the same limiting distributions as those given in (7) and (8). For more general time series processes, we can expect that, as $k \rightarrow \infty$, the autoregressive approximation will give an increasingly accurate representation of the true process. In an important extension of Dickey and Fuller (1979), Said and Dickey (1984) prove the validity of the ADF t-ratio test (ADF_t) in general ARMA processes of unknown order, provided the lag length in the autoregression increases with the sample size at a rate less than $n^{1/3}$, where n = sample size. This statistic has the same limit distribution as the Z_t test given in (12) and thus the same critical values can be used in practical applications.

The limit distribution of the coefficient estimate $\hat{a}$ is dependent on nuisance parameters even as the lag length goes to infinity. Specifically,

$$n\hat{a} \Rightarrow \left[\sigma \int_0^1 W_X dW\right]\left[\omega \int_0^1 W_X^2\right]^{-1},$$

which depends on unknown parameters σ and ω. However, ω and σ can be consistently estimated. In particular, $\hat{\sigma}^2 = \sum \hat{\varepsilon}_t^2/n$ is a consistent estimator of σ^2, and ω^2 can be consistently estimated by the AR estimator (Berk, 1974) $\hat{\omega}^2 = \hat{\sigma}^2/(1 - \sum \hat{\varphi}_j)^2$. Under the null hypothesis that $a = 0$, it is apparent that the modified coefficient-based test statistic, $ADF_a = (\hat{\omega}/\hat{\sigma})n\hat{a}$, has the same limit distribution as that of the Z_a test and that of the original Dickey-Fuller coefficient test. This ADF_a test was developed in Xiao and Phillips (1997).

3. Towards efficient unit root tests

3.1. *The von Neumann ratio and LM tests*

The regression equations of classical unit root tests like (6) and (13) involve redundant trend variables. It is to be anticipated that elimination of redundant components in the deterministic trend may bring an efficiency gain to the unit root tests. One such test that successfully avoids the problem of redundant trend variables is the von Neumann (VN) ratio test.

The von Neumann (VN) ratio is the ratio of the sample variances of the differences and the levels of a time series. For Gaussian data this ratio leads to well known *tests of serial correlation* that have good finite sample properties. Sargan and Bhargava (1983) suggested the use of this statistic for testing the Gaussian random walk hypothesis, and Bhargava (1986) extends it to the case of a time trend. Using nonparametric estimates of the nuisance parameter ω^2, it is a simple matter to rescale the VN ratio to provide a unit root test for model (1) and (2) above. Stock (1995) does this for the case where there is a linear trend. Using a different approach and working with polynomial trends, Schmidt and Phillips (1992) show that for a Gaussian likelihood the Lagrange multiplier (LM) principle leads to a VN test, and can be generalized by using a nonparametric estimate of ω^2. The following discussion gives a generalized VN unit root test for the model (1) and (2), allowing for trends and trend breaks.

If y_t^s were observable, the VN ratio would take the form $VN = \sum_{t=2}^{n}(\Delta y_t^s)^2 / \sum_{t=1}^{n}(y_t^s)^2$. The process y_t^s is, in fact, unobserved but may be estimated from (1). Note that, under the null hypothesis and after differences are taken, we get

$$\Delta y_t = \Delta h_t + \Delta y_t^s. \tag{14}$$

This equation is trend stationary, so that by the Grenander–Rosenblatt theorem (Grenander & Rosenblatt, 1957, Chapter 7) the trend function can be efficiently estimated by an OLS regression. Doing so avoids the problem mentioned earlier of having surplus trend variables in the detrending regression. Intuition suggests that this should increase the power of the test, at least in the neighborhood of the null, and simulations in Schmidt & Phillips (1992) confirm this.

Let $\Delta\hat{y}_t^s = \Delta y_t - \Delta\hat{h}_t$ be the residuals from the efficient detrending regression (14) and let $\hat{y}_t^s = \sum_{j=2}^{t} \Delta\hat{y}_j^s$ be the associated estimate of y_t^s. Let $\hat{\omega}^2$ and $\hat{\sigma}_u^2$ be consistent estimates of ω^2 and σ_u^2. (These may be obtained in the same way as in the construction of the Z_t test, i.e. by using the residuals from the regression (6).) Finally, let $\tilde{y}_t^s = \hat{y}_t^s - \beta' x_t$ be the residuals from an OLS regression of $\hat{y}_t^s$ on x_t. Rescaling the von Neumann ratio then leads to the following two test statistics

$$R_{VN} = \frac{\hat{\omega}^2}{\hat{\sigma}^2} \frac{n^{-1}\sum_{t=2}^{n}(\Delta\hat{y}_t^s)^2}{n^{-2}\sum_{t=1}^{n}(\hat{y}_t^s)^2} \Rightarrow \left[\int_0^1 \hat{V}_X^2\right]^{-1}, \tag{15}$$

$$\tilde{R}_{VN} = \frac{\hat{\omega}^2}{\hat{\sigma}^2} \frac{n^{-1}\sum_{t=2}^{n}(\Delta\tilde{y}_t^s)^2}{n^{-2}\sum_{t=1}^{n}(\tilde{y}_t^s)^2} \Rightarrow \left[\int_0^1 \tilde{V}_X^2\right]^{-1}. \tag{16}$$

The limit process $\hat{V}_X(r)$ in (15) is a detrended generalized *Brownian bridge*, whose precise form depends on the deterministic trend h_t in (1). Specifically,

$$\hat{V}_X(r) = \check{V}_X(r) - \left(\int_0^1 \check{V}_X \check{X}'\right)\left(\int_0^1 \check{X}\check{X}'\right)^{-1} \check{X}(r) \tag{17}$$

is the projection residual of the process $\check{V}_X$ on the space spanned by $\check{X}(r)$, and

$$\check{V}_X(r) = W(r) - \left(\int_0^1 dW\check{X}'\right)\left(\int_0^1 \check{X}\check{X}'\right)^{-1} \int_0^r \check{X} \tag{18}$$

is a generalized Brownian bridge process. When $\check{X}$ has a constant element (as it usually will), it is easy to see that the process $\check{V}_X$ is tied down to the origin at the ends of the $[0, 1]$ interval just like a Brownian bridge, so that both $\check{V}_X(0) = 0$ and $\check{V}_X(1) = 0$. In the case of a simple linear trend, $\check{V}_X(r) = W(r) - rW(1)$ is a standard Brownian bridge, and $\hat{V}_X(r) = \check{V}_X(r) - \int_0^1 \check{V}_X$ is a detrended Brownian bridge. Consonant with the efficient detrending regression $\Delta\hat{y}_t^s = \Delta y_t - \Delta\hat{h}_t$, the limit process $\hat{V}_X$ in (15) is detrended using the limiting trend function $\check{X}(r)$, which, like $\check{x}_t$, involves no redundant trend variables.

Critical values of the limit variates shown in (15) and (16) must be obtained by simulation. The statistics are positive almost surely and the tests are one sided. Schmidt & Phillips (1992) provide tabulations for $\bar{R}_{VN}$ in the case where h_t is a linear trend. The presence of a unit root is rejected at the 5% level if $\bar{R}_{VN} > cv(\bar{R}_{VN}, 5\%)$.

3.2. *Quasi-difference detrended unit root tests and joint estimation of the local parameter and trend*

As discussed in Section 3.1 above, the von Neumann ratio test R_{VN} is constructed using an efficient detrending regression under the null hypothesis in contrast to the regression (6), where there are generally redundant trending regressors. One way to improve the power of unit root tests is to perform the detrending regression in a way that is efficient under the alternative hypothesis as well, an idea that was suggested in Elliot *et al.* (1996) in the context of the removal of means and linear trends. For alternatives that are distant from a unit root, this can be done directly by means of a regression on (1) because y_t^s is stationary with a spectral density that is continuous at the origin and then the Grenander–Rosenblatt theorem applies. To obtain large sample approximations, we can consider alternatives that are closer to unity. Such alternative hypotheses can often be well modelled using the local alternative

$$\alpha = \exp(n^{-1}c) \sim 1 + n^{-1}c \tag{19}$$

for some fixed $c = \bar{c}$, say, given the sample size n. In this case, in order to efficiently estimate the trend coefficient under the alternative hypothesis, we should use quasi-differencing rather than differencing in the construction of the detrending regression. It is known that such a regression leads to estimates of the trend coefficients that are asymptotically more efficient than an OLS regression in levels (Phillips and Lee, 1996), and this result justifies the modified test procedure that follows.

Define the quasi-difference operator as $\Delta_{\bar{c}}$, $\Delta_{\bar{c}} y_t = (1 - L - n^{-1}\bar{c}L)y_t = \Delta y_t - n^{-1}\bar{c}y_{t-1}$, take quasi-differences of (1) and run the detrending regression

$$\Delta_{\bar{c}} y_t = \tilde{\gamma}' \Delta_{\bar{c}} x_t + \Delta_{\bar{c}} \tilde{y}_t^s. \tag{20}$$

We call such detrending procedures quasi-difference (QD) detrending. Using the fitted coefficients $\tilde{\gamma}$ from this OLS regression, the levels data are detrended according to

$$\tilde{y}_t = y_t - \tilde{\gamma}' x_t. \tag{21}$$

The detrended data $\tilde{y}_t$ may be used in the construction of unit root tests. For example, we can construct the modified semi-parametric Z_α test by running the regression of the QD detrended variable $\tilde{y}_t$ on its one-period lagged value $\tilde{y}_{t-1}$ without deterministic trends in the regression, giving

$$\tilde{y}_t = \tilde{\alpha}\tilde{y}_{t-1} + \text{residual}.$$

The modified Z_α test statistic has the following form

$$\tilde{Z}_\alpha = n(\tilde{\alpha} - 1) - \bar{\lambda}\left(n^{-1}\sum_{t=2}^{n} \tilde{y}_{t-1}^2\right)^{-1}$$
$$\Rightarrow \left\{\left[\int_0^1 dW\tilde{W}_{\bar{c}}\right] - \left[\int_0^1 dW_{\bar{c}}X'_{\bar{c}}\right]^{-1}\left[\int_0^1 X_0\tilde{W}_{\bar{c}}\right]\right\}\left[\int_0^1 \tilde{W}_{\bar{c}}^2\right]^{-1} \tag{22}$$
$$= \left[\int_0^1 \tilde{W}_{\bar{c}}^2\right]^{-1}\int_0^1 \tilde{W}_{\bar{c}}d\tilde{W}_{\bar{c}},$$

where $\tilde{W}_{\bar{c}}(r) = W(r) - \int_0^1 dW_{\bar{c}}X'_{\bar{c}}(\int_0^1 X_{\bar{c}}X'_{\bar{c}})^{-1}X(r)$ is the weak limit of $n^{-1/2}\tilde{y}_{[nr]}$, $W_{\bar{c}}(r) = W(r) - \bar{c}\int_0^r W(s)$, $\bar{\lambda}$ is a consistent estimator of λ, $X_{\bar{c}}(r) = X'(r) - \bar{c}X(r)$ is the limit function of the quasi-differenced deterministic trend and $X_0 = X'(r)$ is the limiting deterministic trend function with $\bar{c} = 0$. Note that the simple form of (22) follows because

$$d\tilde{W}_{\bar{c}}(r) = dW(r) - \int_0^1 dW_{\bar{c}}X'_{\bar{c}}\left(\int_0^1 X_{\bar{c}}X'_{\bar{c}}\right)^{-1} X_0(r)dr, \tag{23}$$

since $dX(r) = X'(r)dr = X_0(r)dr$. From expression (23) it is further apparent that $(d\tilde{W}_{\bar{c}}(r))^2 = (dW(r))^2 = dr$. While similar in form to (11), the limit formula (22) depends on the process $\tilde{W}_{\bar{c}}(r)$, rather than $W_X(r)$. One difference here is the dependence of $\tilde{W}_{\bar{c}}(r)$ on $\bar{c}$. A second and more significant difference is that $\tilde{W}_{\bar{c}}(r)$ is formed by taking a non orthogonal Hilbert projection residual of $W(r)$ on the space spanned by $X(r)$ in $L_2[0,1]$. Such projections appear infrequently in Hilbert space analysis, but this is an important example where they do appear, arising from the idempotent operator that gives the optimal direction of the projection in $L_2[0,1]$ function space — see Phillips (1996a) for more discussion.

Using the same idea, we can construct the modified Z_t tests and ADF tests and the corresponding limit theory for these tests is

$$\widetilde{ADF}_t, \tilde{Z}_t \Rightarrow \left[\int_0^1 \tilde{W}_{\bar{c}}^2\right]^{-1/2}\int_0^1 \tilde{W}_{\bar{c}}d\tilde{W}_{\bar{c}}. \tag{24}$$

By a simple application of stochastic calculus using the fact that $(d\tilde{W}_{\bar{c}}(r))^2 = (dW(r))^2 = dr$, it is apparent that the limit distribution (24) can be written in the alternate form

$$\frac{1}{2}\left[\int_0^1 \tilde{W}_{\bar{c}}^2\right]^{-1}[\tilde{W}_{\bar{c}}(1)^2 - 1].$$

In the special case where x_t is a constant or a linear trend, these formulae reduce to those given in Stock (1994).

Since the limit theory in (22) is different from that of (11), new critical values are needed for this test. The limit theory depends explicitly on the trend functions, as it does in (11), but it also depends on the posited value of the localizing parameter $\bar{c}$ that is used in the quasi-differencing. A reasonable default choice of $\bar{c}$ seems to be the value for which local asymptotic power (see Section 3.4 below) is 50% (see King, 1988, and Elliot *et al.*, 1996).

The QD detrending procedure involves the choice of the prespecified local parameter c. There is another way to proceed that does not appear in the literature to date which we will now exposit. If we incorporate the local to unity hypothesis (19) in models (1) and (2), we obtain the following nonlinear regression:

$$\Delta y_t = \beta'\Delta x_t - x\beta'\left(\frac{x_{t-1}}{n}\right) + c\left(\frac{y_{t-1}}{n}\right) + \varepsilon_t. \tag{25}$$

This nonlinear regression provides for joint estimation of the local parameter c and the trend coefficient β. If we denote the nonlinear estimate of β and c in the above regression by $\tilde{\beta}_{nl}$, and $\tilde{c}_{nl}$, denote the limit of $n^{-1/2}D_n(\tilde{\beta}_{nl} - \beta)$ by ξ and the limit of $\tilde{c}_{nl}$ by η, then the asymptotic behavior of these quantities is governed by the following equations:

$$\xi = [\int X_\eta(r)X_\eta(r)'dr]^{-1}\,[\int X_\eta(r)dB(r) + (c-\eta)\int X_\eta(r)J_c(r)dr],$$
$$\eta = c + [\int \underline{J}_{c\xi}(r)^2 dr]^{-1}\,[\int \underline{J}_{c\xi}(r)dB(r) - \xi'\int \underline{J}_{c\xi}(r)X_c(r)dr],$$

where

$$X_\eta(r) = g(r) - \eta X(r),$$
$$X_c(r) = g(r) - cX(r),$$
$$\underline{J}_{c\xi}(r) = J_c(r) - \xi' X(r),$$

and $D_n^{-1}x_{[nr]} \to X(r)$, $nD_n^{-1}\Delta x_{[nr]} \to g(r)$, $J_c(r) = \int_0^r e^{(r-s)c}dB(s)$. There is no direct analytic solution to these equations, but the equations determine (ξ, η) and thereby the limit distributions of $\tilde{\beta}_{nl}$ and $\tilde{c}_{nl}$, which can be found by numerical methods.

Although the local parameter c can not be consistently estimated because of the asymptotic collinearity between y_{t-1} and x_{t-1} in the above nonlinear regression, this regression will still provide a more efficient estimate of the deterministic trend (i.e., the parameter β) than regressions which do not take into account the parameter restrictions in (25). Actually, when the u_t are $N(0, \sigma^2)$ variates, this nonlinear regression delivers the maximum likelihood estimates of β and c for model (1), (2) and (19). The improvement in efficiency from this joint estimation procedure has been confirmed by the authors in Monte Carlo experiments, which we do not report here due to space constraints. Because we do not know the true value of the local parameter c, this maximum likelihood estimate of the deterministic trend can not achieve the efficiency level that applies when the local parameter is known. Nevertheless, the approach certainly seems worthy of use.

