Estimation and testing for cointegration with trended variables : a comparison of a static and a dynamic regression procedure

Boswijk, H.P.

Publication date
1989

Title: ESTIMATION AND TESTING FOR COINTEGRATION WITH TRENDED VARIABLES: A Comparison of a Static and a Dynamic Regression Procedure

Author: H. Peter Boswijk

Address University of Amsterdam
Institute of Actuarial Science and Econometrics
Jodenbreestraat 23
1011 NH Amsterdam
The Netherlands

Date: October 1989

Series and Number: Report AE 11/89

Pages: 38

Price: No Charge

JEL Subject Classification: 211

Keywords: cointegration, deterministic trends, static regression, dynamic regression, Wald test, asymptotic properties, Brownian motions, Monte Carlo simulations

Abstract: In this paper two estimation and testing procedures for cointegration are compared for variables with deterministic trends. The first procedure consists of static regression estimator and a residual unit root test; the second estimator and (Wald) test statistic are derived from a dynamic regression. The asymptotic properties are derived, and critical values for the Wald test are simulated. A Monte Carlo simulation study suggests that in small samples the performance of the two procedures critically depends on the parameters of the data generating process.

1. *Introduction*

The notion of cointegration, which has received much attention in recent publications, was introduced by GRANGER (1981), and further developed in ENGLE AND GRANGER (1987). It captures the empirical observation that many economic variables seem to fluctuate widely, with increasing variance, while particular linear combinations of those variables, representing deviations from economic equilibrium relationships, have finite variance.

Least squares estimation of cointegrating vectors has been studied by PHILLIPS AND DURLAUF (1986) and STOCK (1987), who proved that the short run dynamic properties can be neglected for consistent estimation of the long run parameters. Furthermore, these estimators converge to the true parameter values at a faster rate than usual, a feature which is known as super consistency. In both studies it is assumed that the variables have no deterministic trends. However, many economic variables seem to have some constant drift in the long run. In this paper, the implications of the presence of such variables for the estimation and testing of cointegrating relationships are studied. For the asymptotic derivations, we shall utilize and extend the results obtained by PARK AND PHILLIPS (1988,1989). Since BANERJEE ET AL. (1986) showed that in small samples, least squares estimators from static regressions can have substantial biases, we shall also perform some Monte Carlo experiments.

Two alternative procedures are considered, both based on least squares estimation of single linear regression equations. The first one, proposed by ENGLE AND GRANGER (1987), consists of estimating a cointegrating vector in a static regression equation, and using the residuals from this equation in the Dickey-Fuller test or the Durbin-Watson statistic for a test on integration. The second procedure, a generalization of the one employed by BANERJEE ET AL. (1986), consists of estimating a single error correction equation, from which both an estimator of the cointegrating vector and a test for cointegration can be derived. Both procedures only use limited information; full information procedures, derived in a multivariate setting, have recently been proposed by PHILLIPS AND OULIARIS (1988b) and JOHANSEN (1988). However, the procedures studied in this paper are still of interest, because they are easily implemented and interpreted.

In section 2 the model is stated and some alternative representations are considered. The relationship between cointegrated variables and error correction models, which was established in ENGLE AND GRANGER (1987), is restated for the present case. In the third section, the estimators and test statistics are defined. The asymptotic properties are derived in section 4. Next, in section 5 some Monte Carlo simulation results for the small sample properties of the two procedures are presented and discussed, and finally in section 6 we come to our conclusions.

2. *The model: assumptions and representations*

Consider an N dimensional vector stochastic process $\{z_t, t=1,2,...\}$, with the following properties:

ASSUMPTIONS :

(i) *All components of* z_t *are integrated of order 1, that is,* Δz_t *is stationary.*

(ii) *There exist a linear combination* $v_t = \alpha' z_t$ *which is stationary. The vector* α *is unique up to a scale factor.*

(iii) *The joint distribution of* $\{\Delta z_t, t=1,...\}$ *is Gaussian with finite mean.*

(iv) *The conditional memory of the process* $\{z_t\}_1^\infty$ *is restricted to* p *periods.*

The first two assumptions imply that the components of z_t are cointegrated. The second part of condition (ii) is necessary for single-equation methods to be applicable; otherwise identification problems would arise. Without imposing any exogeneity or causality restrictions, we can partition z_t as $(y_t, x_t')'$ and normalize α as $(1, -\theta')'$. Using these definitions, we can interpret the first two assumptions as follows: they postulate a long run equilibrium relationship $y_t = \theta' x_t$, and state that the deviation from the equilibrium v_t is a stationary stochastic process. The third condition allows the process $\{z_t\}_1^\infty$ to comprise a linear deterministic trend. The last two assumptions are used for the following derivations. Letting F_t denote the σ-field generated by $\{z_t, z_{t-1},...\}$, there are NxN matrices A_j, j=1,...,p, and Ω, and an Nx1 vector m such that we have:

$$E\left[z_t \mid F_{t-1} \right] = \sum_{j=1}^{p} A_j z_{t-j} + m , \tag{1}$$

$$V\left[z_t \mid F_{t-1} \right] = \Omega , \tag{2}$$

$$\varepsilon_t = z_t - E\left[z_t \mid F_{t-1} \right] = z_t - \sum_{j=1}^{p} A_j z_{t-j} - m$$

$$= A(L) z_t - m , \tag{3}$$

where L is the lag or backshift operator, A(L) is an NxN matrix lag polynomial with $A(0) = I_N$ and $\{\varepsilon_t\}_1^\infty$ is, by construction, an N dimensional Gaussian white noise sequence with covariance matrix Ω. Rearranging the terms in (3), the *vector autoregressive representation* of z_t is obtained:

$$A(L) z_t = m + \varepsilon_t . \tag{4}$$

Because of assumption (i), Δz_t has the following Wold representation:

$$\Delta z_t = \mu + \sum_{j=0}^{\infty} C_j \varepsilon_{t-j} = \mu + C(L)\varepsilon_t , \tag{5}$$

where C(L) is an NxN matrix lag polynomial, defined by $(I-L)A(L)^{-1}$, with exponentially decreasing matrices C_j, and

$$\mu = \sum_{j=0}^{\infty} C_j m = C(1)m , \tag{6}$$

which is the drift vector. We can always express C(L) as:

