Some Nonparametric Tests for Unit Roots and Cointegration

Jörg Breitung
Humboldt University Berlin
Institute for Statistics and Econometrics
Spandauer Strasse 1
D-10178 Berlin, Germany

May 3, 1999

Abstract

Following Bierens (1997a,b) and Vogelsang (1998a,b), unit root tests can be constructed which are asymptotically invariant to parameters involved by the short run dynamics of the process. Such an approach is called nonparametric by Bierens (1997b) and can be used to test a wide range of nonlinear models. We consider three different versions of such a test. However, simulation results suggest that only the variance ratio statistic is able to compete with the traditional augmented Dickey-Fuller test. A straightforward generalization of the variance ratio statistic is suggested, which can be used to test the cointegration rank in the spirit of Johansen (1988).

* The research for this paper was carried out within Sonderforschungsbereich 373 at the Humboldt University Berlin and the Training and Mobility of Researchers Programme of the European Commission (contract No. ERBFMRXCT980213). I thank Uwe Hassler and Rolf Tschernig for valuable comments on an earlier version.

1 Introduction

In recent papers by Bierens (1997a,b) and Vogelsang (1998a,b) it was observed that it is possible to construct test statistics that asymptotically do not depend on parameters involved by the short run dynamics of the process. Accordingly, it is not necessary to estimate the nuisance parameters such as the coefficients for the lagged differences in a Dickey-Fuller regression or the "long-run variance" (2π times the spectral density at frequency zero) by using a kernel estimate as in Phillips and Perron (1988). Such an approach is called "model free" in Bierens (1997a) and "nonparametric" in Bierens (1997b). Albeit both terms may be somewhat misleading, we follow Bierens (1997b) and use the term "nonparametric". In fact, asymptotically the tests only involve the parameter under test and it is difficult to think of any test, which is "less parametric".

The idea behind this approach can easily be explained as follows. Let $\{y_t\}_{t=1}^{T}$ be an integrated time series such that under the null hypothesis $\Delta y_t = y_t - y_{t-1}$ is stationary with $E(y_t) = 0$. It is known (e.g. Phillips, 1987) that under suitable conditions

$$T^{-1/2}\sum_{t=2}^{T}\Delta y_t \;\Rightarrow\; \sigma W(1)$$

$$T^{-2}\sum_{t=1}^{T}y_t^2 \;\Rightarrow\; \sigma^2\int_0^1 W(r)^2 dr \ ,$$

where, as usual, the symbol $\Rightarrow$ signifies weak convergence with respect to the associated probability measure, $W(r)$ represents a standard Brownian motion and σ^2 is the long-run variance defined as $\sigma^2 = \sum\limits_{j=-\infty}^{\infty}\gamma_j$, where $\gamma_j = E(\Delta y_t \Delta y_{t+j})$. These results suggest to consider the statistic

$$\psi_T = \frac{T^{-1}\left(\sum\limits_{t=2}^{T}\Delta y_t\right)^2}{T^{-2}\sum\limits_{t=1}^{T}y_t^2} \;\Rightarrow\; \frac{W(1)^2}{\int_0^1 W(r)^2 dr} \ , \qquad (1)$$

which does not depend on the nuisance parameter σ^2 as $T \to \infty$. Unfortunately, a test based on ψ_T is inconsistent. The reason is that under the alternative of a stationary process, the numerator and denominator are of the same order of magnitude so that ψ_T is $O_p(1)$ under the alternative as well. Bierens (1997a) resolves this problem by using the squares of the weighted sum

$$T^{-1/2}\sum_{t=2}^{T} g(t/T)\Delta y_t \;\Rightarrow\; \sigma g(1)W(1) - \sigma\int_0^1 \nabla g(r)dW(r) \qquad (2)$$

as the numerator in (1), where $\nabla g(r)$ denotes the derivative of $g(r)$.[1] For some appropriate weight function $g(r)$ it can be shown that the weighted sum is of the same order of magnitude under both the null and the alternative hypothesis.

In this paper a similar idea is adopted. However, instead of using weighted sums in the numerator of the test statistic we follow Vogelsang (1998a,b) and use functionals on the partial sum $Y_t = \sum_{j=1}^{t} y_j$. The advantage of this approach is that no weights are needed to make the test consistent. Furthermore, it turns out that our tests are more powerful than (the stylized version of) Bierens' (1997a) test and may even outperform the augmented Dickey-Fuller test.

The plan of the paper is as follows. In Section 2, a general framework is suggested which allows for a wide range of nonlinear processes generating the transitory component u_t. The power of Bierens' tests is considered in Section 3. In Section 4, regression and variance ratio statistics based on partial sums are proposed. The inclusion of deterministic terms is considered in Section 5 and Section 6 generalizes the variance ratio statistic to cointegrated systems. Section 7 presents the results of a Monte Carlo comparison of the different test statistics and Section 8 concludes.

[1] In fact, Bierens' (1997a) test is more complicated because he uses vector weights and extra terms to accommodate a nonlinear mean function. However, for our purpose it is sufficient to consider a simplified version of his test given by (2).

2 Basic Assumptions

Let $\{x_t\}_1^T$ be an observed time series with $D_t = E(x_t)$ and define $y_t = x_t - D_t$. Under the null hypothesis the following assumption applies to y_t:

Assumption 1: *There exists a decomposition $y_t = \xi_t + u_t$ with the properties:*

$$(i) \qquad T^{-1/2}\xi_{[rT]} \;\Rightarrow\; \sigma W(r),$$
$$(ii) \qquad T^{-2}\sum_{t=1}^{T} u_t^2 = o_p(1)$$

for some constant σ and $[a]$ denotes the largest integer smaller than a.

For a linear process $\Delta y_t = \sum_{j=0}^{\infty} \gamma_j \varepsilon_{t-j}$ with $\gamma_0 = 1$, $E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) = \sigma_\varepsilon^2 < \infty$ the assumption is satisfied whenever the process admits a Beveridge-Nelson decomposition, i.e., if $\sum_{j=0}^{\infty} j^2\gamma_j^2 < \infty$ (cf Phillips and Solo 1992).