3.3. *A point optimal test*

In the simplest framework where the model is a Gaussian AR(1) with unit error variance, the Neyman-Pearson lemma can be used to construct the most powerful test of a unit root against a simple point alternative. Such a test is point optimal for a unit root at the specific point alternative that is selected. King (1988) provides a general discussion of such point optimal invariant tests, and Dufour and King (1991) developed the family of exact most powerful invariant tests. Elliot *et al.* (1996) apply this idea in the context of unit root tests by using the local alternative (19) for a particular value of $c = \bar{c}$.

If we assume $u_t \equiv \mathrm{iid}N(0, \sigma^2)$ in (2), $y_0^s = 0$, and σ^2 is known, then the likelihood function for the autoregression coefficient α is proportional to

$$L(\alpha) \sim -\tfrac{1}{2}\sigma^{-2}\sum_t (y_t^s - \alpha y_{t-1}^s)^2,$$

where $y_t^s = y_t - h_t$. If h_t were known, then the likelihood function could be calculated and a most powerful test could be constructed directly by the Neyman–Pearson lemma. However, as discussed in Section 3.1, h_t is not known and y_t^s has to be estimated. Moreover, u_t may be a general $I(0)$ process and thus the limit distribution of the point optimal test statistic depends in general on nuisance parameters. In this case, corrections have to be made on the original *LR* test so that the adjusted test statistic is free of nuisance parameters. As Dufour and King (1991) and Elliot *et al.* (1996) discuss, the point optimal invariant test (POI) statistics can be constructed based on the ratio of the sum of squared residuals from the efficient detrending regressions under the null and alternative hypothesis. Specifically, taking a local alternative $\alpha = 1 + n^{-1}c$ with $c = \bar{c}$, using quasi-differencing to detrend, and using a consistent nonparametric estimate $\hat{\omega}^2$ of the nuisance parameter ω^2, the POI test statistic for a unit root in (1) and (2) has the following form:

$$\tilde{P}_{\bar{c}} = \hat{\omega}^{-2}\left\{\bar{c}^2 n^{-2}\sum_{t=2}^{n}(\tilde{y}_{t-1}^s)^2 - \bar{c}n^{-1}\tilde{y}_n^s\right\} \Rightarrow \bar{c}^2\int_0^1 \tilde{W}_{\bar{c}}^2 - \bar{c}\tilde{W}_{\bar{c}}(1), \tag{26}$$

where the notation is the same as that defined above in Section 3.2. The test is performed by comparing the observed value of the statistic with the critical value obtained by simulation. The presence of a unit root in the data is rejected if the calculated value of the statistic $\tilde{P}_{\bar{c}}$ is too small. Note that in the construction of $\tilde{P}_{\bar{c}}$, the estimate $\hat{\omega}^2$ is used and this is obtained in the same way as in the Z_t test, i.e., using residuals from the regression (6). This point is of some importance and affects the consistency of the test — see Section 3.4 below.

3.4. *Asymptotic properties and local power*

All of the above test statistics are asymptotically *similar* in the sense that their limit distributions are free of nuisance parameters. However, the limit distributions do depend on whether the data has been prefiltered in any way by preliminary regression. Thus, if deterministic trends are removed by regression as in (6) or (20), then the limit distributions of the unit root test statistics depend on limiting versions of the deterministic trends that are used in the detrending regressions.

The tests are also consistent against stationary alternatives provided that any nonparametric estimator of ω^2 that is used in the test converges in probability under the alternative to a positive limit as $n \to \infty$. The latter condition is important, and it typically fails when estimates of ω^2 are constructed using first differences or quasi-differences of the data rather than regression residuals. This is because, under the alternative hypothesis, the data are stationary and first differences (or quasi-differences) of stationary data have zero spectrum at the origin. (See Phillips and Ouliaris, 1990, for further discussion of this issue.) The point is especially important when 'detrending after quasi-differencing' is used, as outlined in Section 3.2 above, as in this case there is a natural tendency to estimate ω^2 from the residuals of this regression. In effect, while an efficient regression in quasi-differences may be run to detrend the data, an inefficient regression such as (6), where the autoregressive coefficient is estimated, must be run to estimate the long-run variance parameter ω^2. Thus, some care is needed in the formulation of tests that rely on nonparametric estimates of ω^2. The problem also arises in certain parametric unit root tests when nuisance parameters are estimated using first differenced data (as in the case of Solo's, 1984, LM test — see Saikonnen and Luukkonen, 1993).

Rates of divergence of the statistics under the alternative are also available. For instance, when $|\alpha| < 1$, Z_α, $\tilde{Z}_\alpha$, ADF_α, $VN = O_p(n)$, and Z_t, $ADF_t = O_p(n^{1/2})$ as $n \to \infty$. Thus, coefficient-based tests that rely on the estimated autoregressive coefficient and the von Neumann ratio/LM tests diverge at a faster rate than tests that are based on the regression t-ratio. We may therefore expect such tests to have greater power than t-ratio tests, and this is generally borne out in simulations. Heuristically, the t-ratio tests suffer because there is no need to estimate a scale parameter when estimating the autoregressive coefficient α. The autoregressive estimators $\hat{\alpha}$, and $\tilde{\alpha}$, on the other hand, are already scale invariant. Note also that while $ADF_t^2 = O_p(n)$ and therefore has the same divergence characteristics as Z_α, $\tilde{Z}_\alpha$, ADF_α, and VN, the statistic ADF_t^2 produces a nondirectional test, whereas the coefficient-based tests are directional (against stationarity, or explosive behavior).

Under the local alternative hypothesis (19), the limit theory for the above statistics can be derived and used to analyze local asymptotic power. When (2) and (19) hold, y_t^s behaves asymptotically like a linear diffusion, i.e., $n^{-1/2} y^s_{[nr]} \Rightarrow J_c(r) = \int_0^r e^{(r-s)c} dW(s)$ (see Phillips, 1987b). The limit distributions of the unit root test statistics then involve functionals of $J_c(r)$. For example, the Z_α statistic has the limit

$$Z_\alpha \Rightarrow c + \left[\int_0^1 J_{cX} dW\right]\left[\int_0^1 J_{cX}^2\right]^{-1}, \tag{27}$$

where $J_{cX}(r) = J_c(r) - (\int_0^1 J_c X')(\int_0^1 XX')^{-1}X(r)$. This limit is identical to that of the Dickey-Fuller test that is based directly on the coefficient estimator $n(\hat{\alpha} - 1)$ when u_t has no serial correlation and no corrections are required to make the statistic asymptotically similar. Thus, the corrections for residual serial correlation in the statistic Z_α do not lead to any loss in asymptotic power.

The local asymptotic theory can be used to construct asymptotic power envelopes for unit root tests. Under the hypothesis that the data is Gaussian, the best test of a unit root against the specific local alternative with $c = \bar{c}$ is given by the point optimal test by virtue of the Neyman-Pearson lemma. When efficient detrending under this alternative is used, the resulting test statistic is $\bar{P}_{\bar{c}}$ as given in (26). The limit distribution of this statistic under the specific local alternative $c = \bar{c}$ is

$$\bar{P}_{\bar{c}} \Rightarrow \bar{c}^2 \int_0^1 \tilde{J}_{\bar{c}}^2 - \bar{c}\tilde{J}_{\bar{c}}(1), \tag{28}$$

where $\tilde{J}_{\bar{c}}(r) = J_{\bar{c}}(r) - X(r)(\int_0^1 X_{\bar{c}}X_{\bar{c}}')^{-1}(\int_0^1 X_{\bar{c}}dW - \bar{c}\int_0^1 X_{\bar{c}}J_{\bar{c}})$. As we vary the parameter $\bar{c}$, this distribution delivers a power envelope against which other tests may be compared. Note that the limit given in (28) is attainable by the POI test (26) only when $c = \bar{c}$ exactly. In general, the chosen value of c that is used in the QD detrending procedure on which (26) is based will be different, and hence the power of the POI test will generally be less than that delivered by (28). In the special case of a linear tend, computations in Stock (1995) indicate that the POI test (26) has power that is very close to the power envelope for a wide range of local alternatives.

3.5. *Further issues on size and power*

Two further issues relating to the size and power of unit root tests deserve attention. The first of these arises from some recent work of Phillips (1998a, b) showing that nonstationary time series admit many different representations. The most obvious representation comes directly from the model formulation (2) itself. However, Phillips (1998a) shows that there are valid alternative representations in terms of deterministic functions. These representations originate in the corresponding representation of the limiting Brownian motion for which $n^{-1/2}y^s_{[nr]} \Rightarrow B(r)$. Indeed, for $B(r)$ we have the following L_2-representation

$$W(r) = \omega\sqrt{2}\sum_{k=1}^{\infty} \frac{\sin[(k-1/2)\pi r]}{(k-1/2)\pi}\xi_k = \omega\sum_{k=1}^{\infty}\varphi_k(r)\xi_k, \tag{29}$$

where the components ξ_k are independently and identically distributed (iid) as $N(0,1)$ and the functions $\varphi_k(r)$ form an orthogonal set in $L_2[0,1]$. Phillips shows that empirical regressions of y_t^s on $\varphi_{K,t} = (\varphi_1(t/T), \ldots, \varphi_K(t/T))'$ accurately reproduce in the limit the first K terms of the expansion (29). Further, when $K \rightarrow \infty$ and $K/T \rightarrow 0$ as $T \rightarrow \infty$, such regressions succeed in reproducing the entire representation (29). It follows that these deterministic functions are capable of successfully representing a nonstationary time series like y_t^s in the limit as $T \rightarrow \infty$.

Such regressions on deterministic functions then become an alternate way of modelling a nonstationary time series. A fascinating implication of this work is that unit root tests which involve deterministic functions, like (1) above, will inevitably lead to the rejection of the unit root hypothesis when $K, T \longrightarrow \infty$ and in this sense the conventional critical values used in unit root tests (like those based on the limit functionals (11) and (12)) are invalid asymptotically when the competing deterministic functions that appear in the maintained hypothesis provide an alternative mechanism of modelling the non-stationarity, as in cases like polynomial trends and trend break polynomials they will. These issues are further explored in ongoing research in Phillips (1998b). It is too early to comment on the full implications of this work, but the results are of obvious importance in the empirical assessment of trend/trend-break stationarity versus persistence in economic time series.

The second issue emerges from some recent work by Faust (1996). Faust pointed out that while unit root tests like the Z_a, Z_t, and ADF_t tests have limit distributions under the null that are invariant to the nuisance parameters σ^2, ω^2 and λ, the non-parametric nature of the maintained hypothesis can cause difficulties. In particular, the key condition that underlies the unit root null hypothesis in (2) is that $C(1) \neq 0$. In fact, using the BN expansion of the operator $C(L)$, we have, as in (5) above

$$u_t = C(1)\varepsilon_t + \tilde{\varepsilon}_{t-1} - \tilde{\varepsilon}_t, \quad \text{with } \tilde{\varepsilon}_t = \tilde{C}(L)\varepsilon_t. \tag{30}$$

It follows that if $C(1) = 0$, then $y_t^s = \tilde{\varepsilon}_0 - \tilde{\varepsilon}_t + y_0^s$, which is stationary. Of course, in the $I(1)$ class for y_t^s there will be some error processes u_t satisfying (3) for which $C(1)$ is arbitrarily close to zero. Indeed, Faust shows for any real a, the set of sequences

$$\mathscr{C}_a = \{c = (c_j)_1^\infty \colon c_j = 0 \text{ for all but a finite number of } j;\ C(1) = a\}$$

is dense in l_2, the space of square summable sequences. It follows that given a sequence $c = (c_j)_1^\infty$ of coefficients in the Wold representation of u_t for which $C(1) = a \neq 0$, there is a sequence $c' = (c_j')_1^\infty$ that is arbitrarily close to c in l_2 but for which $C'(1) = \sum_{j=0}^\infty c_j' = 0$. Faust concludes that the $I(1)$ sequences, for which $C(1) \neq 0$, and the $I(0)$ sequences, for which $C(1) = 0$, are both dense in l_2, and hence these classes of processes are nearly observationally equivalent. (Campbell and Perron, 1991, and Blough, 1992, also indicated that this property may affect unit root tests.) One implication of this near observational equivalence, is that the size of semi-parametric unit root tests will not converge to the nominal size given by the limit distribution, at least when the size is computed by taking the supremum of the rejection probability of the unit root test over the set $\mathscr{G}_\eta = \{c \in \mathscr{C}_a \colon |C(1)| > \eta > 0\}$ for any η. (In fact, Faust shows that the actual size of the test converges to unity). In view of (30), the restriction $\eta > 0$ would normally be interpreted as setting up a buffer zone between the class of $I(1)$ and $I(0)$ processes. Faust's result shows that even this buffer zone does not prevent size distortions in a general enough nonparametric setting for the error process.

The reason for the size distortion is that the nonparametric (composite) form of the null hypothesis is too broad as it stands. As is apparent from (30), the error processes that lead to the size distortion involve sequences like $c' = (c_j')_1^\infty$ that are arbitrarily close in l_2 to a null sequence c (for which $C(1) \neq 0$) but which have $C'(1) = \sum_{j=0}^\infty c_j' = 0$. However, for these sequences c' to produce data y_t^s that have $I(1)$-like properties, the second component $\tilde{\varepsilon}_t' = \tilde{C}'(L)\varepsilon_t$ in the BN decomposition of

$$u_t' = C'(L)\varepsilon_t = C'(1)\varepsilon_t + \tilde{\varepsilon}_{t-1}' - \tilde{\varepsilon}_t'$$

must also have $I(1)$-like properties. It is, in fact, quite easy to exclude this possibility by placing a smoothness requirement on spectrum of $\tilde{\varepsilon}_t'$. This can be accomplished by a summability condition on the allowable sequences $\{\tilde{c}_j'\}$, which is in turn assured by a summability condition on the original sequence c'. It turns out that a strengthening of the summability condition used in Phillips and Solo (1992) to validate the BN decomposition is sufficient to rule out the pathology of $I(0)$ sequences with near $I(1)$ behavior.

4. Finite sample properties of unit root tests

Extensive simulations have been conducted to explore the finite sample performance of unit root tests (inter alia, Schwert, 1989; Diebold and Rudebusch, 1991; DeJong *et al.*, 1992; Phillips and Perron, 1988; Ng and Perron, 1995; and Stock, 1995). One general conclusion to emerge is that, although differences exist across tests and these depend on the models generating the data, the discriminatory power in all of the tests between models with a root at unity and a root close to unity is generally low. For instance, power is usually less than 30% for $\alpha \in [0.90, 1.0)$ and $n = 100$. Power is reduced further by detrending the data — even larger values of the test statistics are required to achieve a rejection of the null and the power curve is lower. Both these features mirror the asymptotic theory. However, as the discussion in the previous section indicates, some of the observed power reduction is spurious, because the critical values of the tests are inappropriate when extensive deterministic detrending is done prior to testing for unit roots.

Another interesting finding from simulation studies is the extent of the finite sample size distortion (the difference between the nominal asymptotic size of the test and the actual finite sample size) of the tests in cases where the true model is close to a trend stationary process (Schwert 1989). For example, if u_t in (2) follows a moving average process $u_t = \varepsilon_t + \theta\varepsilon_{t-1}$ whose parameter is large and negative, then the sample trajectories of y_t^s more closely resemble those of a stationary process than a random walk. In such cases there is a tendency for all of the tests to over-reject the null of a unit root. This is an outcome that may not be so serious in practical work if the data are indeed better modeled by a trend stationary process, and so it is easy to overstate the importance of size distortions in such cases.

Tests that are based directly on autoregressive coefficient estimates like the Z_α tend to be more affected by size distortion than the other tests because the bias in the first order autoregressive estimator is large in this case, not only in finite samples but even in the asymptotic distribution (9), where the miscentering is measured by the bias parameter $\lambda = \theta\sigma_\varepsilon^2$. This is large when θ is large, and good estimates of the bias parameter are needed to control the size distortion. The one sided covariance parameter λ is usually estimated in a nonparametric way by kernel methods which often give confidence intervals with low coverage probabilities, especially when the time series has substantial temporal dependence. The parameter is also estimated at a slower rate than $\sqrt{n}$, and it is therefore often difficult to estimate well with samples of the size that are typical in many econometric applications (with $n \leqslant 200$). Similar comments apply to the estimation of the long run variance parameter ω^2, which appears in the other semi-parametric tests.