$$\sum_{j=0}^{\infty} C_j L^j = \sum_{j=0}^{\infty} C_j + (I-L) \sum_{j=1}^{\infty} (- \sum_{i=j}^{\infty} C_i) L^{j-1}$$

$$= C(1) + \Delta C^*(L) , \tag{7}$$

where $C^*(L)$ is a matrix lag polynomial with exponentially decreasing matrices C_j^*. Therefore we can rewrite (5) as

$$\Delta z_t = \mu + C(1)\varepsilon_t + \Delta C^*(L)\varepsilon_t . \tag{8}$$

Integrating both sides of this equation, we obtain the *common trends representation*, introduced by Stock and Watson (1988):

$$z_t = \mu t + C(1) \sum_{s=1}^{t} \varepsilon_s + C^*(L)\varepsilon_t + c , \tag{9}$$

where c is equal to the sum of the initial conditions $z_0 - C^*(L)\varepsilon_0$, which we consider to be non-stochastic. This representation clarifies that z_t is the sum of a linear deterministic trend, a stochastic trend (or random walk) and a stationary part. From (9) it is clear that any linear combination $\alpha' z_t$ can only be stationary if $\alpha' C(1) = 0$ (which implies $\alpha'\mu = 0$ because of equation (6)). Therefore, the existence of a unique cointegrating vector requires that rank C(1) = N-1. The equilibrium error $v_t = \alpha' z_t$ is stationary with the following Wold representation:

$$v_t = \nu + K(L)\varepsilon_t = \nu + u_t , \tag{10}$$

where $\nu = \alpha' c$ and $K(L) = \alpha' C^*(L)$.

The connection between cointegrated systems and *error correction mechanisms* is stated in the *Granger Representation Theorem* (see Engle and Granger (1987)). Engle and Yoo (1987) extended this theorem to the case where z_t has a linear trend. This version is repeated below:

GRANGER REPRESENTATION THEOREM :

If the N dimensional stochastic process $\{z_t\}_1^\infty$ *satisfies assumptions* (i) - (iv) *and equations* (4) *and* (5), *then* :

(i) A(1) *has rank* 1 ,

(ii) *There exist* N×1 *vectors* α *and* γ *such that*

$$\alpha' C(1) = 0, \qquad C(1)\gamma = 0, \qquad A(1) = \gamma\alpha' ,$$

(iii) *There exists an error correction representation* :

$$\Delta z_t = m - \gamma\alpha' z_{t-1} + \sum_{j=1}^{p-1} A_j^* \Delta z_{t-j} + \varepsilon_t . \tag{11}$$

Proof : See ENGLE AND GRANGER (1987), pp. 256-258, with ε_t replaced by $(m + \varepsilon_t)$. □

The error correction representation from this theorem consists of a system of simultaneous equations. Estimation of the parameters of (11), in particular of α, should be performed under non-linear cross-equation restrictions, because α appears in all equations. A solution for this estimation problem is provided by JOHANSEN (1988) for the case of integrated processes with no drift.

Alternatively, we can derive a single error-correction equation, from which the cointegrating vector can be identified. First we partition m as $(m_1, m_2')'$, and A_j, j=1,...,p, and Ω as

$$A_j = \begin{bmatrix} a_{j,11} & a_{j,12} \\ a_{j,21} & A_{j,22} \end{bmatrix}, \qquad \Omega = \begin{bmatrix} \omega_{11} & \omega_{21}' \\ \omega_{21} & \Omega_{22} \end{bmatrix}.$$

Because of the joint normality of y_t and x_t, we have:

$$E\left[\, y_t \mid x_t, F_{t-1} \,\right] = \kappa + \beta_0' x_t + \sum_{j=1}^{p} (\varphi_j y_{t-j} + \beta_j' x_{t-j}) , \tag{12}$$

$$V\left[\, y_t \mid x_t, F_{t-1} \,\right] = \sigma^2 , \tag{13}$$

where

$$\kappa \equiv m_1 - \omega_{21}'\Omega_{22}^{-1}m_2 , \qquad \beta_0 \equiv \Omega_{22}^{-1}\omega_{21} , \qquad \sigma^2 \equiv \omega_{11} - \omega_{21}'\Omega_{22}^{-1}\omega_{21},$$
$$\varphi_j \equiv a_{j,11} - \omega_{21}'\Omega_{22}^{-1}a_{j,21} , \qquad \beta_j' \equiv a_{j,12} - \omega_{21}'\Omega_{22}^{-1}A_{j,22} , \qquad j = 1,2,\dots,p-1.$$

Letting $\varphi(L) = I - \sum_{j=1}^{p}\varphi_j L^j$ and $\beta(L) = \sum_{j=0}^{p}\beta_j L^j$, we can formulate the following autoregressive-distributed lag equation:

$$\varphi(L)y_t = \kappa + \beta(L)'x_t + \eta_t , \tag{14}$$

where $\{\eta_t\}_1^\infty$ is a Gaussian white noise sequence with variance σ^2. Cointegration implies that the roots of the equation det $\varphi(z) = 0$ all lie outside the unit circle, which is the usual *stability condition* and should not be confused with a *stationarity condition*, which is a condition on the joint process $\{(y_t, x_t')'\}_1^\infty$.

Equation (14) can always be rewritten in an error correction form. First, express the lag polynomials as

$$\begin{aligned} \varphi(L) &= (1- \sum_{j=1}^{p}\varphi_j)L + (I-L) + (I-L)\sum_{j=2}^{p}(\sum_{i=j}^{p}\varphi_i)L^{j-1} = \varphi(1)L + (I-L)\varphi^*(L) , \\ \beta(L) &= \sum_{j=0}^{p}\beta_j L + \beta_0(I-L) + (I-L)\sum_{j=2}^{p}(-\sum_{i=j}^{p}\beta_i)L^{j-1} = \beta(1)L + (I-L)\beta^*(L) , \end{aligned} \tag{15}$$

with $\varphi^*(0) = 1$ and $\beta^*(0) = \beta_0$. If we define $\lambda = (\lambda_1, \lambda_2')' \equiv (-\varphi(1), \beta(1)')'$, and substitute these expressions in (14), we have