For nonlinear processes, a Beveridge-Nelson (BN) type of decomposition may be constructed as follows. Let

$$\begin{aligned} E(y_{t+h}|y_t, y_{t-1}, \ldots) &= E(y_{t+h}|\varepsilon_t, y_{t-1}, y_{t-2}, \ldots) \\ &= E(y_{t+h}|y_{t-1}, y_{t-2}, \ldots) + E(y_{t+h}|\varepsilon_t) \ , \end{aligned}$$

where $\varepsilon_t = y_t - E(y_t|y_{t-1}, y_{t-2}, \ldots)$. Accordingly we obtain

$$\hat{y}_{t,h} = \hat{y}_{t-1,h+1} + v_t \qquad (3)$$

where $\hat{y}_{t,h}$ denotes the prediction of y_{t+h} based on the information available at t and $v_t = \pi(\varepsilon_t)$ is a function of the innovation. Note that the well known "updating formula" for linear stationary processes (e.g. Granger and Newbold 1986, p. 131) is a special case of (3), by inserting $v_t = E(y_{t+h}|\varepsilon_t) = \beta_h\varepsilon_t$, where β_h is the coefficient at lag h in the moving average representation of y_t.

If the prediction converges with increasing lag horizon, i.e.,

$$\lim_{h\to\infty} \hat{y}_{t-1,h+1} - \hat{y}_{t-1,h} = 0$$

we have

$$\hat{y}_{t,\infty} = \hat{y}_{t-1,\infty} + v_t$$

and it is seen that $\xi_t = \hat{y}_{t,\infty}$ (the permanent component according to BN (1981)) can be represented as a random walk process with v_t as increments. Whenever v_t satisfies the requirements of functional central limit theorem (e.g. Herrndorf 1984) then ξ_t converges to a Wiener process.

To illustrate the use of the BN decomposition for nonlinear processes consider the bilinear process

$$\Delta y_t = \phi(\varepsilon_{t-1}\Delta y_{t-1} - \sigma^2) + \varepsilon_t$$

where ϵ_t is a white noise process with $E(\varepsilon_t^2) = \sigma^2$. Since $E(\Delta y_{t+1}|y_t, y_{t-1}, \ldots) = \phi(\varepsilon_t \Delta y_t - \sigma^2)$ and $E(\Delta y_{t+j}|y_t, y_{t-1}, \ldots) = 0$ for $j > 1$, the permanent component results as

$$\xi_t = y_t + \phi(\varepsilon_t \Delta y_t - \sigma^2)$$

and the increment of the random walk component is given by

$$\begin{aligned} v_t &= \Delta y_t + \phi(\varepsilon_t \Delta y_t - \varepsilon_{t-1}\Delta y_{t-1}) \\ &= \varepsilon_t + \phi \varepsilon_t \Delta y_t \ . \end{aligned}$$

It turns out that $E(v_t v_{t-j}) = 0$ for $j \neq 0$ and if $E|\varepsilon_t|^{4+\delta} < \infty$ for some $\delta > 0$, the component ξ_t satisfies Assumption 1 (i). Similarly, it can be seen that the "transitory component" $u_t = y_t - \xi_t = -\phi(\varepsilon_t \Delta y_t - \sigma^2)$ satisfies Assumption 1 (ii).

Finally, it may be interesting to note that Assumption 1 allows u_t to be fractionally integregated with $(1-L)^d u_t = \varepsilon_t$ where L is the lag operator, d is a real number and ε_t is white noise. From Sowell (1990, Theorem 1) it follows that for $1/2 < d < 3/2$ the variance of u_t is $O(T^{2d-1})$ and, thus, Assumption 1 (ii) is satisfied for $d < 1$. In such situations, the augmented Dickey-Fuller test is expected to have poor power, because a high augmentation lag is needed to account for the long memory of the errors.

3 Bierens' Approach

In this section we consider a Bierens type of a nonparametric test statistic. As already mentioned, the statistic we consider is a "stylized version" of the test suggested in Bierens (1997a). Since we neglect a (possibly nonlinear) time trend and consider a scalar weight function, the test statistic simplifies to

$$\tau_T = \frac{T\left(\sum_{t=1}^{T} g(t/T)\Delta y_t\right)^2}{\sum_{t=1}^{T} y_t^2} \,. \tag{4}$$

Bierens (1997a) construct the weights using Chebishev time polynomials but any other differentiable weight function may be used as well. It is interesting to consider the effects of the weight function on the the power of the test. The following proposition characterizes the asymptotic behavior of the test statistic under the alternative of a stationary process.

Proposition 1: *Let* $y_t = \sum_{j=0}^{\infty} \gamma_j \varepsilon_t$ *be stationary and ergodic,* $E(\varepsilon_t^2) = \sigma_\varepsilon^2$ *and the roots of the polynomial* $\gamma(z) = \gamma_0 + \gamma_1 z + \gamma_2 z^2 + \cdots$ *are all outside the complex unit circle. For* $T \to \infty$ *we have*

$$T^{-1}\tau_T / v_\tau \;\Rightarrow\; \chi^2$$

where

$$v_\tau = 1 + \sum_{j=1}^{\infty} 2\bar{c}_j \delta_j, \qquad \delta_j = E(\Delta y_t \Delta y_{t+j})/E(y_t^2),$$

$c_{j,T} = T^{-1}\sum_{t=1}^{T} g(t/T) g[(t+j)/T]$, *and* $\bar{c}_j = \lim_{T\to\infty} c_{j,T}$. and χ^2 is a χ^2 distributed random variable with one degree of freedom.