Recent attempts to improve the estimation of this parameter using data-determined bandwidth choices (Andrews, 1991) coupled with prewhitening (Andrews & Monahan, 1992), and data-based model selection and prewhitening (Lee & Phillips, 1994) offer some promise in this direction, as does pretesting for lag length in ADF regressions (Ng and Perron 1998). In particular, prewhitening is shown to bring better accuracy and less variance to kernel estimators. The idea behind prewhitening is to transform the data to reduce temporal dependence before applying kernel estimation. The transformed data typically have a flatter spectrum which can be estimated with less bias than the original spectrum. The kernel density estimator for the original data can then be obtained by applying the inverse transformation. Andrews and Monahan (1992) introduced a class of VAR prewhitened kernel estimators. In the scalar case, Lee and Phillips (1994) extend this idea by employing model selection techniques in the prewhitening stage and implement the Hannan-Rissanen recursion to efficiently estimate an ARMA model prefilter prior to estimating the long run variance ω^2 by kernel techniques. It is shown that, with this method, $\sqrt{n}$-rates of estimation are achievable with nonparametric estimates when consistent *model selection* techniques are used to determine the prefilter and the model for the errors lies within the prefiltering class (in this case the class of finite parameter ARMA models).

The parametric ADF t-ratio test is less affected by size distortions when the true model is close to stationarity, but generally has less power than the other tests. As shown by simulation experiments, the coefficient-based tests and VN ratio tests typically have better power properties than the ADF t-ratio test. Although on theoretic grounds it is known that the lag length of the ADF regression can grow at a rate $o(n^{1/3})$, not much information is provided in this rate criterion about lag length selection for specific sample sizes. With these tests, power is further reduced by the inclusion of additional lagged dependent regressors in (13). It has been found in many Monte Carlo studies that lag length selection has important effects on the finite sample performance of ADF tests. DeJong *et al.* (1992) show in their simulation results that increasing the lag length typically lowers the power in a systematic way, although it may also reduce size distortion. Again, the use of model selection methods like BIC (Schwarz, 1978; Rissanen, 1978) are useful in this respect and provide some improvement in the finite sample performance of the ADF tests. Ng and Perron (1995, 1998) studied the choice of lag length in constructing the ADF t-test and compared information-based model selection rules, such as BIC and AIC (Akaike, 1977), with classical sequential tests in determining lag length, such as F- and t-tests for the significance of the lag coefficients. They show that data-dependent rules which take sample information into account have beneficial effects on the finite sample performance of unit root tests.

Since detrending the data reduces power, it is to be expected that the inclusion of surplus trend variables in regressions like (6) will do so also. Hence, efficient detrending procedures like those discussed in Section 3.2 can be expected to benefit all tests, and this is partly confirmed by simulations in Stock (1995). Of all the procedures studied so far, efficient detrending by regression in quasi-differences seems to be the most successful in increasing finite sample (and asymptotic) power.

We provide some Monte Carlo results here, partially illustrating the findings in existing simulation studies, with an emphasis on studying the effect of the procedures mentioned above on the finite sample performance of common unit root tests. In particular, we examine the effect of model selection procedures, bandwidth selection methods, prewhitening and detrending procedures on the finite sample power of the following unit root tests: ADF_α, ADF_t, Z_α, Z_t tests combined with various detrending procedures; and VN and POI tests. For the ADF tests, the BIC criterion of Schwarz (1978) and Rissanen (1978) is used in selecting the appropriate lag length of the autoregression and the AR spectral estimator of Berk (1974) is used for the estimation of the long run variance parameter. Thus, in this Monte Carlo experiment, the ADF tests are all parametric. For the Z tests, the Andrews and Monahan prewhitened kernel estimation of the long run variance parameter is used. Although comparison has been made for different kernel choices, only those results using the Parzen kernel function are reported because no unambiguous ranking could be found among different kernels. Size-corrected power is reported in the simulations to provide a comparison among the different tests, although it does not reflect empirical rejection frequencies based on the use of asymptotic critical values. The finite sample critical values are calculated as quantiles in the simulations under the null hypothesis of a unit root, given the model selection rules and kernel choices.

The simulation results suggest some general findings. First, using data-based bandwidth choice coupled with prewhitening procedures in the estimation of nuisance parameters significantly improves the finite sample performance of the Z tests. Second, the use of model selection procedures like BIC in choosing lag length helps to improve the ADF tests. Third, unit root tests based on QD detrending have reasonably good finite sample properties, especially in the case where the deterministic trend includes a constant term.

Table A reports the size-corrected power of ADF and Z tests for the case without deterministic trends. Four designs for the data generating process are considered here. In each case, $y_t = y_t^s$, $y_t^s = \alpha y_{t-1}^s + u_t$, and initial values are set to be 0. The four different error structures are: an AR(1) process $u_t = \rho u_{t-1} + \varepsilon_t$, with $\rho = 0.5, -0.5$ and an MA(1) process $u_t = \varepsilon_t - \theta\varepsilon_{t-1}$, with $\theta = 0.5, -0.8$, where ε_t are iid standard normal variates. For the case with a deterministic trend, the size corrected power properties are reported in Figures 1 to 10. Figure 1 depicts the power of four (OLS detrended) unit root tests (Z_α, Z_t, ADF_α, ADF_t) when the error process is AR(1) with $\rho = 0.5$, and Figure 2 reports the results for these tests when u_t is MA(1) process with $\theta = 0.5$. Figures 3 and 4 compare different tests based on the same detrending procedures, and Figures 5 to 8 compare the effects of different detrending procedures on the same tests. All these experiments study the case where u_t is an iid standard normal process. Specifically, power comparisons among five tests (Z_α, Z_t, ADF_α, ADF_t and VN) are given in Figure 3. The power envelope is also provided in the graph for convenience of comparison. The power of the QD detrended versions of these tests are given in Figure 4. Figure 5 compares the power of the OLS detrended Z_α test with those of the QD detrended Z_α tests for different choices of the prespecified local parameter $\bar{c}$, and Figures 6, 7, and 8 compare the power of Z_t, ADF_α, and ADF_t tests respectively for different choices of detrending procedures. Figures 9 and 10 give the power of these tests for another form of deterministic trend.

DeJong *et al.* (1992) find from their Monte Carlo study that the semiparametric Z tests have very low power when there is positive serial correlation, while the ADF_t test is reasonably well-behaved in this case. Their results were obtained based on commonly used estimators of the nuisance parameters without

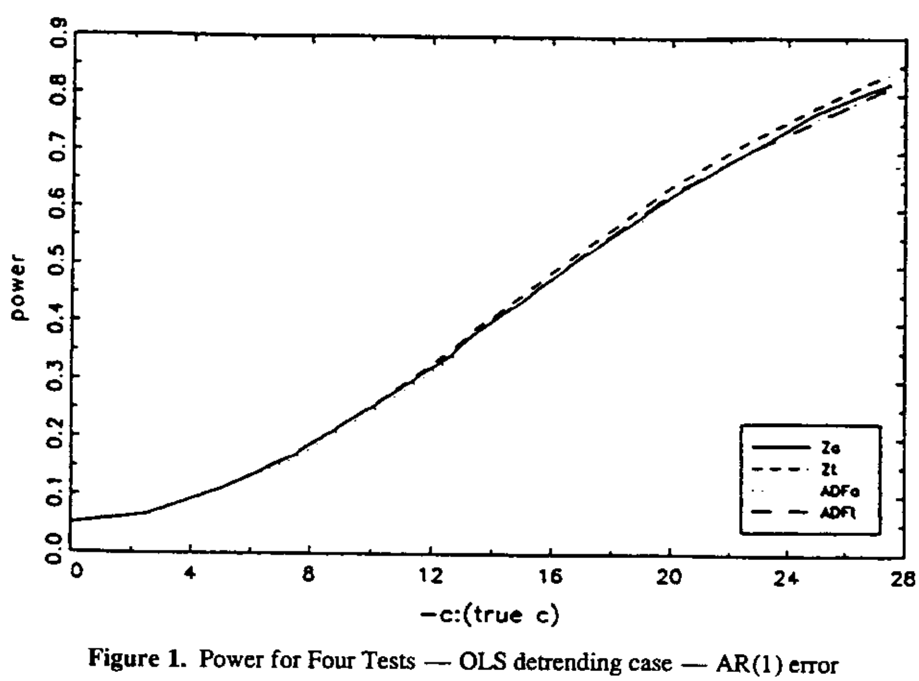

Figure 1. Power for Four Tests — OLS detrending case — AR(1) error

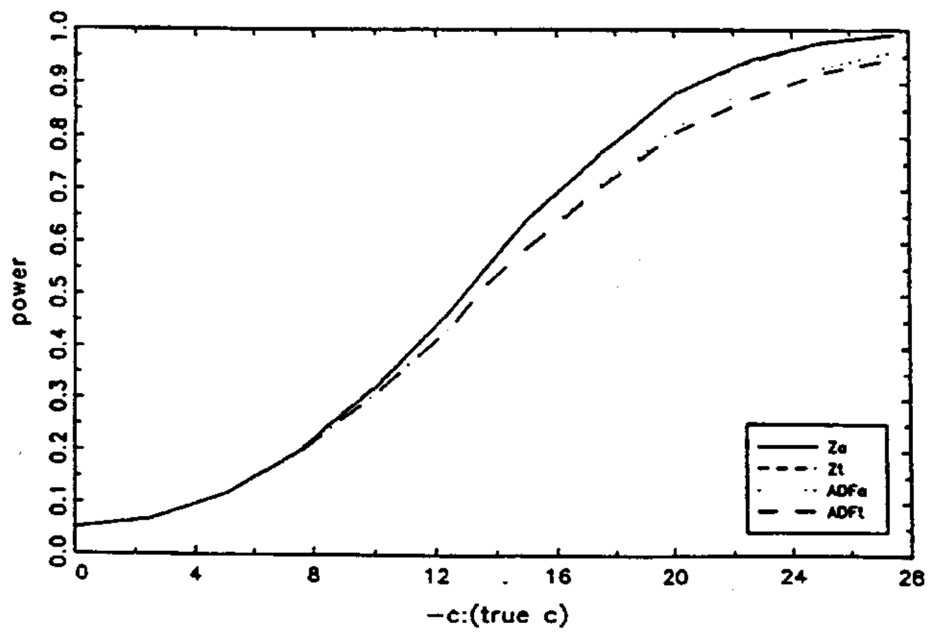

Figure 2. Power for Four Tests — OLS detrending case — MA(1) error

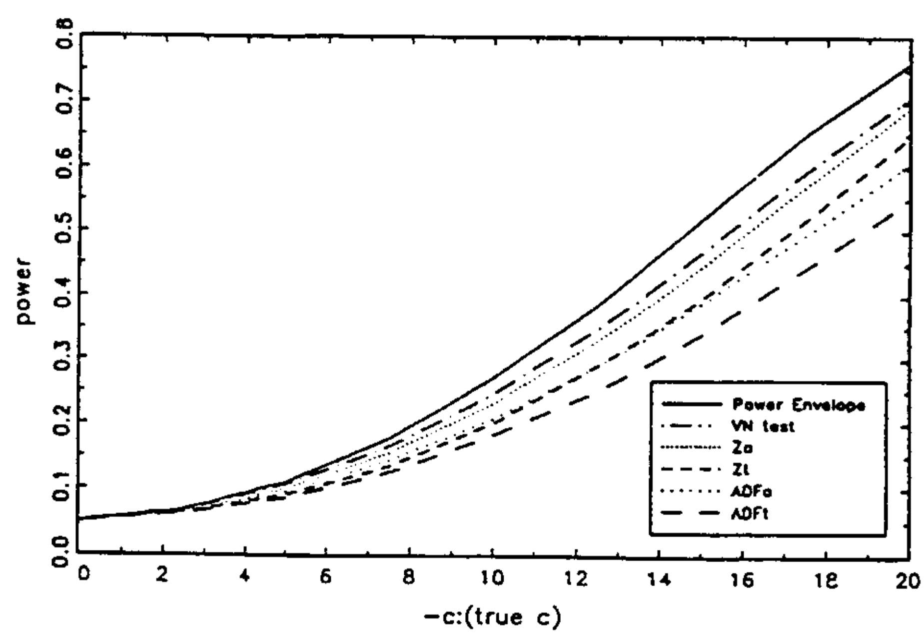

Figure 3. Power for Five Tests — OLS detrending case

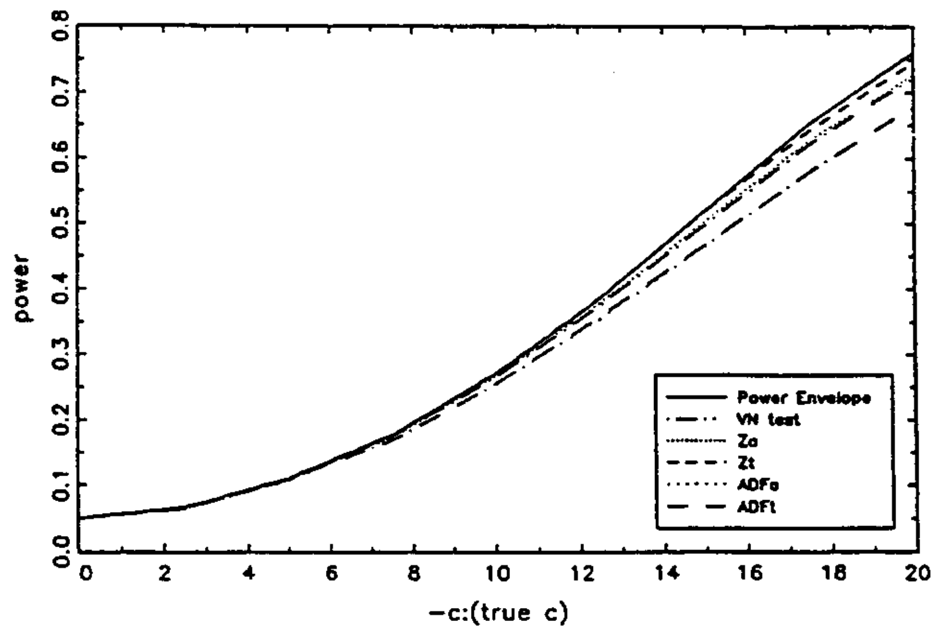

Figure 4. Power for Five Tests — GLS detrending case (c = −10)