$$\begin{aligned} \Delta y_t &= \kappa + \beta_0'\Delta x_t + \lambda_1 y_{t-1} + \lambda_2' x_{t-1} + \sum_{j=1}^{p-1}(\varphi_j^*\Delta y_{t-j} + \beta_j^{*\prime}\Delta x_{t-j}) + \eta_t \\ &= \kappa + \beta_0'\Delta x_t + \lambda_1(y_{t-1}-\theta' x_{t-1}) + \sum_{j=1}^{p-1}(\varphi_j^*\Delta y_{t-j} + \beta_j^{*\prime}\Delta x_{t-j}) + \eta_t , \end{aligned} \tag{16}$$

where it remains to be proved that $\lambda_2 = -\lambda_1\theta$. This follows from the definition of the parameters φ_j and β_j. Because $A_0 = I_N$, we have $\varphi_0 = (a_{0,11} - \omega_{21}'\Omega_{22}^{-1}a_{0,21}) = 1$ and $\beta_0' = -(a_{0,12} - \omega_{21}'\Omega_{22}^{-1}A_{0,22}) = \Omega_{22}^{-1}\omega_{21}$. With these expressions, it is easy to see that

$$\begin{aligned} (-\varphi(1), \beta(1)')' &= -[\ (a(1)_{11} - \omega_{21}'\Omega_{22}^{-1}a(1)_{21}) \ ,\ (a(1)_{12} - \omega_{21}'\Omega_{22}^{-1}A(1)_{22})\ ] \\ &= -(1, -\omega_{21}'\Omega_{22}^{-1})A(1) = -(1, -\beta_0')\gamma\alpha' , \end{aligned} \tag{17}$$

because of the Granger Representation Theorem, so we have $\lambda' = \lambda_1(1, -\theta')$ with $\lambda_1 = (1, -\beta_0')\gamma$.

3. *Two estimation and testing procedures*

I. Static regression

ENGLE AND GRANGER (1987) proposed to estimate θ in a static, so-called *cointegrating regression* equation. Using the definition $v_t = \alpha' z_t = y_t - \theta' x_t$ and equation (10), we have:

$$y_t = \nu + \theta' x_t + u_t . \qquad (18)$$

However, u_t is far from a white noise innovation, since u_t is serially correlated and, more importantly, u_t is generally correlated with x_t. Nevertheless, PHILLIPS AND DURLAUF (1986) and STOCK (1987) showed that because of the non-stationarity of x_t, the ordinary least squares estimator $\hat{\theta}_T$ is consistent and converges to the true value θ at a rate of $O_p(T^{-1})$ instead of the usual rate $O_p(T^{-1/2})$, where T is the sample size. A similar result has been proved for the case where y_t and x_t have deterministic trends by PARK AND PHILLIPS (1988).

The residuals $\hat{u}_t$ of regression (18) can be used in the Durbin-Watson (DW) statistic for a test on integration or in an Augmented Dickey-Fuller (ADF) test, see DICKEY AND FULLER (1979), which is the t-ratio of π in the auxiliary regression equation

$$\Delta\hat{u}_t = \pi\hat{u}_{t-1} + \sum_{j=1}^{q} \delta_j \Delta\hat{u}_{t-j} + e_t . \qquad (19)$$

ENGLE AND GRANGER (1987) recommend this test on the basis of Monte Carlo simulation results; its critical values seem to be less dependent on the parametrization then those of alternative tests are. The Durbin-Watson statistic, which was proposed for a test on (co-)integration by SARGAN AND BHARGAVA (1983), seems to have higher power if the process v_t has a short conditional memory. This is a very special case, even if the process z_t can be represented by a low order VAR.

Alternatively, we can use the residuals $\hat{u}^*_t$ from the regression equation

$$y_t = \nu + \zeta t + \theta' x_t + u^*_t \qquad (20)$$

in the auxiliary regression (19) or in the Durbin-Watson statistic, or equivalently, use detrended data in (18). We shall designate the statistics constructed from these residuals ADF* and DW*, respectively. Although u_t is not trended, so that (20) is actually overparametrized, these test statistics are of interest because it has often been recommended to use detrended data if deterministic trends are present (see e.g. ENGLE AND GRANGER (1987, p.255)). Moreover, their asymptotic properties differ from those of the standard statistics DW and ADF, as we shall see in section 4.

II. Autoregressive-Distributed lag regression

Least squares estimation of (16) yields $\hat{\lambda}_{1T}$ and $\hat{\lambda}_{2T}$. With these, it is obvious to define the estimator

$$\tilde{\theta}_T = -\frac{1}{\hat{\lambda}_{1T}} \hat{\lambda}_{2T} . \tag{21}$$

Although this estimator uses more information than the static regression estimator, it is in general not a maximum likelihood estimator. To see this, note that the joint density $f(y_t, x_t | F_{t-1})$ can be factorized as $f(y_t | x_t, F_{t-1})\ f(x_t | F_{t-1})$; θ enters both the conditional density of y_t and the marginal density of x_t (see equation (11)), so x_t is generally not weakly exogenous for θ. The only exception to this is the case where y_t is the only 'error correcting' variable in the system, that is, when γ in (11) is equal to $(\gamma_1, 0')'$. We shall not impose this restriction here. Note that the presence of the constant term κ in (16) is the consequence of the presence of a deterministic trend in z_t. If μ=0, it can in principle be removed from the regression equation, but we shall not investigate the consequences of this restriction.

BEWLEY (1979) and WICKENS AND BREUSCH (1988) show that $\tilde{\theta}_T$ is identical to the instrumental variable estimator of θ in

$$y_t = \kappa^* + \theta' x_t + \sum_{j=0}^{p-1} (\varphi_j^{**} \Delta y_{t-j} + \beta_j^{**\prime} \Delta x_{t-j}) + \eta_t^* , \tag{22}$$

with $\{y_{t-1},\dots,y_{t-p},x_t,\dots,x_{t-p}\}$ as instruments. Equation (22) is called the *pseudo-structural form*. Because this equation is exactly identified, the estimator $\tilde{\theta}_T$ has no finite moments (see KINAL (1980)). This will have special implications for the interpretation of the Monte Carlo results in section 5.

We can use the Wald test (F) for the joint significance of y_{t-1} and x_{t-1} in (16) as a test for cointegration. Under the null of no cointegration, λ_1 and λ_2 should be equal to zero, while under the alternative hypothesis that y_t and x_t are cointegrated, $-2<\lambda_1<0$ and $\lambda_2 \neq 0$. This test is in line with the one proposed by BANERJEE ET AL. (1986); they consider only the case where N=2 and θ=1, and use the t-ratio of λ_1, the coefficient of the error correction term $(y_{t-1}-x_{t-1})$.