Proof: From the central limit theorem for stationary processes (Hall and Heyde, 1980) it follows that

$$T^{-1/2}\sum_{t=1}^{T} g(t/T)\Delta y_t \;\Rightarrow\; N(0, \sigma_g^2) \ ,$$

where

$$\sigma_g^2 = \sum_{j=-\infty}^{\infty} \bar{c}_j E(\Delta y_t \Delta y_{t+j}).$$

Furthermore, $T^{-1}\sum y_t^2$ converges in probability to $E(y_t^2)$, so that $\tau_T/v_\tau \Rightarrow \chi^2$, where $v_\tau = \sigma_g^2/E(y_t^2)$. ■

This proposition shows that the power of the test crucially depends on the weight function. Thus, it is important to specify the weight function carefully. For illustration consider the trigonometric weights

$$g_k(t/T) = \cos(\omega_k t/T),$$

where $\omega_k = k \cdot 2\pi$, $k = 1, 2, \ldots$. Such a weight function is also considered in Bierens (1997b). The main difference between $g_k(t/T)$ and the Chebychev polynomial used in Bierens (1997a) is that the Chebychev polynomial introduces a phase shift. However, this does not have any effect on our discussion.

A second order Taylor expansion gives

$$\cos\left(\omega_k \frac{t+j}{T}\right) \simeq \cos(\omega_k t/T) + \sin(\omega_k t/T)\frac{j\omega_k}{T} - \cos(\omega_k t/T)\left(\frac{j\omega_k}{T}\right)^2$$

and, thus,

$$2c_{j,T} = 1 - \frac{(j\omega_k)^2}{T} .$$

If y_t is white noise, we have $\delta_1 = -1$ and $\delta_j = 0$ for $j \geq 2$. Accordingly, to achieve a good power of the test, $c_{1,T}$ should therefore be as small as possible, that is, a high frequency should be used for the trigonometric weights. On the other hand, if Δy_t is positively correlated, a low frequency is more appropriate. This example demonstrates, that there is no uniformly optimal weight function for the test and it is difficult to specify the weight function without an idea about the autocorrelation function of the series (see also Tschernig 1997).

Another problem is that the frequency of the trigonometric weight function must be low relative to the sample size. Assume that the frequency grows with the sample size such that $k = T/(2q)$ and, therefore, $g_q(t/T) =$

$\cos(\pi t/q)$, where $q = 1, 2, \ldots$. For the maximal frequency $q = 1$ the weight function flips between the values 1 and -1. From the above reasoning we expect that setting $q = 1$ yields a test with optimal power against a white noise series. However, as shown by the following proposition, the asymptotic theory for such a test is different.

Proposition 2: *Let $g_q(t/T) = \cos(\pi t/q)$, where $q < \infty$ and assume that y_t obeys Assumption 1. Then, as $T \to \infty$ we have*

$$T^{-1} \sum_{t=2}^{T} \cos(t\pi/q)\Delta y_t \;\Rightarrow\; \pi\sqrt{2} f_{\Delta y}(\pi/q) W(1) \;, \tag{5}$$

where $f_{\Delta y}(\pi/q)$ denotes the spectral density of Δy_t at frequency π/q.

PROOF: From eq. (32) of Phillips & Solo (1992) and Assumption 1 we have

$$\begin{aligned}
\sum_{t=2}^{T} \cos(t\pi/q)\Delta y_t &= Re\left[\sum_{t=2}^{T} e^{it\pi/q}\Delta y_t\right] \\
&= Re\left[\gamma(e^{i\pi/q})\sum_{t=2}^{T} e^{i\pi/q}\varepsilon_t + O_p(1)\right] \\
&= \sum_{t=1}^{T}\left\{Re[\gamma(e^{i\pi/q})]\cos(t\pi/q) - Im[\gamma(e^{i\pi/q})]\sin(t\pi/q)\right\} + O_p(1)
\end{aligned}$$

where $Re(a)$ and $Im(a)$ denote the real and imaginary part of the complex number a.

The phase of the filter $\gamma(L)$ is defined as

$$\phi_\gamma(\omega) = \tan^{-1}\left\{-Im[\gamma(e^{-i\omega})]/Re[\gamma(e^{-i\omega})]\right\}.$$

Furthermore,

$$a\cos j\omega + b\sin j\omega = \cos[j\omega + \tan^{-1}(-b/a)].$$

This gives

$$\sum_{t=1}^{T} \cos(t\pi/q)\Delta y_t = \sum_{t=1}^{T} \cos[t\pi/q + \phi_\gamma(\pi/q)]\varepsilon_t + O_p(1).$$

Using the results of Chan and Wei (1988) we have

$$T^{-1/2}\sum_{t=1}^{T}\cos[t\pi/q+\phi_\gamma(\pi/q)]\varepsilon_t \;\Rightarrow\; \frac{2\pi f_{\Delta y}(\pi/q)}{\sqrt{2}}W(1)$$

which yields the desired result. ■

Note that $\lim\limits_{T\to\infty} T^{-1}\sum\limits_{t=1}^{T}\cos(t\pi/q)^2 = 1/2$ so that if Δy_t is white noise the expression (5) is normally distributed with variance $\sigma_\varepsilon^2/2$. Furthermore, it follows from Proposition 2 that, in general, the limiting distribution of the test statistic τ_T depends on $f_{\Delta y}(\pi/q)$.

To summarize, Bierens' (1997a) asymptotic theory is valid for the weight function $g(t/T) = \cos(k2\pi t/T)$. However, if $k \to \infty$ at the rate T, then a different asymptotic theory applies implying that the test statistic depends on parameters characterizing the short run dynamics of Δy_t. We therefore expect that a trigonometric weight function with a high frequencies results in a size bias and, thus, there is a trade off between size bias and power of the test.

4 Nonparametric Unit Root Tests Based on Partial Sums

Assume that y_t is $I(1)$ as defined in Assumption 1 and consider the (unbalanced) regression

$$Y_t = \alpha y_t + e_t \ ,$$

where $Y_t = \sum\limits_{j=1}^{t} y_j$ denotes the partial sum of y_t. The least-squares estimator of α is

$$\widehat{\alpha}_T = \frac{\sum\limits_{t=1}^{T} Y_t y_t}{\sum\limits_{t=1}^{T} y_t^2} \tag{6}$$

and using standard results of the asymptotic theory for unit root processes (e.g. Hamilton 1994) the asymptotic distribution is given by

$$T^{-1}\widehat{\alpha}_T \;\Rightarrow\; \frac{\int_0^1 [\int_0^r W(s)ds] W(r)dr}{\int_0^1 W(r)^2 dr} \;,$$

which does not depend on the nuisance parameter σ^2.