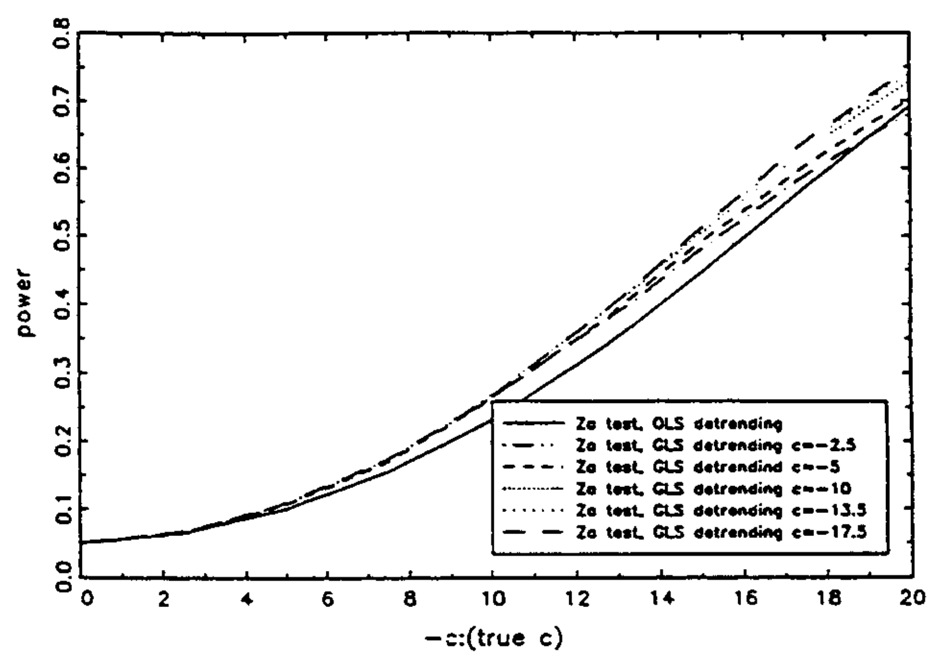

Figure 5. Power of Za tests based on different detrending procedures

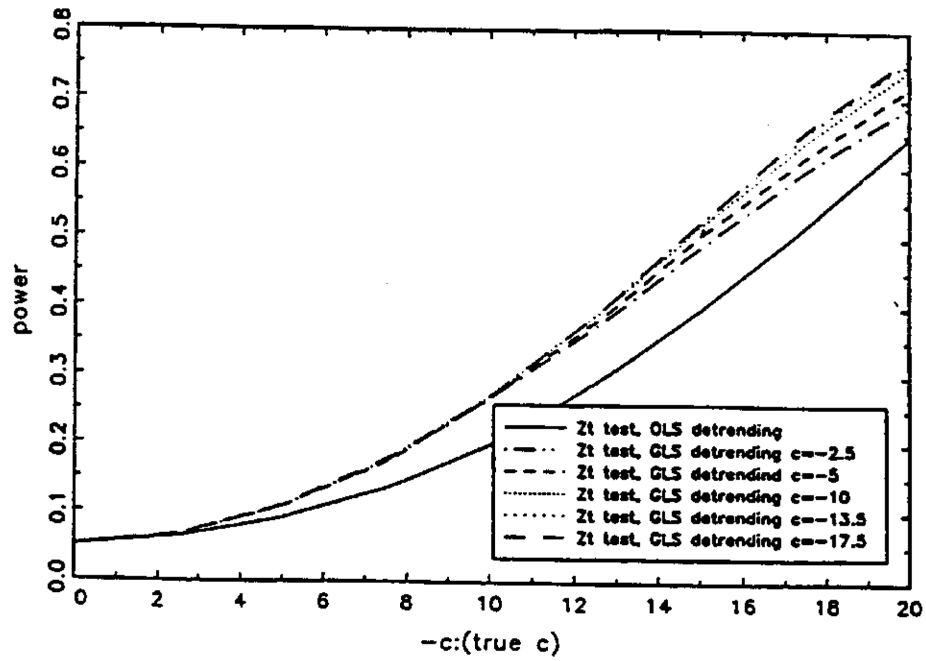

Figure 6. Power of Zt tests based on different detrending procedures

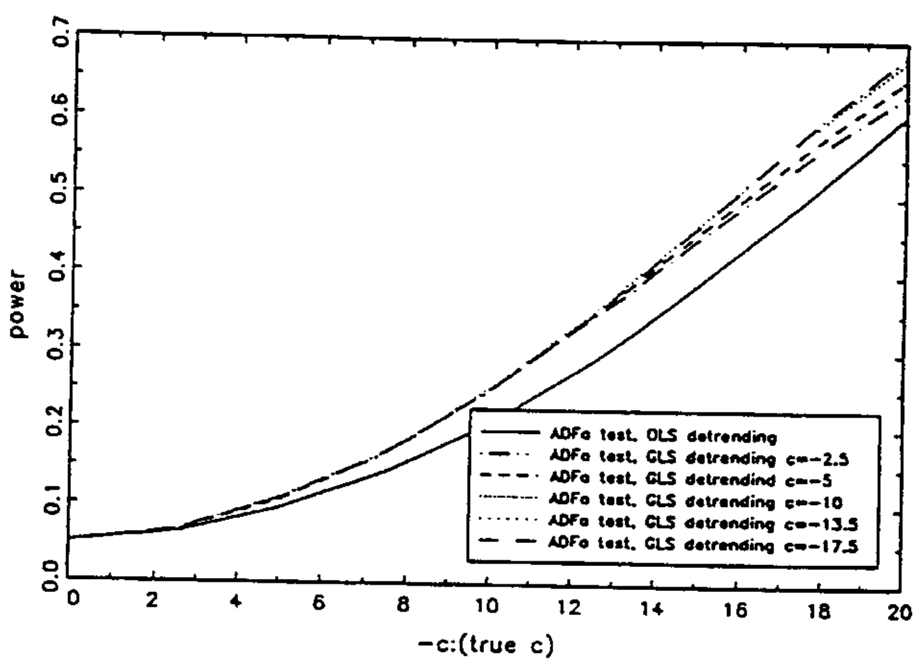

Figure 7. Power of ADFa tests based on different detrending procedures

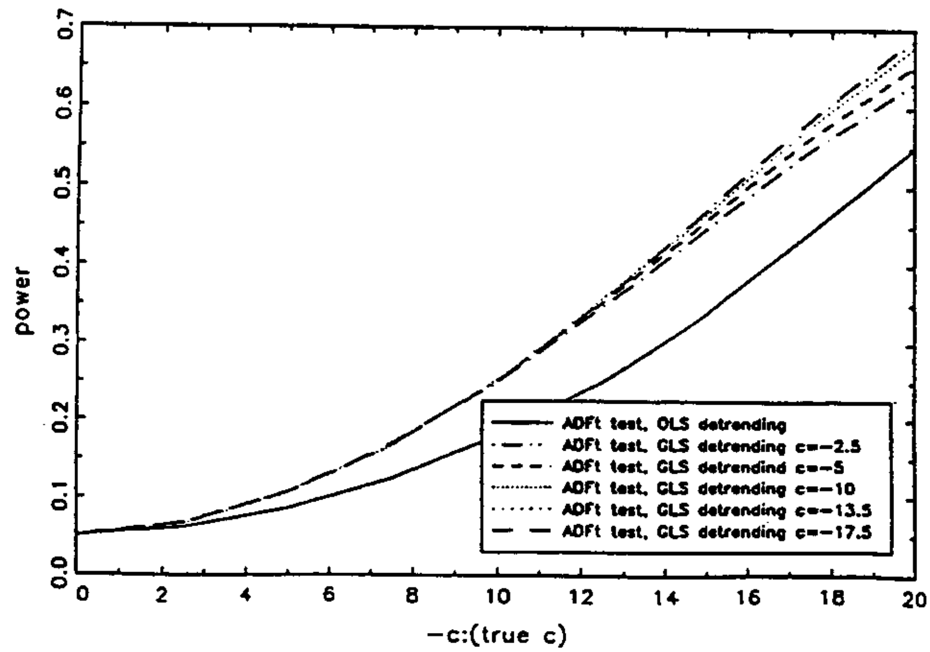

Figure 8. Power of ADFt tests based on different detrending procedures

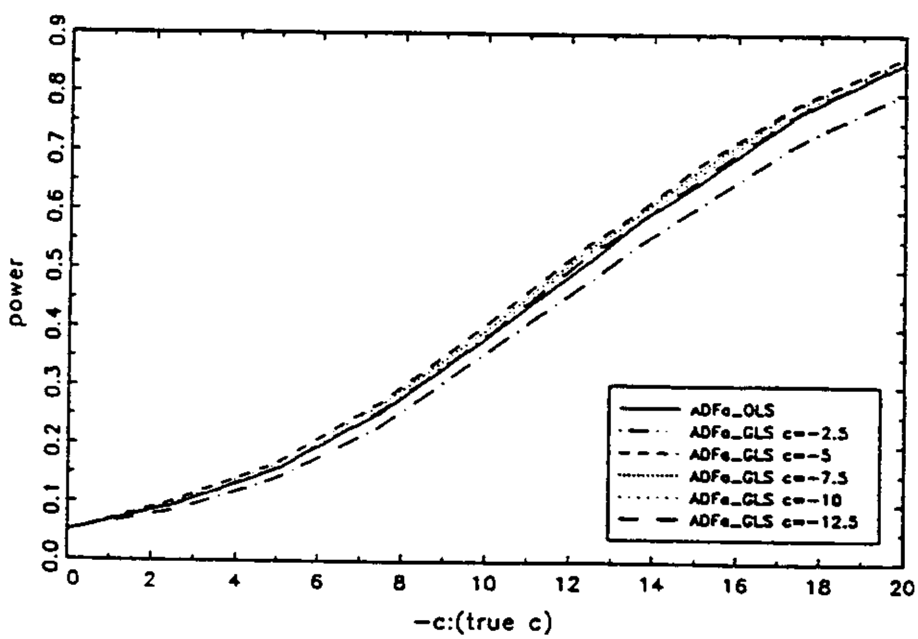

Figure 9. Effect of GLS detrending on ADFa test — iid error

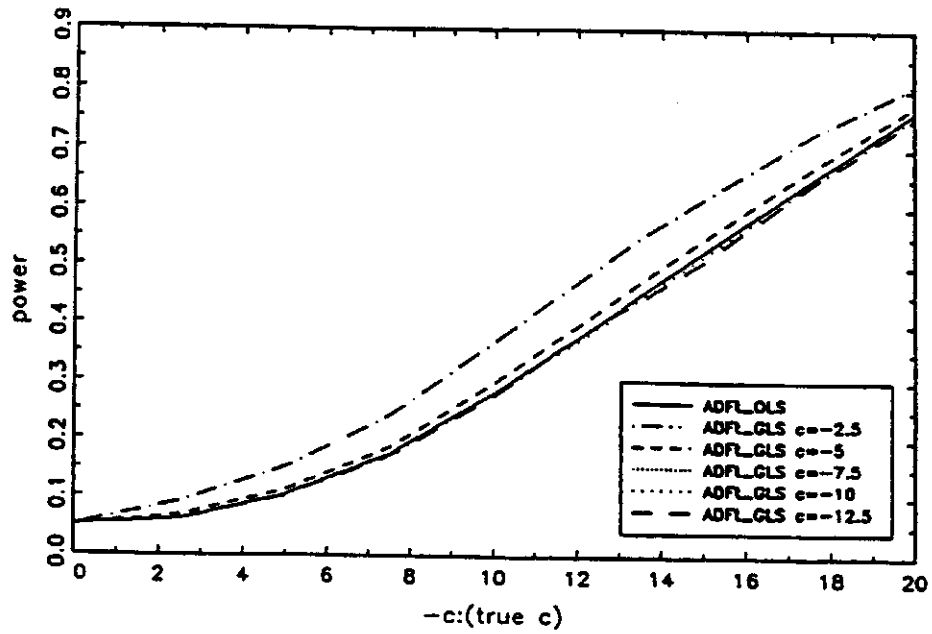

Figure 10. Effect of GLS detrending on ADFt test — iid error

Table A. Size corrected power of unit root tests 5% level, no deterministic trend

Test	α	AR(1) $\rho=0.5$	AR(1) $\rho=-0.5$	MA(1) $\theta=0.5$	MA(1) $\theta=-0.8$
ADF_α	0.95	0.216	0.234	0.181	0.251
	0.9	0.458	0.524	0.478	0.566
	0.85	0.622	0.717	0.650	0.758
ADF_t	0.95	0.211	0.233	0.171	0.248
	0.9	0.448	0.522	0.472	0.565
	0.85	0.618	0.716	0.649	0.761
Z_α	0.95	0.228	0.268	0.198	0.268
	0.9	0.536	0.615	0.520	0.619
	0.85	0.718	0.792	0.698	0.792
Z_t	0.95	0.222	0.268	0.196	0.266
	0.9	0.538	0.618	0.516	0.611
	0.85	0.718	0.806	0.707	0.792

prewhitening and without data-based bandwidth selection. We apply the Andrews-Monahan procedure for estimating the long run variance parameter in the Z tests and found that the data-based bandwidth selection and prewhitening procedures have very important effects on the Z tests. The finite sample performance of the Z tests improves significantly with the use of this estimator of the long run variance, especially for the case with positive serial correlation. The size distortion, although still present, is also decreased after using the prewhitening procedure. Table A shows that the Z tests, with data-based bandwidth choice coupled with prewhitening, generally have higher size corrected power than the ADF tests. Qualitatively similar results can also be found in Figures 1 and 2 when a deterministic trend is included. The results from Figure 1 to Figure 3 also confirm the findings in other simulation experiments that coefficient based tests generally have higher power than the t-ratio tests, at least in large samples, and they show the relatively good finite sample properties of the VN test.

Figures 3 to 10 show the effect of different detrending procedures on the finite sample power of unit root tests. Two forms of deterministic trend were considered in the experiment. The first case, $x_t = a + bt$, includes both an intercept term and a time trend, while the second case, $x_t = t$, does not have a constant term. Figures 3, 4, 5, 6, 7, and 8 correspond to the first case with both a constant term and a time trend, and Figures 9 and 10 correspond to the second case with no constant term. These figures show that QD detrending increases the finite sample power of unit root tests, especially when the deterministic trend includes a constant term. Among the four tests, Z_α, Z_t, ADF_α, and ADF_t, QD detrending brings the largest power gains to the t-ratio based tests. Different choices of the prespecified local parameter $\bar{c}$ were tried and the results show that, in the case $x_t = a + bt$, for quite a wide range of choices of $\bar{c}$, the QD detrended tests have reasonably good power properties against alternatives close to the unit root. Differences among the tests using different $\bar{c}$ occur when the true local parameter c becomes larger (in absolute value), corresponding to alternatives that are distant from a unit root. As the true value of $|c|$ becomes large, a QD detrended test with $\bar{c} = -2.5$ has power lower than tests using a larger $\bar{c}$ (in absolute value). This phenomenon is expected because, from the perspective of a point optimal test, the power should be higher when the prespecified $\bar{c}$ is closer to the true c value.

Another important phenomenon we see from these figures is the difference between the case with a constant term in the deterministic trend and the case with no constant term. Figures 7, 8, 9, and 10 compare the ADF tests combined with different detrending procedures for the two kinds of deterministic trend removal. The QD detrended tests seem to be more sensitive to the choice of $\bar{c}$ in the case without a constant term. A larger power gain from QD detrending is found in the presence of a constant term. Phillips and Lee (1996) provide an analysis of the effects of a fitted intercept. In most practical applications, an intercept will be included in the deterministic trend and so QD detrending can be expected to be successful in improving the finite sample power of unit root tests.

5. Unit root tests against trends with structural breaks

Just as dummy variables are used in regression to deal with unusual observations and shifts in the mean, breaks in deterministic trend functions can be employed to capture changes in trend. This possibility is already included in the specification of h_t in (1). For instance, the trend function

$$h_t = \sum_{j=0}^{p} f_j t^j + \sum_{j=0}^{p} f_{m,j}\, t^j_m, \text{where } t^j_m = \begin{cases} 0 & t \in \{1, \ldots, m\} \\ (t-m)^j & t \in \{m+1, \ldots, n\} \end{cases} \tag{31}$$

allows for the presence of a structural change in the polynomial trend at the data point $t = m + 1$. Suppose $\mu = \lim_{n\to\infty}(m/n) > 0$ is the limit of the fraction of the sample where this structural change occurs. Then the limiting trend function $X_\mu(r)$ corresponding to (31) has a similar break at the point μ. The unit root tests and power functions considered above, including those that make use of efficient detrending procedures, all continue to apply as given for such broken trend functions. Indeed, (31) may be extended further to allow for multiple break points in the sample and in the limit process without affecting the theory. The tests may be interpreted as tests for the presence of a unit root in models where broken trends may be present in the data. The alternative hypothesis in this case is that the data are stationary about a broken deterministic trend of degree p.

In order to construct unit root tests that allow for breaking trends like (31) it is necessary to specify the break point m. (Correspondingly, the limit theory depends on the limit processes $X(r)$ and $\check{X}(r)$ and these depend on the break point μ.) In effect, the break point is exogenously determined. Perron (1989) considered linear trends with single break points in this way. An alternative perspective is that the break point(s) is (are) endogenous to the data and unit root tests should take account of this fact. In this case, alternative unit root tests have been suggested (e.g., Banerjee *et al.*, 1990, and Zivot & Andrews, 1992) that endogenize the break point by choosing the value of m that gives the least favorable view of the unit root hypothesis. This has been done for the parametric ADF test and for linear trends with breaks. If $ADF(m)$ denotes the ADF statistic given by the t-ratio for α in the ADF regression (13) with a broken trend function like (31), then the trend break ADF statistic is

$$ADF(\hat{m}) = \min_{\underline{m} \le m \le \bar{m}} ADF(m), \text{ where } \underline{m} = [n\underline{\mu}],\ \bar{m} = [n\bar{\mu}] \text{ and } 0 < \underline{\mu} < \bar{\mu} < 1, \tag{32}$$

and $[\cdot]$ signifies the integer part of its argument. The limit theory for this trend break ADF statistic is given by

$$ADF(\hat{m}) \Rightarrow \inf_{\mu \in [\underline{\mu}, \bar{\mu}]} \left[\int_0^1 W_{X\mu} dW\right]\left[\int_0^1 W^2_{X\mu}\right]^{-1/2}, \tag{33}$$

where the limit process $X_\mu(r)$ that appears in this functional on the right side is now dependent on the trend break point μ over which the functional is minimized. Similar extensions to trend breaks are possible for the other unit root tests·

considered above. Critical values of the limiting test statistic (33) are naturally further out in the tail than those of the exogenous trend break statistic, so it is harder to reject the null hypothesis of a unit root when the break point is considered to be endogenous.