We shall also consider F^*, which is the Wald test for the joint significance of y_{t-1} and x_{t-1} in equation (16) with a linear trend term added to the regressors. The true coefficient of this regressor is zero both under the null and under the alternative hypothesis, but, as we shall show in the next section, the use of detrended data has the advantage that the asymptotic distribution of the Wald test is free of nuisance parameters.

4. *Asymptotic properties*

In order to derive the asymptotic results, we shall need the following lemmas, based on the *multivariate invariance principle* from PHILLIPS AND DURLAUF (1986), the *continuous mapping theorem* (see BILLINGSLEY (1968)) and a special weak convergence result proved in PHILLIPS (1988). After that, we shall make extensive use of the results from PARK AND PHILLIPS (1988,1989) and PHILLIPS AND OULIARIS (1988a).

LEMMA 1 (Multivariate invariance principle) :

Let $\{x_t\}_1^{\infty}$ *be an* N *dimensional process with*

$$\Delta x_t = u_t = \Theta(L)\varepsilon_t ,$$

a zero-mean stationary Gaussian MA(∞) *process with* $V(\varepsilon_t) = \Omega$, *a positive definite matrix. Next, define* $S_t = x_t - x_0$, *and let*

$$\Sigma = \lim_{T\to\infty} E(\tfrac{1}{T} S_T S_T') = \Theta(1)\Omega\Theta(1)' ,$$

$$X_T(r) = \tfrac{1}{\sqrt{T}} S_{[rT]} = \tfrac{1}{\sqrt{T}} S_{j-1} , \qquad (j-1)/T \leq r < j/T , \qquad (j=1,\dots,T , r \in [0,1]).$$

If Σ *positive definite then, as* $T\to\infty$,

$$X_T(r) \Rightarrow B(r) \equiv \Sigma^{1/2}W(r),$$

where B(r) *and* W(r) *are* N *dimensional Brownian motions with covariance matrices* Σ *and* I_N, *respectively, and the symbol "⇒" denotes weak convergence of the associated probability measures.*

Proof : Proofs are given in appendix A. □

LEMMA 2 (Continuous Mapping Theorem) :

If $X_T(r) \Rightarrow B(r)$ *as* $T\to\infty$ *and* h *is a continuous functional, except for a set* D *for which* $P\{B(r) \in D\} = 0$, *then* $h(X_T(r)) \Rightarrow h(B(r))$ *as* $T\to\infty$.

LEMMA 3 :

Make the assumptions of lemma 1, and let

$$\Sigma_1 = \lim_{T\to\infty} \sum_{t=2}^{T} E(u_1 u_t')$$

Then, as $T\to\infty$, *we have:*

(a) $T^{-3/2} \sum_{t=1}^{T} x_t \Rightarrow \int_0^1 B(r)\, dr$,

(b) $T^{-2} \sum_{t=1}^{T} x_t x_t' \Rightarrow \int_0^1 B(r)B(r)'\, dr$,

(c) $T^{-1} \sum_{t=1}^{T} x_{t-1} u_t' \Rightarrow \int_0^1 B(r)\, dB(r)' + \Sigma_1$,

(d) $T^{-5/2} \sum_{t=1}^{T} t x_t \Rightarrow \int_0^1 rB(r)\, dr$,

(e) $T^{-3/2} \sum_{t=1}^{T} t u_t \Rightarrow \int_0^1 r\, dB(r)$.

The results of lemma 3 form the building blocks of the asymptotic theory for regressions with integrated processes. All estimators and test statistics can be expressed as products of second moment matrices, and with the continuous mapping theorem we can link these separate convergence results together, yielding the required asymptotic expressions.

Estimation

In theorem 1, we state some asymptotic results for the estimators $\hat{\theta}_T$ and $\tilde{\theta}_T$. For the static regression estimator, these results have been proved by PHILLIPS AND DURLAUF (1986) and PARK AND PHILLIPS (1988); they are repeated below for comparison. For $\tilde{\theta}_T$, we use theorems 3.1 and 3.2 from PARK AND PHILLIPS (1989). In order to derive the asymptotic results for this estimator, it is convenient to premultiply the partial regressor vector z_{t-1} and its coefficient vector λ with an orthogonal matrix $G' = (g_1, G_2)'$, where g_1 is the normalized cointegrating vector and

$$G_2 = \begin{bmatrix} \theta' \\ \cdot\cdot \\ I_{N-1} \end{bmatrix} \begin{bmatrix} I_{N-1} + \theta\theta' \end{bmatrix}^{-\frac{1}{2}} , \qquad (23)$$

so the columns of G_2 span the orthogonal complement of the cointegrating vector. The matrix G in effect decomposes z_{t-1} into a stationary variable and an integrated (N-1)-vector.

We provide the asymptotic distributions only for a special case, because the other asymptotic distributions involve quite complicated functionals of Brownian motions and depend on some nuisance parameters, so they are not well suited for efficiency assessment or specification testing. However, those results are available from the author on request.

THEOREM 1 :

Let $\{z_t\}_1^\infty = \{(y_t, x_t')'\}_1^\infty$ *satisfy assumptions* (i)-(iv). *Next, let* $\mu=(\mu_y,\mu_x')'$, *and define* $\sigma_u^2=K(1)\Omega K(1)'$, *where the lag polynomial* K(L) *is defined in equation* (10). *Furthermore, let the* (N-1)×1 *vector* μ_G *be defined by* $G_2'\mu$. *Then we have, as* $T\to\infty$:

(a)
$$\hat{\theta}_T \xrightarrow{p} \theta ,$$
$$\tilde{\theta}_T \xrightarrow{p} \theta .$$

(b) *If* N = 2 *and* $\mu \neq 0$, *then*
$$T^{3/2}(\hat{\theta}_T - \theta) \Rightarrow N(0,12\sigma_u^2/\mu_x^2) ,$$
$$T^{3/2}(\tilde{\theta}_T - \theta) \Rightarrow N(0,12\sigma^2/(\lambda_1\mu_x)^2) .$$

(c) *If* N > 2 *or* $\mu = 0$, *then*
$$\hat{\theta}_T - \theta = O_p(T^{-1}) ,$$
$$\tilde{\theta}_T - \theta = O_p(T^{-1}) .$$