Second, we construct the "variance ratio statistic"

$$\varrho_T = \frac{\sum_{t=1}^{T} Y_t^2}{\sum_{t=1}^{T} y_t^2} \tag{7}$$

which is asymptotically distributed as

$$T^{-2}\varrho_T \;\Rightarrow\; \frac{\int_0^1 [\int_0^r W(s)ds]^2 dr}{\int_0^1 W(r)^2 dr} \tag{8}$$

and, thus, does not depend on nuisance parameters as well.

The following proposition shows that these tests are consistent against stationary alternatives.

Proposition 3: *Let y_t be stationary with Wold representation $y_t = \sum_{j=0}^{\infty} \gamma_j \varepsilon_{t-j}$, where $\gamma_0 = 1$, $\sum_{j=0}^{\infty} \gamma_j^2 < \infty$, and ε_t is white noise with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) = \sigma_\varepsilon^2$. Under this alternative we have as $T \to \infty$*

$$\widetilde{\alpha}_T \;\Rightarrow\; \frac{\bar{\gamma}^2 \sigma_\varepsilon^2 \chi^2 - \sigma_y^2}{2\sigma_y^2}$$

$$T^{-1}\varrho_T \;\Rightarrow\; \frac{\bar{\gamma}^2 \sigma_\varepsilon^2 \int_0^1 W(r)^2 dr}{\sigma_y^2},$$

where $\bar{\gamma} = \sum_{j=0}^{\infty} \gamma_j$, $\sigma_y^2 = \sum_{j=0}^{\infty} \gamma_j^2 \sigma_\varepsilon^2$ and χ^2 denotes a χ^2-distributed random variable with one degree of freedom. Consequently, $Pr\{T^{-1}\widetilde{\alpha}_T < c_\alpha\} \to 1$ and $Pr\{\varrho_T < c_\varrho\} \to 1$ for $c_\alpha > 0$ and $c_\varrho > 0$.

PROOF: Using standard results from the asymptotic theory for unit root processes (e.g. Hamilton 1994) we get

$$T^{-1}\sum_{t=1}^{T} Y_t y_t \;\Rightarrow\; \bar{\gamma}^2\sigma_\varepsilon^2 \int_0^1 W(r)dW(r) + (\bar{\gamma}^2\sigma_\varepsilon^2 + \sigma_y^2)/2 \; ,$$

where we use

$$\lim_{T\to\infty} E(T^{-1}Y_T^2) = \left(\sum_{j=0}^{\infty}\gamma_j\right)^2 \sigma_\varepsilon^2 = \bar{\gamma}^2\sigma_\varepsilon^2$$

and

$$\lim_{T\to\infty} E\left(T^{-1}\sum_{t=1}^{T} Y_t y_t\right) = E\left(\sum_{j=0}^{\infty} y_1 y_{1+j}\right) = (\bar{\gamma}^2\sigma_\varepsilon^2 + \sigma_y^2)/2.$$

It is well known that $[2\int_0^1 W(r)dW(r) + 1]$ is χ^2 distributed.

Furthermore we have

$$T^{-2}\sum Y_t^2 \;\Rightarrow\; \bar{\gamma}^2\sigma_\varepsilon^2 \int_0^1 W(r)^2 dr.$$

Using these results and $T^{-1}\sum_{t=1}^{T} y_t^2 \xrightarrow{p} \sigma_y^2$, the limiting distributions follow immediately. ■

It is interesting to compare these test statistics with Bierens' approach. The latter statistic uses a weighted sum as numerator that is of the same order of magnitude under the null and alternative hypothesis. In contrast we use functionals on the partial sum yielding a numerator that differs by a factor of $O_p(T^{-2})$ when moving from the null to the alternative hypothesis. This difference in the asymptotic properties dominates the difference in the denominator of the test statistic and a consistent test results.

5 Including Deterministic Terms

To accommodate processes with a nonzero mean we assume that the mean function $E(y_t) = D_t = c'd_t$ is a linear function of deterministic variables like a constant, time trend or dummy variables stacked in the $k \times 1$ vector d_t and c is a vector of unknown coefficients. In this case it is natural to remove the mean of the time series by using the residuals from the regression function $y_t = \hat{c}' d_t + \hat{u}_t$, where $\hat{c}$ denotes the least-squares estimator of c. The test statistic $\widetilde{\alpha}_T$ is obtained from a regression of $\widehat{Y}_t$ on u_t, where $\widehat{Y}_t = \sum_{i=1}^t \hat{u}_i$ denotes the partial sum of $\hat{u}_t$. Similarly, the variance ratio statistic $\widetilde{\varrho}_T$ is computed as $\widetilde{\varrho}_T = \sum \widehat{Y}_t^2 / \sum \hat{u}_t^2$.

Unfortunately, this approach fails for the statistic $\tilde{\alpha}_T$ if d_t contains a constant term. This is stated in the following proposition.

Proposition 4: *If an element of d_t is constant, then $\widetilde{\alpha}_T = -1/(2T)$.*

PROOF: We have

$$\sum_{t=1}^{T} (\widehat{Y}_t - \hat{u}_t)^2 = \sum_{t=1}^{T} \widehat{Y}_{t-1}^2 = \sum_{t=1}^{T} \widehat{Y}_t^2 - 2 \sum_{t=1}^{T} \widehat{Y}_t \hat{u}_t + \sum_{t=1}^{T} \hat{u}_t^2$$

so that

$$\sum_{t=1}^{T} \widehat{Y}_t \hat{u}_t = - \left(\widehat{Y}_T^2 + \sum_{t=1}^{T} \hat{u}_t^2 \right) /2 \; .$$

However, since in a regression with a constant $\sum u_t = 0$ so that $\widehat{Y}_T = 0$, and, thus, $\widetilde{\alpha}_T = -1/(2T)$. ∎

Therefore, in this case the series y_t must be adjusted for a constant mean by using a different approach such as subtracting the first observation. In other cases the test statistic can be computed by replacing y_t by the residuals $\hat{u}_t$. As usual the asymptotic distribution of the test statistic is affected by such data transformations. For example, if $d_t = 1$, then the asymptotic distribution of the variance ratio statistic is as in (8) but with Brownian bridges $W(r) - rW(1)$ instead of standard Brownian motions.