Asymptotic and finite sample critical values for the endogenized trend break *ADF* unit root test are given in Zivot & Andrews (1992). Simulations studies indicate that the introduction of trend break functions leads to further reductions in the power of unit root tests and to substantial finite sample size distortion in the tests. Sample trajectories of a random walk are often similar to those of a process that is stationary about a broken trend for some particular breakpoint (and even more so when several break points are permitted in the trend). So continuing reductions in the power of unit root tests against competing models of this type is to be expected. In view of the fact that Brownian motion can be represented as an infinite linear random combination of deterministic functions of time, as shown in (29) above, there are good theoretical reasons for anticipating this outcome. Carefully chosen trend stationary models can always be expected to provide reasonable representations of given random walk data, but such models are certain to fail in post sample projections as the post sample data drifts away from the final trend line. Phillips (1998a,1998b) explores these issues in a systematic way.

6. Fractional integration

Although most attention has been focused on $I(1)$ and $I(0)$ processes in econometric applications, the concept of an integrated process generalizes to higher order integration and fractional integration. These concepts are embodied in the following extended version of (2)

$$(1-L)^d y_t^s = u_t, \tag{34}$$

where d may be fractional and the operator $(1-L)^d$ is defined by the formal binomial expansion

$$(1-L)^d = 1 + \sum_{j=1}^{\infty} \frac{(-d)_j}{j!} L^j, \quad (a)_j = (-a)(-a+1), \ldots, (-a+j-1) \tag{35}$$

whose convergence properties depend on the value of d. Note that (35) terminates when d is a positive integer. The process y_t^s is said to be an $I(d)$ process. With this generalization, there may be one or several unit roots (d integer $\geqslant 1$) or fractional integration ($0<d<1$). Such processes have been the subject of intensive recent research and are reviewed in Robinson (1994a) and Baillie (1996). When $0<d<1/2$, y_t^s is stationary but *strongly correlated* in the sense that its lag-j autocovariance γ_j decays at the rate $|j|^{2d-1}$, which is slower than that of stationary linear processes like u_t. When $1/2 \leqslant d \leqslant 1$, y_t^s is nonstationary, and the value $d=1/2$ provides the nexus between stationary and nonstationary regions. When d is an integer $\geqslant 2$, it is called higher order integration. In this case, y_t^s has two or more real autoregressive unit roots and is stationary after differencing d times. A process with $d \geqslant 1/2$ has nonstationary long-memory and a variance that explodes as $t \to \infty$. Such processes are, in fact, not mean reverting, although their impulse responses, which are obtained from the expansion

$$(1 - L)^{-d} = 1 + \sum_{j=1}^{\infty} \frac{(d)_j}{j!} L^j$$

and have the form

$$\frac{(d)_j}{j!} = \frac{1}{\Gamma(d)} \frac{\Gamma(d+j)}{\Gamma(j+1)} \sim \frac{1}{\Gamma(d)} \frac{1}{j^{1-d}} \text{ as } j \to \infty,$$

decay to zero provided $d < 1$, and so shocks in (34) are not persistent in this case.

Within the family (34) it is possible to test for 'unit root' nonstationarity by estimating d and testing the null hypothesis $d = 1$ against the alternative $d < 1$, or to test for stationarity $d < 1/2$ against $d = 1/2$. At present, the literature has focussed on parametric tests, because of the difficulties of a general treatment that covers both stationary and nonstationary cases in the semiparametric case, i.e., when u_t is treated nonparametrically in estimation and inference about d.

Robinson (1994b) took model (1) and (34) and proposed a unit root test against fractional alternatives based on the LM principle. Consider a test of the null hypothesis H_{d_0}: $d = d_0$ in the simple case where the u_t are iid $N(0, \sigma^2)$ variates. Let $L(\eta)$ be the negative of the log-likelihood of u_t, where $\eta = (d, \gamma, \sigma^2)$, then an LM (score) statistic is

$$\tilde{R} = \frac{\partial L(\eta)}{\partial \eta'} \left[E\left(\frac{\partial L(\eta)}{\partial \eta} \frac{\partial L(\eta)}{\partial \eta'} \middle| H_{d_0} \right) \right]^{-1} \frac{\partial L(\eta)}{\partial \eta} \Bigg|_{d = d_0, \gamma = \tilde{\gamma}, \sigma^2 = \tilde{\sigma}^2} = \tilde{r}'\tilde{r}.$$

Let $\Delta^d = (1 - L)^d$ and take fractional differences of equation (1) under the null, leading to

$$\Delta^{d_0} y_t = \gamma' \Delta^{d_0} x_t + \Delta^{d_0} y_t^s. \tag{36}$$

Under H_{d_0}, $\Delta^{d_0} y_t^s = u_t$. Estimating the trend coefficient γ by (36) and calculating the fitted residuals $\tilde{u}_t$ using this trend estimator, we obtain $\tilde{\sigma}^2 = n^{-1} \sum_t \tilde{u}_t^2$. The LM test statistic $\tilde{R}$ can then be calculated. Under certain regularity conditions, Robinson showed that the statistic $\tilde{r}$ has a standard normal limit distribution, and thus the LM test $\tilde{R}$ is asymptotically distributed as χ^2. The asymptotic theory justifies a one-sided test for a unit root H_0: $d = 1$ against a fractional alternative H_1: $d < 1$ which rejects the null hypothesis when $\tilde{r} < z_\alpha$, where z_α is the corresponding critical value of standard normal distribution. Tanaka (1999) considered similar parametric LM tests for the nonstationary case $d \geqslant 1/2$.

Under Gaussian assumptions, efficient parametric estimation of d can be obtained based on the maximum likelihood principle, provided the model is fully specified. However, since calculating these Gaussian estimates requires numerical methods of estimation and the good large sample properties rely on correct specification of the short memory components of the model, simpler estimates have been suggested that do not rely on full specification of the short memory components.

If u_t in (34) were a white-noise sequence with spectrum $f_u(\lambda) = (\sigma^2/2\pi)$, then the spectral density of $y_t^s, f(\lambda, d, \sigma^2)$ satisfies

$$\log f(\lambda, d, \sigma^2) = \log(\sigma^2/2\pi) - d \log|1 - e^{i\lambda}|^2. \tag{37}$$

If we denote the periodogram of y_t^s by $I(\lambda_j) = (2\pi n)^{-1}|\sum_{t=1}^n (y_t^s - \bar{y}^s)e^{it\lambda_j}|^2$, with $\lambda_j = 2\pi j/n$, and $\bar{y}^s = n^{-1}\sum_{t=1}^n y_t^s$, then, the form of (37) suggests a log periodogram regression of the type

$$\log I(\lambda_j) = c - d \log|1 - e^{i\lambda}|^2 + v_s. \tag{38}$$

If we confine attention to a set of $m < n$ fundamental frequencies $\{\lambda_j\}_{j=1}^m$ in this regression, then we can expect this procedure to give satisfactory results even when the spectrum of u_t is quite general because in that case $\log(f_u(\lambda)) - \log(f_u(0)) = c$ for $\lambda \to 0+$, which holds provided $m/n \to 0$. Such log periodogram regressions have been extensively used in empirical research largely because they are so convenient. However, the asymptotic properties of such estimates of d have only recently been obtained (Robinson, 1995; Hurvich *et al.*, 1998) and then only in the stationary Gaussian case. The essential difficulty was pointed out by Kunsch (1986), viz. that the periodogram ordinates $I(\lambda_j)$ in (38) are asymptotically correlated for fixed j. At present, there is no published asymptotic theory for the nonstationary case.

Another approach is to use an explicit model, like the following, to approximate the spectrum of a process with long-range dependence (see Robinson, 1994a)

$$f(\lambda, d, \beta) = \exp\left[\sum_{k=1}^{p-1} \beta_k \cos\{(k-1)\lambda\}\right]|1 - e^{i\lambda}|^{-2d}.$$

Notice that $\log|1 - e^{i\lambda}| = \sum_{k=1}^{\infty}(\cos k\lambda/k)$, so the above representation can be reparameterized as

$$f(\lambda, d, \beta) = \exp\left[\sum_{k=1}^{p-1} \theta_k \cos\{(k-1)\lambda\} - 2d\sum_{k=1}^{\infty} \frac{\cos k\lambda}{k}\right],$$

where $\theta = (\theta_1, \ldots, \theta_{p-1}, d)'$. Thus, the logarithm of $f(\lambda, d, \theta)$ is a linear function of θ and linear regression of $\log I(\lambda_j)$ on these components can be applied to estimate the parameters.

Another quite different approach is to locally approximate the Gaussian likelihood in the frequency domain, leading to the following objective function suggested by Künsch (1986)

$$Q_m(G, d) = \frac{1}{m}\sum_{j=1}^{m}\left[\log(G\lambda_j^{-2d}) + \frac{\lambda_j^{2d}}{G} I_x(\lambda_j)\right] \tag{39}$$

(39) are d and $G = f_u(0)$ and are estimated by minimizing $Q_m(G, d)$, so that

$$(\hat{G}, \hat{d}) = \arg \min_{0 < G < \infty, \hat{d} > 0} Q_m(G, d),$$

which involves numerical optimization. Concentrating (39) with respect to G, we find that the estimate $\hat{d}$ satisfies

$$\hat{d} = \arg \min_d R(d),$$

where

$$R(d) = \log \hat{G}(d) - 2d \frac{1}{m} \sum_{j=1}^{m} \log \lambda_j, \quad \hat{G}(d) = \frac{1}{m} \sum_{j=1}^{m} \lambda_j^{2d} I_x(\lambda_j).$$

Recently, Robinson (1995a) analyzed the above estimators in the stationary case where $d \in (-\frac{1}{2}, \frac{1}{2})$. Under rather weak regularity conditions on the smoothness of $f_u(\lambda)$, the innovations in the Wold representation of u_t and an expansion rate condition on m, which requires that $m \to \infty$ but $(m/n) \to 0$ as $n \to \infty$, Robinson showed that $\hat{d} \to_p d_0$ and $\hat{G}(\hat{d}) \to_p G_0$. Under a slight strengthening of these conditions, Robinson also established that $\hat{d}$ is asymptotically normally distributed with the limit distribution

$$m^{1/2}(\hat{d} - d) \xrightarrow{d} N(0, \tfrac{1}{4}). \tag{40}$$

This limit theory makes testing and the construction of confidence intervals for d_0 a straightforward matter in the stationary case.

In recent unpublished work, Phillips (1998c) has dealt with the nonstationary case where $d \in (\frac{1}{2}, 1]$ and, under regularity conditions that are broadly similar to those of Robinson (1995b), has established that

$$\hat{d} \xrightarrow{p} d, \quad \hat{G}(\hat{d}) \xrightarrow{d} G_0 + C(d), \tag{41}$$

and

$$m^{1/2}(\hat{d} - d_0) \Rightarrow MN\left(0, \frac{1}{4} \frac{G_0^2}{(G_0 + C(d_0))^2}\right), \tag{42}$$

where $C(d) > 0$ is a random and depends on the true value of d. Since the variance in (42) is smaller than 1/4, conservative confidence intervals can be constructed for d that utilize the limit theory (40) and apply for both stationary and nonstationary d.

Another approach to estimation of d stems from the properties of the autocovariances. If we approximate the spectral density $f(\lambda)$ by $c\lambda^{-2d}$ for λ close to 0, and denote $\mathrm{cov}(y_j, y_{j+h})$ by γ_h, then, under certain conditions (Yong, 1974), the relation $f(\lambda) \sim G\lambda^{-2d}$ as $\lambda \to 0$ is equivalent to '$\gamma_h \sim gh^{2d-1}$ as $h \to \infty$', for suitable g, providing semiparametric estimates of d based on estimates of γ_h for large h. For example, if $\hat{\gamma}_h$ are consistent estimates of γ_h, d can be consistently estimated by the following semiparametric estimate

$$\hat{d} = \frac{-\frac{1}{2}\sum_{h=n-p}^{n-1} \log \hat{\gamma}_h(\log h - \overline{\log h})}{\sum_{h=n-p}^{n-1} (\log h - \overline{\log h})^2},$$

where $\overline{\log h} = p^{-1}\sum_{h=n-p}^{n-1} \log h$, p increase suitably with n. Alternatively, d can be estimated by minimizing $\sum_{h=n-p}^{n-1}(\hat{\gamma}_h - Ch^{2d-1})^2$.

Other procedures than the ones discussed above are available and the subject is still under intensive study. Two recent surveys on the topic are given by Beran (1992) and Robinson (1994), and Baillie (1996) reviews many of the empirical aspects of fractional integration. Some applications to conditional heterogeneity are discussed in Baille *et al.* (1996).

7. Seasonal unit root tests

Another important extension of (2) is the seasonal unit root model

$$(1 - L^4)y_t = u_t. \tag{43}$$

Notice that the polynomial $1 - L^4$ can be expressed as $(1 - L)(1 + L)(1 + L^2)$. Thus, the unit roots (or roots on the unit circle) in (43) are 1, -1, i, and $-i$, corresponding to the annual ($L = 1$) frequency, the semi-annual ($L = -1$) frequency, and the quarter and three quarter annual ($L = i, -i$) frequency respectively. The model (43) is often relevant in practical work with quarterly data. Quarterly differencing, as in (43), is sometimes used as a seasonal adjustment device, and it is of interest to test whether the data supports the implied hypothesis of the presence of unit roots at these seasonal frequencies. Other types of seasonal processes, say monthly data, can be analyzed in the same way.

Dickey *et al.* (1984) proposed a test for the presence of a single unit root at a seasonal lag by considering the following model

$$y_t = \alpha y_{t-s} + \varepsilon_t.$$

The null hypothesis is $\alpha = 1$, and the alternative is $\alpha < 1$. The limit distribution of the least squares estimate of α is given and small-sample distributions for several values of s are provided based on Monte Carlo experiments. A more general case was studied by Hylleberg *et al.* (1990), extending the parametric *ADF* test to the case of seasonal unit roots. In order to accommodate fourth differencing as in (43) the autoregressive model is written in the new form

$$\Delta_4 y_t = \alpha_1 y_{1t-1} + \alpha_2 y_{2t-1} + \alpha_3 y_{3t-2} + \alpha_4 y_{3t-1} + \sum_{j=1}^{p} \varphi_j \Delta_4 y_{t-j} + \varepsilon_t, \tag{44}$$

where $\Delta_4 = 1 - L^4$, $y_{1t} = (1 + L)(1 + L^2)y_t$, $y_{2t} = -(1 - L)(1 + L^2)y_t$, $y_{3t} = -(1 - L^2)y_t$. The transformed data y_{1t}, y_{2t}, y_{3t} retain the unit root at the zero frequency (long run), the semi-annual frequency (two cycles per year), and the annual frequency (one cycle per year). When $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$, there are unit roots at the zero and seasonal frequencies. To test the hypothesis of a unit root ($L = 1$) in this seasonal model, a t-ratio test of $\alpha_1 = 0$ is used. Similarly, the test for a semi-annual root ($L = -1$) is based on a t-ratio test of $\alpha_2 = 0$, and the test for an annual root on the t-ratios for $\alpha_3 = 0$ or $\alpha_4 = 0$. If each of the α's is different from zero, then the series has no unit roots at all and is stationary. Details of the implementation of this procedure are given in Hylleberg *et al.* (1990) and the limit theory for the tests is developed in Chan and Wei (1988).

As an alternative approach to the conventional seasonal unit root analysis, periodic models have been used in the study of seasonality. Under the framework of a periodic model, parameters are allowed to vary according to the time at which the series are observed. A definition of periodic integration and some testing procedures are provided in Osborn *et al.* (1988). Franses (1996) provides a useful reference on periodicity in the context of models with stochastic trends. There has also been work on seasonal versions of the fractional integration model (Porter Hudak, 1990), and seasonal versions of error correction models and reduced rank regressions (McAleer and Franses 1998). Hylleberg (1994) is a useful general reference on the topic.