(d) *If* $\mu \neq 0$, *then*
$$\mu_x'(\hat{\theta}_T - \theta) = O_p(T^{-3/2}) ,$$
$$\mu_G'(\tilde{\theta}_T - \theta) = O_p(T^{-3/2}) .$$

The results of theorem 1 indicate that the presence of deterministic trends can only improve the convergence of the estimator $\hat{\theta}_T$ and $\tilde{\theta}_T$. Note that if the cointegrating relationship involves only two variables, both x_t and $G_2'z_t$ asymptotically behave as linear trends, so the convergence rate is even faster. In that case the limiting distribution of both $\hat{\theta}_T$ and $\tilde{\theta}_T$ is normal, which is quite exceptional for regressions with integrated processes. If $N > 2$ and $\mu \neq 0$, the distributions of $T(\hat{\theta}_T - \theta)$ and $T(\tilde{\theta}_T - \theta)$ lie in (N-2) dimensional subspaces of $\mathbf{R}^{N-1}$; therefore the results of part (d) apply.

Testing

In order to derive the asymptotic distributions of the test statistics under the null hypothesis of no cointegration, we shall replace assumption (ii) with

(ii)' *There are no stationary linear combinations of the components of* z_t, *so* C(1) *is non-singular.*

For the asymptotic distributions of the residual based tests DW and ADF we need the distribution of the static regression estimator $\hat{\theta}_T$. This will be given in lemma 4, but first we make the following definitions. We define the N dimensional zero-mean stationary process v_t and its partial sum process S_t as

$$v_t = \Delta z_t - \mu = C(L)\varepsilon_t , \tag{24}$$

$$S_t = \sum_{s=1}^{t} v_s = z_t - \mu t - z_0 . \tag{25}$$

Next, we shall be using the following moment matrices:

$$\Sigma_0 = E\, v_1 v_1' = \sum_{j=0}^{\infty} C_j \Omega C_j' , \tag{26}$$

$$\Sigma_1 = \lim_{T\to\infty} \sum_{t=2}^{T} E v_1 v_t' = \sum_{j=0}^{\infty} \sum_{i=j+1}^{\infty} C_j \Omega C_i' , \tag{27}$$

$$\Sigma = \lim_{T\to\infty} E(\tfrac{1}{T} S_T S_T') = C(1)\Omega C(1)' = \Sigma_0 + \Sigma_1 + \Sigma_1' , \tag{28}$$

Next, define $\pi = \mu_y(\mu_x'\mu_x)^{-1}\mu_x$ and $h_1 = (\mu_x'\mu_x)^{-1/2}\mu_x$. Furthermore, let $H = (h_1, H_2)$ be an orthogonal (N-1)×(N-1) matrix, so that the columns of H_2 span the orthogonal complement of μ_x. We define the (N-1)×N matrix R as

$$R = \begin{bmatrix} 1 : & -\pi' \\ 0 : & H_2' \end{bmatrix}. \tag{29}$$

Let $W_1(r) = (W_{11}(r), W_{12}(r)')'$ denote the standard N dimensional Brownian motion (with covariance matrix I_N), and let $W_2(r) = (W_{21}(r), W_{22}(r)')'$ denote the standard (N-1) dimensional Brownian motion. We define $B_1(r)$ and $B_2(r)$ as follows:

$$B_1(r) = (B_{11}(r), B_{12}(r)')' = \Sigma^{1/2} W(r) , \tag{30}$$

$$B_2(r) = (B_{21}(r), B_{22}(r)')' = R\, B_1(r) . \tag{31}$$

For any process X(r) on [0,1] we let $X^*(r)$ and $X^{**}(r)$ denote the demeaned and detrended processes, respectively:

$$X^*(r) = X(r) - \int_0^1 X(s)ds , \tag{32}$$

$$X^{**}(r) = X(r) - 4(\int_0^1 X(s)ds - \tfrac{3}{2} \int_0^1 sX(s)ds) + 6r(\int_0^1 X(s)ds - 2 \int_0^1 sX(s)ds) . \tag{33}$$

Finally, we shall use the following scalar process:

$$P(r) = r - \tfrac{1}{2} - \int_0^1 sB_{22}^*(s)'ds \, (\int_0^1 B_{22}^*(s)B_{22}^*(s)'ds)^{-1} B_{22}^*(r) . \tag{34}$$

With these definitions, we can state the asymptotic properties of $\hat{\theta}_T$ under the null hypothesis of no cointegration.

LEMMA 4 :

Let $\{z_t\}_1^\infty$ *satisfy assumptions* (i), (ii)', (iii) *and* (iv). *Then, as* $T \rightarrow \infty$,

(a) *If* $N = 2$ *and* $\mu \neq 0$, *then*

$$\sqrt{T}(\hat{\theta}_T - \pi) \Rightarrow \frac{12}{\mu_x} \int_0^1 (r-\tfrac{1}{2})B_{21}^*(r)dr = \xi_1 ,$$

(b) *If* $N > 2$ *and* $\mu \neq 0$, *then*

$$\hat{\theta}_T - \pi \Rightarrow H_2 \, (\int_0^1 B_{22}^{**}(r)B_{22}^{**}(r)'dr)^{-1} \int_0^1 B_{22}^{**}(r)B_{21}^{**}(r)dr = \xi_2 ,$$

$$\sqrt{T}\mu_x'(\hat{\theta}_T - \pi) \Rightarrow (\int_0^1 P(r)^2 dr)^{-1} \int_0^1 P(r)B_{21}^*(r)dr = \xi_3 ,$$

(c) *If* $\mu = 0$, *then*

$$\hat{\theta}_T \Rightarrow \left(\int_0^1 B_{12}^*(r)B_{12}^*(r)'dr \right)^{-1} \int_0^1 B_{12}^*(r)B_{11}^*(r)dr = \xi_4 ,$$

(d) *For* $\hat{\theta}_T^*$, *the static estimator with a trend term included in the regression, we have*

$$\hat{\theta}_T^* \Rightarrow \left(\int_0^1 B_{12}^{**}(r)B_{12}^{**}(r)'dr \right)^{-1} \int_0^1 B_{12}^{**}(r)B_{11}^{**}(r)dr = \xi_5 .$$

The results of lemma 3 are similar to the 'spurious regression' results of PHILLIPS (1986); statement (c) is in fact part of his theorem 2. If $k > 1$ and the variables are trended, the estimator $\hat{\theta}_T$ converges to a non-degenerate random variable; with only one regressor however, it converges to the ratio of the drift parameters. In both cases the residuals $\hat{u}_t$ do not have linear trends, because $\hat{\alpha}_T \equiv (1, -\hat{\theta}_T')'$ converges to a (random) variable that is orthogonal to the drift vector μ.