6 Testing the Cointegration Rank

The variance ratio statistic for a nonparametric unit root test can be generalized straightforwardly to test hypotheses on the cointegration rank in the spirit of Johansen (1988, 1991). As in Section 2, it is assumed that the process can be decomposed into a q-dimensional stochastic trend component ξ_t and a $(n-q)$-dimensional transitory component u_t.

Assumption 2: *There exists an invertible matrix $Q = [\gamma, \beta]$, where γ and β are linearly independent $n \times q$ and $n \times (n-q)$ matrices, respectively, with $0 \leq q < n$ such that*

$$
\begin{aligned}
Q' y_t = \begin{bmatrix} \gamma' y_t \\ \beta' y_t \end{bmatrix} &\equiv \begin{bmatrix} \xi_t \\ u_t \end{bmatrix} = z_t \\
T^{-1/2} \xi_{[aT]} &\Rightarrow W_q(a) \\
T^{-2} \sum_{t=1}^{T} u_t u_t' &= o_p(1) \ ,
\end{aligned}
$$

where $W_q(a)$ is a q–dimensional Brownian motion with unit covariance matrix.

Note that to allow for some general nonlinear processes generating ξ_{1t}, we do not assume that the "error correction term" u_t is stationary. Instead we assume that the trend component ξ_t is "variance dominating" in the sense that the variance of ξ_t diverges with a faster rate than u_t.

The dimension of the stochastic trend component ξ_t is related to the cointegration rank of a linear system by $q = n - r$, where r is the rank of the matrix Π in the so-called vector error correction representation

$$\Delta y_t = \Pi y_{t-1} + v_t, \tag{9}$$

and v_t is a stationary error vector. In a linear system, the hypothesis on the number of stochastic trends is equivalent to a hypothesis on the cointegration rank as in Johansen (1988). However, since we do not assume that the process is linear, the representation of the form (9) may not exist.

Our test statistic is based on the eigenvalues λ_j of the problem

$$|\lambda_j B_T - A_T| = 0 \ , \tag{10}$$

where

$$A_T = \sum_{t=1}^{T} y_t y_t' \ , \quad B_T = \sum_{t=1}^{T} Y_t Y_t'$$

and $Y_t = \sum_{j=1}^{t} y_j$ denotes the n-dimensional partial sum with respect to y_t. The eigenvalues of (10) are identical to the eigenvalues of the matrix $R_T = A_T B_T^{-1}$. For $n = 1$ the eigenvalue is identical to the statistic $1/\varrho_T$ and, thus, the test can be seen as a generalization of the variance ratio statistic to multivariate processes.

The eigenvalues of (10) are given by

$$\lambda_j = \frac{\eta_j' A_T \eta_j}{\eta_j' B_T \eta_j} \ , \tag{11}$$

where η_j is the eigenvector associated with the eigenvalue λ_j. If the vector η_j falls inside the space spanned by the columns of γ, then $\eta_j' A_T \eta_j$ is $O_p(T^2)$ and $\eta_j' B_T \eta_j$ is $O_p(T^4)$ so that the eigenvalue is $O_p(T^{-2})$. On the other hand, if the eigenvector η_j falls into the space spanned by the columns of β, it follows that $T^2\lambda_j$ tends to infinity, as $T \to \infty$. Therefore, the test statistic

$$\Lambda_q = T^2 \sum_{j=1}^{q} \lambda_j \tag{12}$$

has a nondegenerate limiting distribution, where $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ denote the eigenvalues of the matrix R_T. In contrast, if the number of stochastic trends is smaller than q, then Λ_q eventually diverges T to infinity. The following proposition gives the limiting null distribution for the test statistic Λ_q.

Proposition 5: *Assume that y_t admits a decomposition as in Assumption 2 with $0 < q \leq n$. Then, as $T \to \infty$*

$$\Lambda_q \ \Rightarrow \ tr \left\{ \int_0^1 W_q(a) W_q(a)' da \left[\int_0^1 V_q(a) V_q(a)' da \right]^{-1} \right\}$$

W_q is a q-dimensional standard Brownian motion and $V_q(a) = \int_0^a W_q(s)ds$.

PROOF: Let $Z_t = \sum_{j=1}^{t} z_j$ denote the partial sum with respect to $z_t = Q'y_t = [\xi_t', u_t']'$. Then, the eigenvalues of problem (10) also solves the problem

$$|\lambda_j D_T - C_T| = 0 \ ,$$

where

$$C_T = \sum_{t=1}^{T} z_t z_t' \ , \quad D_T = \sum_{t=1}^{T} Z_t Z_t' \ .$$

Partition the corresponding eigenvectors $\tilde{\eta}_j = [\tilde{\eta}_{1j}', \tilde{\eta}_{2j}']'$ such that $\tilde{\eta}_j' z_t = \tilde{\eta}_{1j}' \xi_{1t} + \tilde{\eta}_{2j}' u_{2t}$, and Z_t is partitioned accordingly. We normalize the eigenvectors as

$$\tilde{\eta}_1 = [\tilde{\eta}_{11}, \ldots, \tilde{\eta}_{1q}] = \begin{bmatrix} I_q \\ \Phi_T \end{bmatrix}$$

so that $\tilde{\eta}_{1j}' z_t = \xi_{jt} + \Phi_T' u_t$, where ξ_{jt} denotes the j-th component of the vector ξ_t. It follows that

$$\begin{aligned} \lambda_j &= \frac{\tilde{\eta}_j' C_T \tilde{\eta}_j}{\tilde{\eta}_j' D_T \tilde{\eta}_j} \\ &= \frac{\sum_{t=1}^{T} \xi_{jt}^2 + o_p(T^2)}{\sum_{t=1}^{T} Z_{jt}^2 + o_p(T^4)} \\ &= \frac{\sum_{t=1}^{T} \xi_{jt}^2}{\sum_{t=1}^{T} Z_{jt}^2} + o_p(1), \end{aligned}$$

where $Z_{jt} = \sum_{s=1}^{t} \xi_{js}$. As $T \to \infty$ we therefore have

$$T^2 \sum_{j=1}^{q} \lambda_j \ \Rightarrow \ tr \left\{ \int_0^1 W_q(a) W_q(a)' da \left[\int_0^1 V_q(a) V_q(a)' da \right]^{-1} \right\}$$

■

From this proposition it follows that the distribution of the q smallest eigenvalues of the problem (10) does not depend on nuisance parameters and, thus, we do not need to select the lag order of the VAR process as in Johansen's approach or the truncation lag as for the test of Quintos (1998).