8. Bayesian unit root tests

While most practical work on unit root testing has utilized classical procedures of the type discussed above, Bayesian methods offer certain advantages that are useful in empirical research. Foremost among these is the potential that these methods offer for embedding the unit root hypothesis in the much wider context of model specification. Whether or not a model such as (4) has a unit root can be viewed as part of the overall issue of model determination. Model comparison techniques like *posterior odds* and *predictive odds* make it easy to assess the evidence in the data in support of the hypothesis $\alpha = 1$ at the same time as decisions are made concerning other features of model specification, such as the lag order in the autoregression (6), the degree of the deterministic trend component, and the presence of trend breaks. A common asymptotic theory (see Phillips and Ploberger, 1996) further facilitates this approach to model selection and leads to an extension of the Schwarz (1978) BIC criterion to models with some nonstationary data that is based on the idea of selecting the model that is *a posteriori* the most probable. The approach has connections with *prequential probability* (Dawid, 1984) and *stochastic complexity* (Rissanen, 1986). It can be shown that model choices that are made in this way are completely consistent in the sense that the probability of type I and type II errors goes to zero as $n \to \infty$ (Phillips and Ploberger, 1994).

In the context of Bayesian analysis, a model may be selected based on the posterior odds ratio

$$\left(\frac{\pi_2}{\pi_1}\right)\frac{\Pr[Y^n \mid \mathcal{M}_2]}{\Pr[Y^n \mid \mathcal{M}_1]}$$

where $\mathcal{M}_1$ and $\mathcal{M}_2$ are the two candidate models, and π_1 and π_2 are corresponding prior weights on $\mathcal{M}_1$ and $\mathcal{M}_2$. If $\mathcal{M}_1$ and $\mathcal{M}_2$ are specified as $I(0)$ and $I(1)$ models respectively, the unit root hypothesis can be tested by choosing between models $\mathcal{M}_1$ and $\mathcal{M}_2$. More rigorously, consider a linear regression

$$y_t = \theta' z_t + \varepsilon_t,$$

where y_t and ε_t are real valued stochastic processes on a probability space $(\Omega, \mathcal{F}, P)$, and $\mathcal{F}_t \subset \mathcal{F}(t = 0, 1, 2, \ldots)$ is a filtration to which y_t and ε_t are adapted, z_t is a $k \times 1$ vector defined on the same space and is $\mathcal{F}_{t-1}$ measurable). Suppose $\varepsilon_t \equiv$ iid $N(0, \sigma^2)$ and then, conditional on $\mathcal{F}_0$ and θ, the joint density of $Y_n = [y_1, \ldots, y_n]'$ with respect to Lebesgue measure v is

$$\begin{aligned}\mathrm{pdf}(Y_n \mid \mathcal{F}_0, \theta) &= dP_n^\theta / dv \\ &= (2\pi\sigma^2)^{-n/2} \exp\{-(1/2\sigma^2)[\hat{U}_n'\hat{U}_n + (\hat{\theta}_n - \theta)' A_n (\hat{\theta}_n - \theta)]\}, \quad (45)\end{aligned}$$

where $\hat{U}_n = Y_n - Z_n\hat{\theta}_n$, $\hat{\theta}_n = [Z_n'Z_n]^{-1}Z_n'Y_n$, $Z_n = [z_1, \ldots, z_n]$, $A_n = Z_n'Z_n$, and P_n^θ is the probability measure of Y_n. Combining (45) with a prior density, $\pi(\theta)$, for θ, we get the joint density of (θ, Y_n). Conditional on $\mathcal{F}_0$, the data density for Y_n can then be obtained by integrating out θ in the joint density of (θ, Y_n). For $\pi(\theta) = \pi_0$,

$$\mathrm{pdf}(Y_n \mid \mathcal{F}_0) = \pi_0 (2\pi\sigma^2)^{-(n-k)/2} |A_n|^{-1/2} \exp\{-(1/2\sigma^2)\hat{U}_n'\hat{U}_n\}.$$

Let Q_n be the (probability) measure whose density with respect to v is $\mathrm{pdf}(Y_n \mid \mathcal{F}_0)$ and choose $P_n = P_n^0$ as the reference measure, then

$$\begin{aligned}\frac{dQ_n}{dP_n} &= \int \pi(\theta) \frac{dP_n^\theta}{dP_n} d\theta \\ &= \pi_0 (2\pi\sigma^2)^{k/2} |A_n|^{-1/2} \exp\{(1/2\sigma^2)\hat{\theta}_n' A_n \hat{\theta}_n\}.\end{aligned}$$

For all $n > k$, Q_n as given above leads to a proper conditional probability measure and this measure can be interpreted as the Bayesian version of the data generating mechanism. In other words, Q_n gives us the Bayesian model for the data.

A natural measure of model adequacy is provided by the data density dQ_n/dP_n. If we denote Q_n^k as the 'Bayes model' measure given by Q_n for a model with k parameters and incorporate the index 'k' in what follows to signify the number of regressors, then

$$\frac{dQ_n^k}{dP_n} = |(1/2\sigma^2)A_n(k)|^{-1/2} \exp\{(1/2\sigma^2)\hat{\theta}_n'(k)A_n(k)\hat{\theta}_n(k)\}, \quad k = 1, 2, \ldots, K.$$

Let K be some maximum number of regressors, then Q_n^K corresponds to the 'least restricted' option and we may use it as the reference measure. Multiplying the Radon-Nikodym derivatives we obtain the likelihood ratio

$$\frac{dQ_n^k}{dQ_n^K} = \left(\frac{dQ_n^k}{dP_n}\right)\left(\frac{dP_n}{dQ_n^K}\right)$$

corresponding to the two 'Bayes models'

$$H(Q_n^k):\ y_{n+1} = \hat{\theta}_n(k)' z_{n+1} + v_{n+1}^k,$$
$$H(Q_n^K):\ y_{n+1} = \hat{\theta}_n(K)' z_{n+1} + v_{n+1}^K.$$

In model (13), if we set $K = p + q$, and $k = p + q - 1 = K - 1$, where q is the dimension of the deterministic trend x_t, we get the following two 'Bayes models':

$$H(Q_n^{K-1}): \Delta y_{n+1} = \sum_{j=1}^{p-1} \tilde{\varphi}_{jn} \Delta y_{n+1-j} + \tilde{\beta}_n' x_{n+1} + \varepsilon_{n+1}, \text{ and}$$

$$H(Q_n^K): \Delta y_{n+1} = \tilde{a} y_n + \sum_{j=1}^{p-1} \tilde{\varphi}_{jn} \Delta y_{n+1-j} + \tilde{\beta}_n' x_{n+1} + \varepsilon_{n+1}.$$

The first model incorporates a unit root. If we assign equal prior odds to the two models we can test the hypothesis of a unit root, i.e., $H(Q_n^{K-1})$, against trend stationarity, i.e., $H(Q_n^K)$, using the criterion:

$$\text{Accept } H(Q_n^{K-1}) \text{ in favor of } H(Q_n^K) \text{ if } dQ_n^{K-1}/dQ_n^K > 1. \tag{46}$$

Thus, under the wider context of model specification, testing for a unit root becomes an issue of whether or not to include the lag variable (y_n above) as a regressor in the *ADF* regression. This criterion, as shown in Phillips and Ploberger (1994), gives a completely consistent 'Bayes model' test in the sense that the probabilities of both types of error tend to zero as $n \to \infty$.

A second advantage of Bayesian techniques in nonstationary models with unit roots is that the asymptotic form of the posterior density is normal under rather general conditions, a result that facilitates large sample Bayesian inference and that contrasts with the non-standard asymptotic distribution theory of classical estimators and tests. The fact that posterior densities have limiting normal forms in a wider class of models than those for which the maximum likelihood estimator is asymptotically normal has long been known (Heyde and Johnstone, 1979) but its relevance for models with unit roots has only recently been recognized (Sims, 1990; Kim, 1994; Phillips and Ploberger, 1996). For instance, a large sample Bayesian confidence set for the autoregressive parameter α in (11) can be constructed in the conventional way without having to appeal to any nonstandard limit theory. In this respect, Bayesian theory (which leads to a symmetric confidence set for α) differs from classical statistical analysis where the construction of valid confidence regions is awkward because of the discontinuity of the limit theory at $\alpha = 1$ (but may be accomplished using local asymptotics as in Stock, 1991). This divergence can lead to quite different inferences being made from the two approaches with the same data even when the influence of the prior is negligible, as it is in very large samples. In small samples, the role of the prior is important and time series models raise special concerns about the construction of uninformative priors, primarily because a great deal is known about the properties of simple time series models like autoregressions and their characteristic features in advance of data analysis. How this knowledge should be used or ignored is a matter on which there is ongoing debate (see Phillips, 1991, and two themed issues of the *Journal of Applied Econometrics*, 1991, and *Econometric Theory*, 1994).

Third, Bayesian methods offer flexibility and convenience in analyzing models with possible unit roots and endogenous trend breaks. In such cases, a prior distribution of break points is postulated (a uniform prior across potential break points may be appealing in the absence of other information), the posterior mass function is calculated, and the Bayes estimate of the break point is taken as the one with highest posterior mass (Zivot and Phillips, 1994). This approach makes the analysis of multiple break points straightforward, a problem where classical asymptotic theory is much more complex (For a classical analysis of multiple breaks, see Bai, 1997; Lumsdaine and Papell, 1995, among others).

9. Bootstrapping unit root tests

The study of bootstrapping time series regression models was started by Freedman (1984). Bose (1988) shows that under certain regularity conditions, the bootstrap approximation to the distribution of the least squares estimator in a stationary autoregressive model is of order $o(n^{-1/2})$ a.s., thereby improving the normal approximation. The validity of the bootstrap for unit root models has been studied by several authors recently. Basawa *et al.* (1991a) show that the standard bootstrap least squares estimate is asymptotically invalid in unit root models, even if the error distribution is assumed to be normal. Consider the following AR(1) process:

$$y_t = \alpha y_{t-1} + u_t,\ y_0 = 0,\ u_t \equiv \text{iid } N(0, 1).$$

Let $\hat{\alpha}$ be the OLS estimator of α, then under the null hypothesis of a unit root and when $n \longrightarrow \infty$,

$$Z_n = \left[\sum y_{t-1}^2\right]^{-1/2} (\hat{\alpha} - \alpha) \Rightarrow \tfrac{1}{2}\{W(1)^2 - 1\}\left\{\int_0^1 W(r)^2 dr\right\}^{-1/2}. \tag{47}$$

A parametric bootstrap sample y_t^* is obtained recursively from the following recursion:

$$y_t^* = \hat{\alpha} y_{t-1}^* + u_t^*,\ y_0^* = 0, \tag{48}$$

where $\{u_t^*\}$ is a random sample from $N(0, 1)$. The bootstrap estimator of $\hat{\alpha}$ can then be calculated from the bootstrap sample $\{y_t^*\}$, as $\hat{\alpha}^* = [\sum y_t^* y_{t-1}^*] \times [\sum y_{t-1}^{*2}]^{-1}$, and the bootstrap version of Z_n is $Z_n^* = [\sum y_{t-1}^{*2}]^{1/2}(\hat{\alpha}^* - \hat{\alpha})$. Basawa *et al.* (1991a) show that the limit distribution of Z_n^* is not the same as that of Z_n, thus invalidating the bootstrap. Specifically, consider the triangular array

$$y_{k,n} = b_n y_{k-1,n} + u_k, \quad y_0 = 0, \tag{49}$$

where ε_k are independent $N(0, 1)$ variates and $\{b_n\}$ is a sequence of numbers such that $n(b_n - 1) \longrightarrow \lambda$. Then, (49) is a triangular system with roots $b_n = 1 + (\lambda/n)$

that are local to unity. It follows from our earlier analysis that if $b_n^* = [\sum_k y_{k,n} y_{k-1,n}][\sum_k y_{k-1,n}^2]^{-1}$, and $\tau_n(b_n) = [\sum_k y_{k-1,n}^2]^{1/2}(b_n^* - b_n)$ is the corresponding sequence of scaled and centered estimators from (49), then

$$\tau_n(b) \Rightarrow Z_n = \left[\int_0^1 J_\lambda(r)^2 dr\right]^{-1/2} \int_0^1 J_\lambda(r) := \eta(\lambda),$$

where $J_\lambda(r) = \int_0^r e^{(r-s)\lambda} dW(s)$ is a linear diffusion. If the bootstrap approximation were asymptotically valid, then along almost all paths $\tau_n(\hat{\alpha})$ would converge to the same distribution as that of $Z_n = [\sum y_{t-1}^2]^{1/2}(\hat{\alpha} - \alpha)$ itself, viz. (47). However, in fact, $\tau_n(\hat{\alpha}) \Rightarrow \eta(\xi)$, as $n \to \infty$, where

$$\xi = \tfrac{1}{2}\{W(1)^2 - 1\}\left\{\int_0^1 W(r)^2 dr\right\}^{-1}$$

since $n(\hat{\alpha} - 1) \Rightarrow \xi$. A similar invalidity of the bootstrap occurs for the coefficient based bootstrap statistic $n(\hat{\alpha}^* - \hat{\alpha})$.

To circumvent the asymptotic invalidity of the bootstrap, there are several possibilities. Basawa *et al.* (1991b) suggested resampling the restricted residuals under the null hypothesis of a unit root. If u_t^{**} is the restricted residual $y_t - y_{t-1}$ and the bootstrap sample is generated from the following resampling scheme under the null

$$y_t^* = y_{t-1}^* + u_t^{**},$$

then the bootstrap t-ratio statistic Z_n^* has the same asymptotic distribution as Z_n. As an alternative, Ferretti and Romo (1994) consider the unrestricted residuals $u_t^* = y_t - \hat{\alpha} y_{t-1}$ and generate the pseudo data y_t^* under the null of a unit root. They show that the corresponding bootstrap test statistics are asymptotically valid. Another possibility is to use model selection to determine whether to sample from a unit root process or the fitted regression. In this case, a natural procedure is to use the model selection method outlined in (46) above, which gives a consistent model choice procedure in which the type I and type II errors both tend to zero in probability as $n \to \infty$. In consequence, the correct model is chosen asymptotically and the resulting bootstrap test statistics are asymptotically valid.

None of the above methods work when the underlying model is a near integrated process, because there in this case there is no way in which the localizing parameter can be consistently estimated from the sample data.

For unit root models with deterministic trends, the problem of redundant deterministic trend variables discussed in Section 2 surfaces again. Although we use regression models like (6), which include redundant trend variables, to calculate the DF test statistics and to obtain the bootstrap residuals ε_t^*, the redundant variables should be excluded from the resampling scheme to make the bootstrap DF test asymptotically valid. Nankervis and Savin (1994) present some simulation results on bootstrapping unit root tests for the following model

$$y_t = \mu_0 + \mu_1 t + \alpha y_{t-1} + \varepsilon_t$$

under different error distributions. Their sampling scheme is based on restricted residuals and the bootstrap sample data is generated under the null hypothesis. The bootstrap DF test has basically the same power as the original DF test, except for some non-Gaussian distribution cases where the bootstrap tests perform slightly better. Another study of the asymptotic properties of bootstrap procedures in unit root models with a drift is Giersbergen (1995).

10. Testing stationarity

Many empirical analyses, like those of Nelson and Plosser (1982) and subsequent studies, lead to the conclusion that aggregate economic time series are unit root nonstationary. One explanation that has been suggested for these empirical outcomes is that the standard tests are all based on the null hypothesis of a unit root, which assures that the hypothesis will be accepted at conventional significance levels (of 5% and 1%) unless there is strong evidence against it. As a result, there is considerable interest in tests for which the null hypothesis is trend stationary.

Such tests are easily developed by working from the components representation of the time series y_t. In particular, if we add a stationary component v_t to (1) and (2), we get the so-called *components* model

$$y_t = h_t + y_t^s + v_t, \quad y_t^s = y_{t-1}^s + v_t, \tag{50}$$

which decomposes the time series y_t into a deterministic trend h_t, a stochastic trend y_t^s, and a stationary residual v_t. The stochastic trend in (50) is annihilated when $\sigma_r^2 = \mathrm{var}(r_t) = 0$, which therefore corresponds to a null hypothesis of trend stationarity. Under Gaussian assumptions and iid error conditions, the hypothesis can be tested in a simple way using the LM principle. Let $\hat{e}_t$ be the residuals from the regression of y_t on the deterministic trend x_t and $\hat{\sigma}_v^2 = n^{-1}\sum \hat{e}_t^2$, then the LM statistic can be constructed as follows:

$$LM = \frac{n^{-2}\sum S_t^2}{\hat{\sigma}_v^2},$$

where S_t is the partial sum process of the residuals $\sum_{j=1}^{t} \hat{e}_j$. Under the null hypothesis of stationarity, this LM statistic converges to $\int_0^1 V_X^2$, where $V_X(r) = W(r) - [\int_0^r X'][\int_0^1 XX']^{-1}[\int_0^1 XdW]$ is a generalized Brownian bridge process, like (18) above. This procedure can easily be extended to more general cases where there is serial dependence by replacing $\hat{\sigma}_v^2$ with corresponding estimates of the long run variance of v_t based on nonparametric methods. This was done in Kwiatkowski *et al.* (1992), where a general approach was developed.