In theorem 2, we state the asymptotic distributions of the static regression residual based tests ADF and DW. We let ζ_i, i=1,..,4, denote the limits of the estimator $\hat{\alpha}_T \equiv (1, -\hat{\theta}_T')'$ for the separate cases considered in lemma 4. Thus,

$$\zeta_1 = (1, -\pi)' , \qquad \zeta_2 = (1, -(\pi+\xi_2)')' , \qquad \zeta_3 = (1, -\xi_4')' , \qquad \zeta_4 = (1, -\xi_5')' . \tag{35}$$

Although the process z_t can be represented by a finite vector autoregression, this is generally not the case for u_t. Therefore, we need the following condition:

CONDITION 1 : *For* q, *the order of the autoregression in* (18), *we require that* $q \to \infty$ *as* $T \to \infty$ *and* $q = o_p(T^{1/3})$.

SAID AND DICKEY (1984) have shown that under this condition the asymptotic distribution of the ADF test, applied to original data, is free of nuisance parameters, even if these data follow general ARIMA(p,1,q) processes. PHILLIPS AND OULIARIS (1988a) extended this result to the case where residuals instead of original data are used. We reiterate and generalize their results in theorem 2.

THEOREM 2 :

Let $\{z_t\}_1^\infty$ *satisfy the assumptions of lemma 4, and assume that condition 1 holds. Then, as* $T \to \infty$,

(a) *If* N = 1 *and* $\mu \neq 0$, *then*

$$T\,DW \Rightarrow \zeta_1'\Sigma_0\zeta_1 \Big/ \int_0^1 B_{21}^{**}(r)^2 dr ,$$

$$ADF \Rightarrow \Big(\int_0^1 W_2^{**}(r)^2 dr \Big)^{-1/2} \int_0^1 W_2^{**}(r)\, dW_2^{**}(r) ,$$

(b) *If* N > 2 *and* $\mu \neq 0$, *then*

$$T\,DW \Rightarrow \zeta_2'\Sigma_0\zeta_2 \Big/ \int_0^1 B_{21\bullet 2}^{**}(r)^2 dr ,$$

$$ADF \Rightarrow \Big[(\tau_2^{**\prime}\tau_2^{**}) \int_0^1 W_{21\bullet 2}^{**}(r)^2 dr \Big]^{-1/2} \int_0^1 W_{21\bullet 2}^{**}(r)\, dW_{21\bullet 2}^{**}(r) ,$$

(c) *If* $\mu = 0$, *then*

$$T\,DW \Rightarrow \zeta_4'\Sigma_0\zeta_4 \Big/ \int_0^1 B_{11\bullet 2}^{*}(r)^2 dr ,$$

$$ADF \Rightarrow \Big[(\tau_1^{*\prime}\tau_1^{*}) \int_0^1 W_{11\bullet 2}^{*}(r)^2 dr \Big]^{-1/2} \int_0^1 W_{11\bullet 2}^{*}(r)\, dW_{11\bullet 2}^{*}(r) ,$$

(d) *For* DW* *and* ADF*, *we have*

$$T\,DW^{*} \Rightarrow \zeta_5'\Sigma_0\zeta_5 \Big/ \int_0^1 B_{11\bullet 2}^{**}(r)^2 dr ,$$

$$ADF^{*} \Rightarrow \Big[(\tau_1^{**\prime}\tau_1^{**}) \int_0^1 W_{11\bullet 2}^{**}(r)^2 dr \Big]^{-1/2} \int_0^1 W_{11\bullet 2}^{**}(r)\, dW_{11\bullet 2}^{**}(r) ,$$

where

$$B_{i1\bullet 2}^{*}(r) = B_{i1}^{*}(r) - \int_0^1 B_{i1}^{*}(s)B_{i2}^{*}(s)'ds \Big(\int_0^1 B_{i2}^{*}(s)B_{i2}^{*}(s)'ds\Big)^{-1} B_{i2}^{*}(r) ,$$

for i=1,2, *with* $B_{i1\bullet 2}^{**}(r)$, $W_{i1\bullet 2}^{*}(r)$ *and* $W_{i1\bullet 2}^{**}(r)$ *defined analogously, and*

$$\tau_1^{*\prime} = \Big[1 , \; - \int_0^1 W_{11}^{*}(r)W_{12}^{*}(r)'dr \Big(\int_0^1 W_{12}^{*}(r)W_{12}^{*}(r)'dr \Big)^{-1} \Big] ,$$

with τ_1^{**} *and* τ_2^{**} *defined in the same way from the detrended standard Brownian motions* $W_1^{**}(r)$ *and* $W_2^{**}(r)$.

In all cases, the Durbin-Watson statistic converges to zero, and the asymptotic distribution of the normalized DW test statistic depends on unknown parameters. The reason for this is that DW does not correct for the serial correlation of Δz_t. It can be proved that if the system can be represented by a first order VAR, so that $\Sigma = \Sigma_0$, the asymptotic distribution of T DW is free of nuisance parameters. This is however a very restrictive case.

The asymptotic distribution of the Augmented Dickey-Fuller test statistic is in all cases free of nuisance parameters. The asymptotic expressions are all examples of the functional

$$\left(\int_0^1 X(r)^2 dr \right)^{-1/2} \int_0^1 X(r)\, dX(r) \,, \tag{36}$$

where X(r) is a univariate process on [0,1]. If X(r) is W(r), $W^*(r)$ or $W^{**}(r)$, then (36) gives the distribution of the Dickey-Fuller statistics $\hat{\tau}$, $\hat{\tau}_\mu$ and $\hat{\tau}_\tau$, tabulated in FULLER (1976, p.373). For a multivariate process $X(r) = (X_1(r), X_2(r)')'$, the process $X_{1\bullet 2}(r)$ can be interpreted as $X_1(r)$, 'corrected for' $X_2(r)$; it is the projection of $X_1(r)$ on the orthogonal complement of $X_2(r)$.