7 Small Sample Properties

In this section we present the results of some Monte Carlo experiments. It is not intended to give a comprehensive account of the merits and drawbacks of our test relative to other unit root tests based on a parametric or nonparametric adjustment for short run dynamics. Rather, we try to give a rough idea of the relative performance of the tests, where the augmented Dickey-Fuller test is used as a benchmark.

For the univariate tests, the data is generated by the process

$$y_t = \phi y_{t-1} + \varepsilon_t - \beta\varepsilon_{t-1} \tag{13}$$

and $x_t = y_t + D_t$, where $\varepsilon_t \sim niid(0,1)$ and D_t is a constant or a linear trend. The sample size is $T = 200$. Under the null hypothesis $\phi = 1$ and $\beta < 1$. For $\beta \neq 0$ the model does not possess a finite AR representation and following Said and Dickey (1984) and Schwert (1989) we account for correlation by including, respectively, $p = 4$ and $p = 12$ lagged differences in the autoregression. The test is denoted by ADF(p).

For a stylized version of Bierens' (1997a) test we use a trigonometric weight function given by $g_k(t/T) = \cos(k2\pi t/T)$, where $k = 1, 4, 16, 32$. The respective test is labeled as $\tau_T(k)$. The critical values with respect to a significance level of 0.05 are obtained from 10.000 Monte Carlo runs of the model with $\phi = 1$ and $\beta = 0$. To adjust for the mean D_t, the test statistic is constructed using the residuals from a regression of y_t on a constant or a linear trend.

For the regression statistic $\widetilde{\alpha}_T$ we subtract the first observation to correct for a constant mean and in the case of a linear time trend we regress $(y_t -$

y_1) on t (without a constant) and form the statistic with the residuals of this regression. This modification is necessary to sidestep the difficulties mentioned in Section 5.

Finally, the variance ratio statistic is computed using the residuals from a regression of y_t on a constant or a linear trend. The respective test statistic is denoted by $\widetilde{\rho}_T$. Selected critical values for this test are presented in the Appendix.

Table 1 a) presents the empirical sizes computed as the relative rejection frequencies for $H_0 : \phi = 1$ and various values of β in a model with a constant mean. Since the critical values are computed from the same random draws, the empirical sizes are exact 0.05 for $\tau_T(k)$, $\widetilde{\alpha}_T$ and $\widetilde{\varrho}_T$. For low frequencies it turns out that the empirical sizes of the Bierens type tests are close to the nominal ones for all values of β. However, if the frequency increases to $k = 32$ the test shows a serious size bias for β different from zero. This is due to the fact that for high frequencies, the asymptotic distribution involves the parameter β (see section 2).

The test statistic $\widetilde{\alpha}_T$ is quite robust against different values of β and even for $\beta = 0.8$ only a small size bias is observed. The actual null distribution of the statistic $\widetilde{\varrho}_T$ is much more affected by a positive value of β. For $\beta = 0.5$ the size bias is moderate but for $\beta = 0.8$ the test is severely biased towards a rejection of the null hypothesis. A similar outcome is observed for the ADF(4), however, if the ADF test is augmented with 12 lagged differences, the empirical size is close to the nominal size for values up to $\beta = 0.8$.

The empirical powers of the test procedures for different values of ϕ are presented in Table 1 b). It turns out that using a trigonometric weight function with a low frequency yields a poor performance of the Bierens' type of test. As expected (see Section 3) the power of the test improves with an increasing frequency. However, since the actual size of the test increases as well, it is quite difficult to select an appropriate frequency. The regression test using $\widetilde{\alpha}_T$ performs as poor as the former test with a low frequency but the power of the variance ratio test $\widetilde{\varrho}_T$ is much better. For $\phi = 0.95$ the test is even slightly better than the ADF(4) test whereas for other values of ϕ, the power of the variance ratio test is larger than the ADF(12) statistic but smaller than the ADF(4) statistic. The results for a model with a linear time trend are qualitatively similar (see Table 2).

To investigate the properties of the nonparametric cointegration test we generate data according to the "canonical" process (Toda 1994)

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \begin{bmatrix} \phi_1 & 0 \\ 0 & \phi_2 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} - \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{bmatrix}, \qquad (14)$$

where $E(\varepsilon_{1t}^2) = E(\varepsilon_{2t}^2) = 1$ and $E(\varepsilon_{1t}\varepsilon_{2t}) = \theta$. To test the hypothesis $r = 1$, we let $\psi_1 = 1$ and $\psi_2 = 0.8$. Under the alternative we set $\psi_1 \in \{0.95, 0.9, 0.8, 0.5\}$. Furthermore, we let $\theta = 0$ and $\theta = 0.8$ to investigate the impact of the error correlation. The sample size is $T = 200$ and 10.000 samples are generated to compute the rejection frequencies of the tests.

For Johansen's LR trace test, the process is approximated by a VAR(p) process, where p is 4 and 12, respectively. The respective tests are denoted by LR(4) and LR(12). The nonparametric test statistic is Λ_q and the critical values are taken from Table A.2 in the appendix. First, consider the results for testing $H_0 : q = r = 1$. From empirical sizes it turns out that for $\theta = 0$, a VAR(4) model is not sufficient to approximate the infinite VAR process, whereas a VAR(12) approximation yields an accurate size. The nonparametric statistic Λ_q possesses a negligible size bias, only. The power of Λ_q is substantially smaller than the power of LR(4) but clearly higher than the power of LR(12). Similar results apply for the tests letting $\theta = 0.8$. However, the LR(12) statistic now possesses a moderate size bias, whereas Λ_q is nearly unbiased. Moreover, the power of Λ_q is closer to the (favorable) LR(4) statistic than in the case of $\theta = 0$.