Defining $w_t = y_t^s + v_t$ and writing differences as $\Delta w_t = (1 - \theta L)\eta_t$, where η_t is stationary, it is clear that $\sigma_u^2 = 0$ in (50) corresponds to the null hypothesis of a moving average unit root $\theta = 1$ in this representation. Thus, there is a formal correspondence between testing for stationarity and testing for a moving average unit root (Saikonnen and Luukkonen, 1993). The asymptotic theory for the maximum likelihood estimator in the moving average unit root case is known, but has a complex point process representation (Davis and Dunsmuir, 1995). This makes a likelihood ratio approach awkward and the LM test attractive in practice. Leybourne and McCabe (1994) suggested a similar test for stationarity which differs from the test of Kwiatkowski *et al.* (1992) in its treatment of autocorrelation and applies when the null hypothesis is an AR(k) process.

11. Applications and empirical evidence

Most empirical applications of unit root tests have been in the field of economics. Martingales play a key role in the mathematical theory of efficient financial markets (Duffie, 1988) and in the macroeconomic theory of the aggregate consumption behavior of rational economic agents (Hall, 1978). In consequence, economists have been intrigued by the prospect of testing these theories. In the first modern attempt to do so using unit root tests, Nelson and Plosser (1982) tested fourteen historical macroeconomic time series for the United States by the ADF test. The time series start around 1860 to 1909 and end in 1970. Nelson and Plosser analyzed the logarithms of all of these series (except for interest rates, which was treated in levels) and found empirical evidence to support a unit root for thirteen of them (the exception being unemployment). Since then, these series have been re-tested hundreds of times with other methods, and thousands of other time series have been examined in the literature. Meese and Singleton (1982) studied various exchange rate time series and could not reject the null hypothesis of a unit root; and Perron (1988) applied the semiparametric Z tests to the Nelson–Plosser data and some other macroeconomic time series and basically confirmed the conclusion reached by Nelson and Plosser. While it is recognized that the discriminatory power of unit root tests is often low, there is a mounting body of evidence that many economic and financial time series are well characterized by models with roots at or near unity.

Although standard tests of the unit root hypothesis against trend stationary alternatives usually cannot reject the null hypothesis, other approaches do find different results. For example, performing a test for the null hypothesis of stationarity against the alternative of a unit root, Kwiatkowski *et al.* (1992) revisited the Nelson-Plosser data and could not reject the hypothesis of trend stationarity in many of these time series (including real per capita GNP, employment, unemployment rate, GNP deflator, wages and money). Tests based on efficient detrending by quasi-differencing have also been applied to macroeconomic time series and various results have been reported. For instance, applying the QD detrended ADF test to the U.S. GNP data, Cheung and Chin (1995) could not reject the unit root hypothesis in quarterly data but did get different results with annual data.

Gil-Alana and Robinson (1997) applied the LM test for a unit root against fractional alternatives to the extended Nelson-Plosser series (Schotman and van Dijk, 1991). Although their results vary across the fourteen series and across different model structures for the stationary component u_t, they found the most nonstationary evidence for the consumer price and money stock series, trend stationary evidence for industrial production and stationary evidence for the unemployment rate data.

Using the Nelson-Plosser data and a U.S. postwar quarterly real GNP series, Perron (1989) argues that if the Great Depression in 1929 and the oil price shock in 1973 are treated as exogenous events that caused structural changes, then a trend stationary representation with structural change is favored over a unit root representation with structural change. By allowing for these structural breaks, Perron rejected the unit root hypothesis at the 5% level of significance for all of the Nelson-Plosser series except consumer prices, velocity, and bond yields. Christiano (1992), Banerjee *et al.* (1990), and Zivot and Andrews (1992) argue that Perron's tests for a unit root with structural change are biased because the choices of the break points are correlated with the data. Christiano (1992) suggested that the date of the break should be treated as unknown and, by using tests based on bootstrap critical values, reached different conclusions from Perron (1989). Zivot and Andrews (1992) allowed the breakpoint to be endogenous and suggested pre-testing procedures to estimate the structural change points, finding less compelling evidence against the unit root hypothesis.

In recent years, various Bayesian analyses have been conducted on unit root-testing. Using flat prior Bayesian techniques, DeJong and Whiteman (1989a,b,c) tested the Nelson-Plosser series, stock prices and dividend data, and postwar quarterly real GNP for the U.S.A{50 }. Their results challenged the classical unit root tests results in many cases. Schotman and van Dijk (1991) analyzed the random walk hypothesis for real exchange rates and found more evidence in favor of the trend stationary model than classical unit root tests. Contrary to the conclusion of DeJong and Whiteman, Phillips (1991) provided an alternative Bayesian approach using a Jeffreys' prior, and found more support for the unit root model for some series. Using a modified information matrix-based prior, Zivot and Phillips (1994) considered autoregressive models with fitted deterministic trends allowing for certain types of structural change. Their results are generally in accord with those of Phillips (1991). In addition, their Bayesian analysis also shows evidence of trend breaks in some of the macroeconomic series with breaks occurring around 1929, partially supporting the conclusion reached by Perron (1989). It is also shown in Zivot and Phillips (1994) and in other work that the choice of prior can be important in distinguishing between different models. Using the extended Nelson-Plosser data and a Bayesian procedure that consistently classifies the time series as $I(1)$ or $I(0)$, Stock (1994) obtain results largely supporting the unit root hypothesis. Applying the model selection criterion 'PIC' to the Nelson-Plosser data and allowing for model selection of deterministic trend components and lag length in the autoregressions, Phillips and Ploberger (1994) found eleven out of the fourteen time series to be stochastically nonstationary.

Of course, unit root issues in multivariate time series have also attracted a good deal of research, and the *ADF* and semi-parametric Z tests have been extensively used to test for the presence of cointegration using residual based approaches. The tests are used in the same way as standard unit root tests and have the same null hypothesis, but the data are the residuals from a least squares cointegrating regression, and the alternative hypothesis (of cointegration) is now the main hypothesis of interest (Engle and Granger, 1987; Phillips and Ouliaris, 1990). The model is analogous to (1), but both variables y_t and x_t have unit roots and y_t^s is stationary under the alternative hypothesis and unit root nonstationary under the null. The limit theory for these residual based tests was developed in Phillips and Ouliaris (1990). There are also approaches to cointegration testing that rely on likelihood ratio methods (Johansen, 1996) in vector autoregressions and these lead to tests with asymptotic distributions that are simple multivariate analogues of those given in (8) and (12). A large empirical literature has developed around these techniques. More recently, model selection methods have been advocated in Phillips (1996) and Chao and Phillips (1996, 1997). In this work, model selection is used to simultaneously choose the lag length and cointegrating rank in a VAR of possible reduced rank. The method is extremely easy to use and like the PIC test for a unit root that is discussed in Section, produces consistent estimates of lag length and cointegrating rank. The methods have been used with some success in simulations (Phillips, 1998d) and in ex ante forecasting exercises with macroeconomic data for the USA and several Asia-Pacific countries (Phillips, 1995).

Unit root theory plays a major role in modern time series econometrics and weak convergence methods and function space asymptotics have opened up the econometric analysis of nonstationary regression models. While a multitude of test procedures are available for evaluating evidence in support of unit root nonstationarity, fractional integration and short memory stationarity, the main principles of statistical testing are analogous to those in stationary time series and many of the same issues figure in the analysis. However, the nonstandard limit theory of unit root tests does complicate classical inference and there are important new issues that arise from the nonparametric treatment of the stationary component and the existence of valid alternative models for nonstationary data, as discussed in Section 3.5. Further, unit root models provide an interesting case of divergence between the asymptotic behavior of Bayesian and classical estimators and tests. They also provide an instance of the asymptotic failure of the bootstrap. With these interesting characteristics, it is hardly surprising that the field has attracted so much attention in the last 15 years. Additionally, most economic time series have clearly evident nonstationary empirical characteristics, and there are strong reasons in economic theory for giving attention to the martingale hypothesis and for wanting to distinguish between models with persistent and non persistent shocks. For all these reasons, the field has attracted a full spread of participants from empirical macroeconomists interested in growth and finance theorists interested in efficient markets, through to econometricians and statisticians interested in the development of new testing procedures, asymptotic theory and unified methods of inference for data of this type.

Acknowledgements

This review paper is a considerably expanded version of 'A primer on Unit Root Tests' by Phillips and draws on material published in the brief review of unit root testing given in Phillips (1997). We thank Les Oxley and Michael McAleer for comments on an early version of this paper. The paper was keyboarded by the authors in Scientific Word 2.5. Phillips thanks the NSF for research support under grant no. SBR94–22922 and Xiao thanks the Cowles Foundation for hospitality during a visit in May–June 1998.

References

Akaike, H. (1977) On entropy maximization principle, In P. R. Krishnarah (ed.), *Applications of Statistics*. Amsterdam: North-Holland, 1977, pp. 27–41.

Andrews, D. W. K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica*, **59**, 817–858.

Andrews, D. W. K. and J. C. Monahan (1992) An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator, *Econometrica*, **60**, 953–966.

Bai, Jushan (1997) Estimating multiple breaks one at a time, *Econometric Theory*, **13**, 315–352.

Baillie, R., T. (1996) Long memory processes and fractional integration in econometrics, *Journal of Econometrics*, **73**, 5–59.

Baillie, R., T. Bollerslev and H. Mikkelsen (1996) Fractionally integrated generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, **74**(1), September, 3–30.

Banerjee, A., R. L. Lumsdaine and J. H. Stock (1990) Recursive and sequential tests for a unit root: Theory and international evidence, manuscript, Harvard University, Cambridge, MA.

Beran, J. (1992) Statistical Models for Data with long-range dependence (with discussion), *Statistical Science*, **4**, 404–427.

Berk, K. N. (1974) Consistent Autoregressive Spectral Estimates, *Annals of Statistics*, **2**, 489–502.

Basawa, I. V., A. K. Mallik, W. P. McCormick and R. L. Taylor (1991a) Bootstrapping unstable first order autoregressive process, *Annals of Statistics*, **19**, 1098–1101.

Basawa, I. V., A. K. Mallik, W. P. McCormick and R. L. Taylor (1991b) Bootstrapping test of significance and sequential bootstrap estimation for unstable first order autoregressive processes, *Communication in Statistics*, **20**, 1015–1026.

Beveridge, S. and R. Nelson (1981) A new approach to decomposition of time series in permanent and transitory components with particular attention to measurement of the 'business cycle', *Journal of Monetary Economics*, **7**, 151–174.

Bhargava, A. (1986) On the theory of testing for unit roots in observed time series, *Review of Economic Studies*, **53**, 369–384.

Blough, S. R. (1992) The relationship between power and level for generic unit root tests in finite samples, *Journal of Applied Econometrics*, **7**, 295–308.

Bose, A. (1988) Edgeworth correction by bootstrap in autoregressions, *Annals of Statistics*, **16**, 1709–1722.

Canjels, N. And M. Watson (1997) Estimating deterministic trends in the presence of serially correlated errors, *Review of Economics and Statistics*, **79**, 184–200.

Chan, N. H. and C. Z. Wei (1987) Asymptotic inference for nearly nonstationary AR(1) processes, *Annals of Statistics*, **15**, 1050–1063.

Chan N. H. & C. Z. Wei (1988) Limiting distributions of least squares estimates of unstable autoregressive processes, *Annals of Statistics*, **16**, 367–401.

Chao, J. and P. C. B. Phillips (1994) Bayesian model selection in partially non-stationary vector autoregressive processes with reduced rank structure, Yale University, mimeographed.

Chao, J. and P. C. B. Phillips (1996). An empirical Bayesian approach to cointegration rank selection and test of the present value model for stock prices. In J. C. Lee and A. Zellner (eds.), *Prediction, Forecasting and Modeling in Statistics and Econometrics*. Springer-Verlag, 1996.

Chao, J. and P. C. B. Phillips (1997) Model selection in partially nonstationary vector autoregressive processes with reduced rank structure, Cowles Foundation Discussion Paper No. 1155.

Cheung, Y. and M. Chinn (1996) Deterministic, stochastic and segmented trends in aggregate output, *Oxford Economic Papers*, **48**, 134–167.

Christiano, L. (1992) Searching for a break in GNP, *Journal of Business and Economic Statistics*, **10**, 237–250.

Davis, R. A. and W. T. M. Dunsmuir, (1996) Maximum likelihood estimation for MACD processes with a root on or near the unit circle, *Econometric Theory*, **12**, 1–29.

Dawid, A. P. (1984) The prequential approach, *Journal of Royal Statistical Society A*, **147**, 278–292.

DeJong, D. N. and C. H. Whiteman (1989a) Trends and random walks in macro-economic time series: A reconsideration based on the likelihood principle, Working Paper No. 1989–4, University of Iowa.

DeJong, D. N. and C. H. Whiteman (1989b) The temporal stability of dividends and stock prices: Evidence from the likelihood function, *American Economic Review*, **81**, 600–617.

DeJong, D. N. and C. H. Whiteman (1989c) Trends and cycles as unobserved components in real GNP: A Bayesian perspective, *Proceedings of the American Statistical Association*, 63–70.

DeJong, D. N., J. C. Nankervis, N. E. Savin and C. H. Whiteman (1992) The power problems of unit root tests for time series with autoregressive errors, *Journal of Econometrics*, **53**, 323–43.

Dickey, D. A. and W. A. Fuller (1979) Distribution of estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association*, **74**, 427–431.

Dickey, D. A. & W. A. Fuller (1981) Likelihood ratio tests for autoregressive time series with a unit root, *Econometrica*, **49**, 1057–1072.

Dickey, D. A., H. P. Hasza, and W. A. Fuller (1984) Testing for unit roots in seasonal time series, *JASA*, **79**, 355–367.

Diebold, F. and G. D. Rudebusch (1991) On the power of Dickey-Fuller tests against fractional alternatives, *Economic Letters*, **35**, 155–160.

Dufour, J. and M. L. King (1991) Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary and nonstationary AR(1) errors, *Journal of Econometrics*, **47**, 115–143.

Duffie, D. (1988) *Security Markets: Stochastic Models*. San Diego: Academic Press.

Durlauf, S. and P. C. B. Phillips (1988) Trends versus random walks in time series analysis, *Econometrca*, **56**, 656–676.

Elliot, G., T. J. Rothenberg and J. H. Stock (1996) Efficient tests for an autoregressive unit root, *Econometrca*, **64**, 813–836.

Engle, R. F. and C. W. J. Granger (1987) Cointegration and error correction: representation, estimation and testing, *Econometrica*, **55**, 251–276.

Evans, G. B. A. and N. E. Savin (1981) The calculation of the limiting distribution of the least squares estimator of the parameter in a random walk model, *Annals of Statistics*, **9**, 1114–1118.

Faust, J. (1996) Near observational equivalence and theoretical size problems with unit root tests, *Econometric Theory*, **12**, 724–731.