Part (c) and (d) of theorem 2 have been proved in PHILLIPS AND OULIARIS (1988a). The effect of the deterministic trends in (a) and (b) is a reduction of the dimension of the Brownian motion. Therefore, if N=2 and $\mu \neq 0$, ADF is asymptotically distributed as $\hat{\tau}_\tau$, the ADF statistic for a detrended univariate process. In this case, the ADF test (and also the DW test) actually checks whether the linear combination that eliminates the deterministic trends is stationary. Similarly, when N>2 it is tested whether one of the N-1 linear combinations, orthogonal to the drift vector, is stationary. Note that although no unknown parameters enter the asymptotic expressions for ADF, it still depends on whether or not $\mu=0$. The 'detrended' version ADF^* does not have this disadvantage. Asymptotic critical values for ADF and ADF^* are tabulated in PHILLIPS AND OULIARIS (1988a, table 2a-2c).

We now derive the asymptotic distribution of F and F^* under the null of no cointegration.

THEOREM 3 :

Let $\{z_t\}_1^\infty$ *satisfy assumptions* (i), (ii)', (iii) *and* (iv). *Next, let* K_2 *be an* N×(N-1) *matrix such that* $K_2'\mu = 0$ *and* $K_2'K_2 = I_{N-1}$. *Define the following process on* [0,1]:

$$M(r) = \begin{bmatrix} (\mu'\mu)^{1/2} (r-\frac{1}{2}) \\ K_2'B_1^*(r) \end{bmatrix} .$$

Finally, let $s = \sigma^{-1} \Omega^{1/2}(1, -\omega_{21}'\Omega_{22}^{-1})'$. *Then, as* $T \to \infty$,

(a) *If* $\mu \neq 0$, *then*

$$F \Rightarrow s' \left[\int_0^1 dW_1^*(r)\, M(r)' \left(\int_0^1 M(r)M(r)'\, dr \right)^{-1} \int_0^1 M(r)\, dW_1^*(r) \right] s \quad ,$$

(b) *If* $\mu = 0$, *then*

$$F \Rightarrow \int_0^1 dW_{11}^*(r)\, W_1^*(r)' \left(\int_0^1 W_1^*(r)W_1^*(r)'\, dr \right)^{-1} \int_0^1 W_1^*(r)\, dW_{11}^*(r) \quad ,$$

(c) *For* F^*, *the Wald test in a regression with a trend term included, we have*

$$F^* \Rightarrow \int_0^1 dW_{11}^{**}(r)\, W_1^{**}(r)' \left(\int_0^1 W_1^{**}(r)W_1^{**}(r)'\, dr \right)^{-1} \int_0^1 W_1^{**}(r)\, dW_{11}^{**}(r) \quad .$$

From part (a) of theorem 3, it is clear that the asymptotic distribution of F depend on unknown parameters if the process z_t has a deterministic trend. Therefore, in that case F^* is to be preferred. Critical values for the asymptotic distributions of (b) and (c) have been obtained for various values of N by Monte Carlo simulation (using 10000 replications and T=500) and are tabulated in tables B1-B3 in appendix B. The expressions are equal to the first diagonal element of the matrix valued mapping

$$\int_0^1 dX(r)\, X(r)' \left(\int_0^1 X(r)X(r)'\, dr \right)^{-1} \int_0^1 X(r)\, dX(r)' \,, \qquad (37)$$

where X(r) is a demeaned or detrended Brownian motion. Note that this is the quadratic form of a multivariate version of the functional given by (36). Full information tests such as the multivariate unit root test of PHILLIPS AND DURLAUF (1986) or the cointegration test of JOHANSEN (1988) are asymptotically distributed as the trace of this functional, because they are derived from a vector autoregression instead of a single autoregressive-distributed lag equation.

In the next theorem we analyze the asymptotic behavior of all test statistics under the alternative hypothesis that z_t is cointegrated.

THEOREM 4 :

Let $\{z_t\}_1^\infty$ *satisfy assumptions* (i)-(iv). *Then we have*

(a) $DW = O_p(1)$,
$DW^* = O_p(1)$,

(b) $ADF = O_p(T^{1/2})$,
$ADF^* = O_p(T^{1/2})$,

(c) $F = O_p(T)$,
$F^* = O_p(T)$.

From the results in theorem 4 we deduce that all tests are consistent. The normalized Durbin-Watson statistic T DW and the Wald test diverge more rapidly than the ADF test. This suggest that the latter is less powerful in moderate samples; however, this may be misleading, because in the classical case with a single restriction, the F-test also diverges more rapidly that the t-test, but it is well known that their rejection frequencies are identical.

On the basis of the asymptotic results of this section we cannot make a clear choice between the two procedures. Both estimators are 'super'-consistent and have the same convergence rate. As for the test statistics, all tests are consistent, but one of them, the Durbin-Watson test, has the disadvantage that its asymptotic distribution cannot be used for critical values, because it depends on unknown parameters. The detrended test F^* is to be preferred over the standard version F, because its asymptotic distribution does not depend on nuisance parameters. For the ADF test, both versions are independent of unknown parameters. In this case, the standard version may be preferable, because it tests whether a linear combination that eliminates the deterministic trend also eliminates the stochastic trend. Because the asymptotic distribution of this test depends on whether or not $\mu=0$, this should be investigated before the cointegration test is carried out.

5. *Small sample properties : some Monte Carlo experiments*

Before presenting our own results, we shall review the findings of some earlier Monte Carlo studies. BANERJEE ET AL. (1986) compared the static and dynamic regression procedure for a bivariate system of cointegrated variables with no deterministic trends. They found that the dynamic regression estimator has a substantially lower bias than the static estimator. They also compared the powers of the Durbin-Watson test and a t-test on the significance of the restricted error correction term and concluded that in this respect the dynamic regression procedure also performs better. STOCK (1987) presented some Monte Carlo results on the small sample behavior of the static regression estimator (again, for variables with no deterministic trends) and found quite large biases. ENGLE AND GRANGER (1987) compared various tests on cointegration. They concluded that residual based tests preform better than tests based on a vector autoregression; whether the ADF-test or the DW-test is more powerful depends on the presence of serial correlation in Δz_t.