We now turn to the test of $H_0 : r = 0$. Under the null hypothesis the difference of the variables are generated by a multivariate MA process. In this case, all three test statistic are substantially biased, where the size bias does not depend on the parameter θ. Although the sizes bias differs for the three test, the differences are moderate and some general conclusions with respect to the relative power of the tests can be drawn. For $\theta = 0$ and

ϕ_1 close to unity, the nonparametric test Λ_q is slightly more powerful than the LR(4) test, whereas for $\phi_1 = 0.8$ the power of LR(4) is slightly higher. Finally, the power of LR(12) is much smaller than the power of the other two tests. For $\theta = 0.8$ a different picture emerges. The relative power of Λ_q drops substantially and for ϕ_1 close to one, the power is even lower than the power of the LR(12) test. The results for a model with a linear time trend are qualitatively similar and are not presented for reasons of space.[2]

8 Concluding Remarks

Following Bierens (1997a,b) and Vogelsang (1998a,b), unit root tests can be constructed which, asymptotically, do not depend on parameters involved by the short run dynamics of the process. We have considered three different versions of such a test. Our simulation results suggest, however, that among these statistics only the variance ratio statistic is able to compete with the traditional augmented Dickey-Fuller test.

For practical applications of the tests, several points deserve attention. First, the invariance to the short run dynamics of the process is an asymptotic property and need not be encountered in small or moderate samples. In particular, if the variance of the transitory component is important relative to the variance of the random walk component, the size bias may be severe. Second, we have shown that under the alternative of a stationary process, the appropriately normalized test statistics converge to a random variable as T tends to infinity. On the other hand, the normalized Dickey-Fuller test converge to a constant under the null hypothesis and, therefore, the test generally has more favorable properties than the nonparametric counterparts. Finally, in many empirical applications it is not difficult to select an appropriate augmentation lag or the test statistic turns out to be quite robust against different lag orders or truncation lags for the Phillips-Perron type of tests. In these cases the nonparametric test statistics have nothing to offer

[2] The results for further values of ϕ_2 and θ as well as the results for the model with a time trend are available on request.

and the conventional unit root statistics are clearly preferable.

However, there are a number of situations, the nonparametric approach may be attractive. Since the short run component does not affect the asymptotic null distribution of the test statistic, the test is robust against deviations from the usual assumption of linear short run dynamics. Thus, whenever the sample size is large, there is reason to expect that the random walk component dominate the sampling behavior of the test statistic and the asymptotic theory provides a reliable approximation to the actual null distribution. If, in addition, a high augmentation lag is needed and the results depend sensitively on the number of lags included in the Dickey-Fuller regression, it may be useful to apply nonparametric tests.

References

Bierens, H.J. (1997a), Testing the Unit Root With Drift Hypothesis Against Nonlinear Trend Stationarity, With an Application to the US price Level and Interest Rate, Journal of Econometrics, 81, 29–64.

Bierens, H.J. (1997b), Cointegration Analysis, in: C. Heij, H. Schumacher, B. Hanzon, and C. Praagman (eds), System Dynamics in Economic and Financial Models, New York: John Wiley.

Chan, N.H. and C.Z. Wei (1988), Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes, Annals of Statistics, 16, 367–401.

Granger, C.W.J. and P. Newbold (1986), Forecasting Economic Time Series, 2nd ed., Orlando: Academic Press.

Hamilton, J.D. (1994), Time Series Analysis. Princeton: Princeton University Press.

Hall, P. and C.C. Heyde (1980), Martingale Limit Theory and Application, New York: Academic Press.

Hamilton, J.D. (1994), Time Series Analysis, Princeton: Princeton University Press.

Herrndorf, N. (1994), A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables, Annals of Probability, 12, 141–153.

Johansen, S. (1988), Statistical Analysis of Cointegration Vectors, Journal of Economic Dynamics and Control, 12, 231–254.

Johansen, S. (1991), Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models, Econometrica, 59, 1551–1580.

Phillips, P.C.B. (1987), Time Series Regression With a Unit Root, Econometrica 55, 277–301.

Phillips, P.C.B. and P. Perron (1988), Testing for a Unit Root in Time Series Regression, Biometrika 75, 335–346.

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

Quintos, C.E. (1998), Fully Modified Vector Autoregressive Inference in Partially Nonstationary Models, Journal of the American Statistical Association, 93, 783–795.

Said, S.E. and D.A. Dickey (1984), Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order, Biometrika, 71, 599–607.

Schwert, G.W. (1989), Tests for Unit Roots: A Monte Carlo Investigation, Journal of Business & Economic Statistics, 7, 147–159.

Sowell, F. (1990), The Fractional Unit Root Distribution, Econometrica, 58, 495–505.

Toda, H.Y. (1994), Finite Sample Properties of Likelihood Ratio Tests for Cointegrating Ranks When Linear Trends are Present, Review of Economics and Statistics, , 66–79.

Tsay, R.S. (1997), Unit-Root Tests With Threshold Innovations, University of Chicago, mimeo.

Tschernig, R. (1997), Comments to "Cointegration Analysis" by H.J. Bierens, in: C. Heij, H. Schumacher, B. Hanzon, and C. Praagman (eds), System Dynamics in Economic and Financial Models, New York: John Wiley, 239–241.

Vogelsang, T.J. (1998a), Trend Function Hypothesis Testing in the Presence of Serial Correlation, Econometrica, 66, 123–148.

Vogelsang, T.J. (1998b), Testing for the Shift in Mean Without Having to Estimate Serial-Correlation Parameters, Journal of Business & Economic Statistics, 16, 73–80.

Appendix: Critical values

The critical values are computed from the empirical distribution of 10.000 realizations of the limiting expressions of the test statistics, with random walk sequences instead of Brownian motions. Regarding the poor performance of the regression test based on α_T, we only present the critical values of the variance ratio statistic. Critical values for $T^{-1}\widetilde{\alpha}_T$ are available from the author on request.