Ferretti, N. and J. Romo (1994) Unit root bootstrap tests for AR(1) models, working paper, Division of Economics, Universidad Carlos III de Madrid.
Freedman, D. A. (1984) On bootstrapping two-stage least squares estimates in stationary linear models, *Annals of Statistics*, **12**, 827–842.
van Giersbergen, N. P. A. (1995) Bootstrapping unit root tests in the AR(1) model with drift, manuscript, University of Amsterdam.
Gil-Alana, L. A. and P. M. Robinson (1997) Testing of unit root and other nonstationary hypothesis in macroeconomic time series, *Journal of Econometrics*, 241--268.
Grenander, U. and M. Rosenblatt (1957) *Statistical Analysis of Stationary Time Series*. New York: John Wiley.
Hall, R. E. (1978) Stochastic implications of the life cycle-permanent income hypothesis: Theory and evidence, *Journal of Political Economy*, **86**, 971–987.
Hall, P. and C. C. Heyde (1980) *Martingale Limit Theory and its Applications*. New York: Academic Press.
Hasan, M. N. and R. W. Koenker (1997) Robust rank tests of the unit root hypothesis, *Econometrica*, **65**, 133–161.
Heyde, C. C. and I. M. Johnstone (1979). On asymptotic posterior normality for stochastic processes, *Journal of the Royal Statistical Society* **41**, 184–189.
Hurvich, C. M., R. Deo and J. Brodsky (1998) The mean squared error of Geweke and Porter Hudak's estimator of the memory parameter of a long-memory time series, *Journal of Time Series Analysis*, **19**, 19–46.
Hylleberg, S., R. F. Engle, C. W. J. Granger and S. Yoo (1990) Seasonal integration and cointegration, *Journal of Econometrics*, **44**, 215–238.
Hylleberg, S. (ed.) (1992) *Modelling Seasonality*. Oxford: Oxford University Press.
Jeganathan, P. (1991) On the asymptotic behavior of least squares estimators in AR time series with roots near the unit circle, *Econometric Theory*, **7**, 269–306.
Jeganathan, P. (1995) Some aspects of asymptotic theory with applications to time series models, *Econometric Theory*, **11**, 818–887.
Kashyap, R. and K. Eom (1988) Estimation in long-memory time series models, *Journal of Time Series Analysis*, **9**, 35–41.
Kim, J-Y. (1994) Bayesian asymptotic theory in a time series model with a possible nonstationary process, *Econometric Theory*, **10**, 764–773.
King, M. L. (1988) Towards a theory of point optimal testing, *Econometric Reviews*, **6**, 169–218.
Kunsch, H. R. (1986) Discrimination between long range dependence and monotonic trends, *Journal of Applied Probability*, **23**, 1025–1030.
Kwiatkowski, D., P. C. B. Phillips, P. Schmidt and Y. Shin (1992) 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?', *Journal of Econometrics*, **54**, 159–178.
Lai, T. L. and C. Z. Wei (1982) Least squares estimation in stochastic regresssion models with applications to identification and control of dynamic systems, *Annals of Statistics*, **10**, 154–166.
Lai, T. L. and D. Seigmund (1983) Fixed accuracy estimation of an autoregressive parameter, *Annals of Statistics*, **11**, 478–485.
Lee, C. C. and P. C. B. Phillips (1994) An ARMA pre-whitened long run variance estimator, Yale University, mimeographed.
Leybourne S. J. and B. P. M. McCabe (1994) A consistent test for a unit root, *Journal of Business and Economic Statistics*, **12**, 157–166.
Lumsdaine, R. L., and D. H. Papell (1995) Multiple trend breaks and the unit root hypothesis, manuscript, Princeton University.
MacNeill, I. B. (1978) Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times, *Annals of Statistics*, **6**, 422–433.

McAleer M. and P. H. Franses (1998) Cointegration Analysis of Seasonal Time Series, in *Practical Issues in Cointegration Analysis*, Blackwell, Oxford (forthcoming).

Meese, R. A. and K. J. Singleton (1982) On unit roots and the empirical modeling of exchange rate, *Journal of Finance*, **37**, 1029–1035.

Nankervis, J. C. and Savin, N. E. (1994) The level and power of the bootstrap t-test in the AR(1) model with trend, manuscript, Department of Economics, University of Surrey and University of Iowa.

Nelson, C. R. and C. Plosser (1982) Trends and random walks in macro-economic time series: some evidence and implications, *Journal of Monetary Economics*, **10**, 139–162.

Nabeya, S. and K. Tanaka (1988) Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative, *Annals of Statistics*, **16**, 218–235.

Ng, S. and P. Perron (1995) Unit root tests in ARMA models with data dependent methods for the selection of the truncation lag, *Journal of the American Statistical Association*, **90**, 268–281.

Ng. S., and P. Perron (1998) Properties of the autoregressive spectral density estimator at frequency zero in ARIMA process, forthcoming, *Econometric Theory*.

Osborn, D. R., A. P. L. Chui, J. P. Smith, and C. R. Birchenhall (1988) Seasonality and the order of integration for consumption, *Oxford Bulletin of Economics and Statistics*, **50**, 361–368.

Osborn, D. R. and J. P. Smith (1989) The performance of periodic autoregressive models in forecasting seasonal U. K. consumption, *Journal of Business and Economic Statistics*, **7**, 117–128.

Ouliaris, S., J. Y. Park and P. C. B. Phillips (1989) Testing for a unit root in the presence of a maintained trend. In B. Raj (ed.), *Advances in Econometrics and Modelling*. Amsterdam: Kluwer, pp. 7–28.

Ouliaris, S. and P. C. B. Phillips (1994) *Coint 2.0*. Maple Valley, WA: Aptech Systems.

Park, J. Y. and J. Sung (1994) Testing for unit roots in models with structural change, *Econometric Theory*, **10**, 917–936.

Perron, P. (1988) Trend and random walks in macroeconomic time series: Further evidence from a new approach, *Journal of Economic Dynamics and Control*, **12**, 297–332.

Perron, P. (1989) The great crash, the oil price shock and the unit root hypothesis, *Econometrica*, **57**, 1361–1401.

Phillips, P. C. B. (1986) Understanding spurious regression in econometrics, *Journal of Econometrics*, **33**, 311–340.

Phillips, P. C. B. (1987a) Time series regression with a unit root, *Econometrica*, **55**, 277–302.

Phillips, P. C. B. (1987b) Towards a unified asymptotic theory of autoregression. *Biometrika*, **74**, 535–547.

Phillips, P. C. B. (1991) To criticize the critics: an objective Bayesian analysis of stochastic trends, *Journal of Applied Econometrics*, **6**, 333–364.

Phillips, P. C. B. (1995), Automated forecasts of Asia-Pacific economic activity, *Asia Pacific Economic Review*, **1**, 92–102.

Phillips, P. C. B. (1996a) Example of an optimal idempotent non orthogonal projector in $L_2[0, 1]$, research note, Yale University.

Phillips, P. C. B. (1996b) Econometric model determination, *Econometrica*, **64**, 763–812.

Phillips, P. C. B. (1998a) New tools for understanding spurious regression, *Econometrica* (forthcoming).

Phillips, P. C. B. (1998b) The invalidity of critical values of tests for unit roots against trend alternatives, in progress.

Phillips, P. C. B. (1998c) Gaussian semiparametric estimation of fractional integration — the nonstationary case, Yale University, under construction.

Phillips, P. C. B. (1998d) Impulse response and forecast error variance asymptotics in nonstationary VAR's, *Journal of Econometrics*, **83**, 21–56.

Phillips, P. C. B. And C. C. Lee (1996) Efficiency gains from quasi-differencing under nonstationarity. In P. M. Robinson and M. Rosenblatt (eds.), *Athens Conference on Applied Probability and Time Series: Essays in Memory of E.J. Hannan*, Springer-Verlag: New York.

Phillips, P. C. B. and S. Ouliaris (1990) Asymptotic properties of residual based tests for cointegration, *Econometrica*, **58**, 165–194.

Phillips, P. C. B. and P. Perron (1988) Testing for unit roots in time series regression, *Biometrika*, **75**, 335–346.

Phillips, P. C. B. and W. Ploberger (1994) Posterior odds testing for a unit root with data-based model selection, *Econometric Theory*, **10**, 774–808.

Phillips, P. C. B. and W. Ploberger (1996) An asymptotic theory of Bayesian inference for time series, *Econometrica*, **64**, 381–412.

Phillips, P. C. B. and V. Solo (1992) Asymptotics for linear processes, *Annals of Statistics*, **20**, 971–1001.

Porter-Hudak, Susan (1990) An application of the seasonal fractionally differenced model to the monetary aggregates, *Journal of the American Statistical Association*, **85**, 338–344.

Rissanen, J. (1978) Modeling by shortest data description, *Automatica*, **14**, 465–471

Rissanen, J. (1986) Stochastic complexity and modeling, *Annals of Statistics*, **14**, 1080–1100

Robinson, P. (1994a) Time series with strong dependence. In C. Sims (ed.), *Advances in Econometrics*, Sixth World Congress. Cambridge: Cambridge University Press.

Robinson, P. M. (1994b) Efficient tests of nonstationary hypothesis, *Journal of the American Statistical Association*, **89**, 1420–1437.

Robinson, P. M. (1995a) Log periodogram regression of time series with long range dependence, *Annals of Statistics*, **23**, 1048–1072.

Robinson, P. (1995b) 'Gaussian semiparametric estimation of long range dependence', *Annals of Statistics*, **23**, 1630–1661.

Said, S. E. and D. A. Dickey (1984) Testing for unit roots in autoregressive-moving average models of unknown order, *Biometrika*, **71**, 599–608.

Saikkonen, P. and R. Luukkonen (1993) Testing for a moving average unit root in autoregressive integrated moving average models, *Journal of the American Statistical Association*, **88**, 596–601.

Sargan, J. D. and A. Bhargava (1983) Testing residuals from least squares regression for being generated by the Gaussian random walk, *Econometrica*, **51**, 153–174.

Schmidt, P. and P. C. B. Phillips (1992) Testing for a unit root in the presence of deterministic trends, *Oxford Bulletin of Economics and Statistics*, **54**, 257–288.

Schotman, P. and H. K. van Dijk (1991) A Bayesian analysis of the unit root in real exchange rates, *Journal of Econometrics*, **49**, 195–238.

Schwarz, G. (1978) Estimating the dimension of a model, *Annals of Statistics*, **6**, 461–464.

Schwert, G. W. (1989) Tests for unit roots: A Monte Carlo investigation, *Journal of Business and Economic Statistics*, 7, 147–159.

Sims, C. A. (1990) Asymptotic behaviour of the likelihood function in an autoregression with a unit root, mimeographed, Yale University.

Sims, C., J. Stock and M. Watson (1990) Inference in linear time series models with some unit roots, *Econometrica*, **58**, 113–144.

Solo, V. (1984) The order of differencing in ARIMA models, *Journal of the American Statistical Association*, **79**, 916–921.

Stock, J. H. (1991) Confidence intervals for the largest autoregressive root in U.S. economic time series, *Journal of Monetary Economics*, **28**(3), 435–460.

Stock, J. H. (1994) Deciding between I(1) and I(0), *Journal of Econometrics*, 63, 105–131.

Stock, J. H. (1995) Unit roots, structural breaks and trends. In R. F. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Vol. 4, 2739–2841, Amsterdam: North-Holland.

Tanaka, K. (1999) The nonstationary fractional unit root, *Econometric Theory*, (forth coming).

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

Xiao, Z. and P. C. B. Phillips (1997) An ADF coefficient test for ARMA models with unknown orders, Cowles Foundation Discussion Paper No. 1161.

Young, C. H. (1974) 'Asymptotic behaviour of trigonometric series, Chinese University of Hong Kong.

Zivot, E. and D. W. K. Andrews (1992) Further evidence on the great crash, the oil price shock, and the unit root hypothesis, *Journal of Business and Economic Statistics*, **10**, 251–270.

Zivot, E. and P. C. B. Phillips (1994) A Bayesian analysis of trend determination in economic time series. *Econometric Reviews*, **13**, 291–336.

Annotated bibliography on unit root nonstationarity — textbooks, review articles and computer software

Abadir, K. (1999) *Mathematics of Unit Root Econometrics*, New York: John Wiley. (Forthcoming introduction to the mathematics of unit root distribution theory).

Banerjee, A., J. Dolado, J. W. Galbraith and D. F. Hendry (1992) *Cointegration, Error Correction, and the Econometric Analysis of Non-Stationary Time Series*. Oxford: Oxford University Press. (A textbook introduction to unit roots and cointegration that is designed to be accessible to graduate students of economics.)

Campbell, J. Y. and P. Perron (1991) Pitfalls and opportunities: What macroeconomists should know about unit roots *NBER Macroeconomics Annual*, 141–200. (A survey of research on unit roots in the 1980's that is intended for applied researchers.)

Davidson, J. (1994) *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford University Press. (An introduction to asymptotic theory for econometricians, including functional limit theory and stochastic integration with an emphasis on mixing sequences and functions of mixing sequences).

Dhrymes, P. (1998) *Time Series, Unit Roots and Cointegration*. San Diego: Academic Press. (A rigorous treatment of time series methods that deals with unit root theory and cointegration).

Diebold, F. X. and M. Nerlove (1990) Unit roots in economic time series, *Advances in Econometrics*, **8**, 3–70. (A survey of unit root literature in the 1980s.)

Dolado, J., T. Jenkinson and S. Sosvilla-Rivero (1990) Cointegration and unit roots, *J. Economic Surveys*, **4**, 249–273. (A general survey of literature in the field of unit roots and cointegration in the 1980s.)

Econometric Theory, **10**(3&4) (1994) (Symposium of recent work in the field of Bayesian unit roots.)

Econometric Reviews (1994) (Two special issues on unit roots and cointegration reporting recent research.)

Franses, P. H. (1996) *Periodicity and Stochastic Trends in Economic Time Series*, Oxford University Press, Oxford. (A general reference on periodic models for seasonally observed time series with stochastic trends)

Fuller, W. A. (1976) *Introduction to Statistical Time Series*. New York: Wiley. (Accessible technical introduction to time series. The first textbook treatment of unit roots. Contains the original tabulations of the critical values of the Dickey-Fuller tests.)

Fuller, W. A. (1996) *Introduction to Statistical Time Series*. New York: Wiley. (2nd

Edition of Fuller, 1976, expanded to include some coverage of systems with several unit roots)

Hamilton, J. D. (1994) *Time Series Analysis*. Princeton: Princeton University Press. (A textbook treatment of time series with coverage of unit root theory and cointegration with examples and applications. Designed mainly for graduate students in economics.)

Hatanaka, M. (1996) *Time-Series Based Econometrics*. Oxford: Oxford University Press. (Textbook of univariate tests for unit roots and cointegration analysis, which relates the econometric methods to economic ideas and empirical applications).

Journal of Applied Econometrics, **6**(4) (1991) (A special issue on Bayesian unit root models with a focus on the formulation of priors in stationary and nonstationary time series models.)

Journal of Business and Economic Statistics, **10** (1992) (A special issue devoted to structural change models and unit roots with broken trends.)

Johansen, S. (1995) *Likelihood Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press. (Textbook coverage of Gaussian maximum likelihood estimation and inference in vector autoregressive models with reduced rank structure, dealing with sequential tests for cointegrating rank and restrictions on the cointegrating space.)

Ouliaris, S. and P. C. B. Phillips (1993) *COINT 2.0: GAUSS Procedures for Cointegrating Regression*. Aptech Systems, Washington. (A suite of GAUSS computer programs for nonstationary time series and model selection, which provides routines for all of the major unit root tests and has built in kernel procedures for long run variance estimation and spectral density estimation.)

Oxford Bulletin of Economics and Statistics, **48** and **54** (1986 and 1992) (Two special issues of empirical and theoretical articles on cointegration and unit roots.)

Phillips, P. C. B. (1988) Multiple Regression with Integrated Time Series, *Contemporary Mathematics*, **80**, 79–105. (A detailed review article that covers functional limit theory, weak convergence to stochastic integrals, the asymptotic theory of unit roots, regression with integrated processes, spurious regression, cointegration, and local-to-unity asymptotics.)

Phillips, P. C. B. (1992) Unit roots, *New Palgrave Dictionary of Money and Finance*. London: Macmillan. (A short introduction to unit root testing and unit root models in economics and finance, written for economists.)

Phillips, P. C. B. (1995) Cointegration and unit roots: Recent books and themes for the future, *Journal of Applied Econometrics*, **10**, 87–94. (Reviews recent themes in the literature and points to directions of future research.)

Phillips, P. C. B. (1997) Unit Root Tests, *Encyclopedia of Statistical Science, Update*, Volume 1, pp. 531–542. (Review of unit root tests and overview of the field with an empirical illustration).

Reinsel, G. (1993) *Elements of Multivariate Time Series Analysis*. New York: Springer-Verlag. (A modern treatment of multivariate time series with some attention to VAR models with reduced rank structure and cointegration.)

Stock, J. H. and M. W. Watson (1988) Variable Trends in Economic Time Series, *Journal of Economic Perspectives*, **2**(3), 147–174. (Introductory outline of stochastic trends and unit roots and cointegration for economists.)

Tanaka, K. (1996) *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*. New York: Wiley. (Textbook coverage at an advanced level of the limit distribution theory for unit root models and small cointegration systems, with a focus on the Fredholm approach to characterising the distribution of functionals of stochastic processes like Brownian motion.)