Table A.1: Critical Values for $T^{-2}\varrho_T$

T	0.1	0.05	0.01
mean adjusted			
100	0.01435	0.01004	0.00551
250	0.01433	0.01003	0.00561
500	0.01473	0.01046	0.00536
trend adjusted			
100	0.00436	0.00342	0.00214
250	0.00442	0.00344	0.00223
500	0.00450	0.00355	0.00225

Note: The hypothesis of a unit root process is rejected if the test statistic falls below the respective critical values reported in this table.

Table A.2: Critical Values for Λ_q

$r = n - q$	0.1	0.05	0.01
mean adjusted			
1	67.89	95.60	185.0
2	261.0	329.9	505.8
3	627.8	741.1	1024
4	1200	1360	1702
5	2025	2255	2761
6	3177	3460	4045
7	4650	5049	5905
8	6565	7061	8032
trend adjusted			
1	222.4	281.1	443.6
2	596.2	713.3	976.1
3	1158	1330	1689
4	1972	2184	2699
5	3107	3429	4120
6	4572	4954	5780
7	6484	6984	8012
8	8830	9388	10714

Note: The hypothesis $r = r_0$ is rejected if the test statistic exceeds the respective critical values. The simulation are based on a sample size of $T = 500$.

Table 1: Rejection Frequencies for a Constant Mean

a) Empirical size				
test statistic	$\beta = -0.5$	$\beta = 0$	$\beta = 0.5$	$\beta = 0.8$
$\tau_T(1)$	0.050	0.050	0.042	0.020
$\tau_T(4)$	0.049	0.050	0.042	0.023
$\tau_T(16)$	0.048	0.050	0.076	0.197
$\tau_T(32)$	0.033	0.050	0.167	0.451
$\widetilde{\alpha}_T$	0.050	0.050	0.055	0.066
$\widehat{\varrho}_T$	0.047	0.050	0.072	0.223
ADF(4)	0.049	0.052	0.062	0.373
ADF(12)	0.058	0.060	0.055	0.074
b) Empirical power				
test statistic	$\phi = 0.95$	$\phi = 0.90$	$\phi = 0.80$	$\phi = 0.50$
$\tau_T(1)$	0.033	0.016	0.007	0.006
$\tau_T(4)$	0.179	0.212	0.170	0.052
$\tau_T(16)$	0.204	0.325	0.447	0.494
$\tau_T(32)$	0.211	0.335	0.483	0.638
$\widetilde{\alpha}_T$	0.169	0.172	0.154	0.106
$\widehat{\varrho}_T$	0.292	0.543	0.808	0.990
ADF(4)	0.262	0.680	0.982	1.000
ADF(12)	0.202	0.417	0.715	0.923

Note: The entries of the table display the rejection frequencies based on 10.000 replications of model (13), where D_T is constant. The sample size is $T = 200$ and the nominal size of the test ist 0.05. Since the critical values are computed from the same random draws, the empirical sizes are exact 0.05 for $\tau_T(k)$, $\widetilde{\alpha}_T$ and $\widetilde{\varrho}_T$. The sample size is $T = 200$ and the nominal size of the test ist 0.05.

Table 2: Rejection Frequencies for a Linear Trend

a) Empirical size				
test statistic	$\beta = -0.5$	$\beta = 0$	$\beta = 0.5$	$\beta = 0.8$
$\tau_T(1)$	0.053	0.050	0.036	0.008
$\tau_T(4)$	0.052	0.050	0.036	0.008
$\tau_T(16)$	0.047	0.050	0.071	0.146
$\tau_T(32)$	0.033	0.050	0.169	0.399
$\widetilde{\alpha}_T$	0.052	0.050	0.056	0.068
$\widehat{\varrho}_T$	0.045	0.050	0.102	0.452
ADF(4)	0.048	0.055	0.073	0.533
ADF(12)	0.057	0.058	0.057	0.088
b) Empirical power				
test statistic	$\phi = 0.95$	$\phi = 0.90$	$\phi = 0.80$	$\phi = 0.50$
$\tau_T(1)$	0.013	0.002	0.001	0.001
$\tau_T(4)$	0.101	0.116	0.074	0.013
$\tau_T(16)$	0.118	0.206	0.313	0.365
$\tau_T(32)$	0.120	0.214	0.350	0.518
$\widetilde{\alpha}_T$	0.087	0.107	0.114	0.089
$\widehat{\varrho}_T$	0.182	0.420	0.788	0.995
ADF(4)	0.162	0.454	0.901	1.000
ADF(12)	0.127	0.260	0.491	0.765

Note: The entries of the table display the rejection frequencies based on 10.000 replications of model (13), where D_t is a linear time trend.

Table 3: Testing Hypotheses on the Cointegration Rank

$H_0 : r = 1\ ,\quad \phi_2 = 0.8$				
test statistic	$\phi_1 = 1$	$\phi_1 = 0.95$	$\phi_1 = 0.90$	$\phi_1 = 0.80$
$\theta = 0$				
Λ_T	0.059	0.346	0.604	0.853
LR(4)	0.072	0.428	0.894	0.999
LR(12)	0.048	0.180	0.389	0.636
$\theta = 0.8$				
Λ_T	0.043	0.295	0.566	0.853
LR(4)	0.057	0.310	0.793	0.999
LR(12)	0.063	0.190	0.382	0.636
$H_0 : r = 1\ ,\quad \phi_2 = 1$				
test statistic	$\phi_1 = 1$	$\phi_1 = 0.95$	$\phi_1 = 0.90$	$\phi_1 = 0.80$
$\theta = 0$				
Λ_T	0.107	0.300	0.582	0.900
LR(4)	0.083	0.241	0.558	0.962
LR(12)	0.094	0.166	0.290	0.506
$\theta = 0.8$				
Λ_T	0.107	0.240	0.508	0.854
LR(4)	0.083	0.511	0.949	1.000
LR(12)	0.094	0.352	0.581	0.768

Note: The entries of the table report the rejection frequencies based on 10.000 replications of model (14), where $E(y_t)$ is constant.