In order to investigate the small sample performance of the two estimation and testing procedures for trended data, we use simulated data from the following bivariate error correction system:

$$\Delta y_t = \beta \Delta x_t - \lambda(y_{t-1} - \theta x_{t-1}) + \varepsilon_{1t} , \tag{36}$$

$$\Delta x_t = \mu + \gamma(y_{t-1} - \theta x_{t-1}) + \varepsilon_{2t} , \tag{37}$$

with

$$\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \quad \text{i.i.d. } N\left(0 , \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \right) ,$$

$\beta = 0.5$, $\lambda \in \{0, 0.1, 0.5\}$,
$\theta = 1$, $\gamma \in \{0, 0.1\}$,
$\mu = 0.01$, $\sigma_1 \in \{0.01, 0.05\}$,
$\sigma_2 = 0.02$, $T \in \{24, 48, 100, 200\}$.

Although the parameter values are entirely imaginary, they may be associated with a system of logarithms of quarterly time series, such as aggregate consumer's expenditure and income; in that case standard errors are fractions of the dependent variables and the drift parameters represent mean quarterly growth rates. The parameter β is called the impact multiplier, and θ the long run multiplier; in a preliminary Monte Carlo study we also tried a value of β=0.1, but this did not change the performance significantly. Which parameter values are the most relevant depends on the variables being modelled; inspecting the consumption function of DAVIDSON ET AL. (1978) and the money demand equation of HENDRY AND RICHARD (1983) suggests that a standard error σ_1 of 0.01 and an adjustment rate λ of 0.1 may be considered realistic for those models. Note that if $\gamma \neq 0$, x_t is not weakly exogenous for θ in equation (16); therefore, in that case $\tilde{\theta}_T$ is not a full information estimator. Whether this is a relevant case for the forenamed relationships is hard to establish; usually weak exogeneity is assumed instead of tested.

From all possible parameter value combinations, we exclude those with $(\lambda,\gamma) = (0, 0.1)$. Although this case is feasible, it is difficult to interpret because y_t and x_t then are cointegrated, but it does not show up in a conditional model of y_t given x_t. We use the parameter values $\lambda=0$ and $\gamma=0$ to compare the empirical with the nominal significance levels. The lag length p in (16) is set to its true value 1. The lag length q in the auxiliary regression is set to 0 for all sample sizes, which may seem to be in conflict with condition 1, which requires q to be a slowly increasing function of T. However, under the null hypothesis u_t follows a first order autoregressive process, so no inclusion of extra lags is necessary. In both regression equations a constant term is included. For all experiments we use 1000 replications.

Estimation

In table 1 and 2, the bias and root mean squared error of the estimators $\hat{\theta}_T$ and $\tilde{\theta}_T$ are compared for various parameter values. In table 1, $\lambda = 0.1$, indicating that 10 percent of the equilibrium error is corrected in the next period. In this case therefore, cointegration of y_t and x_t will show up weakly in the data relative to the case where $\lambda=0.5$. The results for the sample sizes 24 and 48 are quite bad for both procedures: in all cases, either the bias or RMSE (or both) is larger than 0.1, which is 10% of the true value of θ.

On average, $\tilde{\theta}_T$ seems to have a smaller bias, but a larger RMSE. Closer examination reveals that this is the consequence of the estimator being defined as the *ratio* of two estimators. Because the true value of the denominator is relatively close to zero, there are quite some outliers, giving rise to very large variances. As argued in section 3, $\tilde{\theta}_T$ can be interpreted as an exactly identified instrumental variable estimator, and consequently it has no finite moments. Therefore the bias and RMSE of this estimator actually do not exist, and their Monte Carlo estimators will do a poor job in measuring the location and dispersion of $\tilde{\theta}_T$. For the larger sample sizes and for $\lambda=0.5$, this problem does not seem to be very serious, which is the reason why we still use the bias and root mean squared error for comparison of the two estimators.

In table 2, $\lambda = 0.5$. Now the true value of the denominator of the dynamic regression estimator is sufficiently far from zero, so there is no problem of outliers. The adjustment speed of the system is higher, which leads to better performance of both estimators, but the dynamic regression estimator seems to be superior in this case.

TABLE 1 : Bias and RMSE of $\hat{\theta}_T$ and $\tilde{\theta}_T$, $\lambda = 0.1$

	T	$\hat{\theta}_T$ bias	$\hat{\theta}_T$ RMSE	$\tilde{\theta}_T$ bias	$\tilde{\theta}_T$ RMSE
$\gamma = 0$,	24	-0.2091	0.2609	-0.0793	0.6403
$\sigma_1 = 0.01$	48	-0.0935	0.1327	-0.0229	0.1644
	100	-0.0305	0.0502	-0.0056	0.0402
	200	-0.0088	0.0184	-0.0013	0.0130
$\gamma = 0.1$,	24	-0.2220	0.2837	0.0284	3.2461
$\sigma_1 = 0.01$	48	-0.0981	0.1479	-0.0294	0.5399
	100	-0.0309	0.0545	-0.0110	0.0598
	200	-0.0088	0.0196	-0.0035	0.0197
$\gamma = 0$,	24	-0.1680	0.6877	-0.0622	2.8112
$\sigma_1 = 0.05$	48	-0.0814	0.3716	-0.0560	1.5305
	100	-0.0343	0.1549	-0.0110	0.1774
	200	-0.0105	0.0605	-0.0034	0.0632
$\gamma = 0.1$,	24	-0.1972	0.6460	-0.9150	2.8315
$\sigma_1 = 0.05$	48	-0.1032	0.3633	-0.5376	6.0746
	100	-0.0415	0.1583	-0.5663	0.7566
	200	-0.0168	0.0691	-0.4817	0.6715

The standard error σ_1 has an asymmetric effect on the estimators: the RMSE of $\hat{\theta}_T$ approximately triples when σ_1 is multiplied by 5, whereas the RMSE of the second estimator seems to be proportional with σ_1. The presence of an error correction term in the equation for x_t does not have a large effect on the performance of $\hat{\theta}_T$; on the dynamic regression estimator, the effect is negative, especially if $\sigma_1 = 0.05$. The case {λ=0.1, γ=0.1, σ_1=0.05} is quite disastrous for $\tilde{\theta}_T$: even with the larger sample sizes, the bias and RMSE hardly show any decrease. As noted before, whether this is relevant depends on the model and the data being modelled; however, the value of the standard error does seem to be quite large for log-linear models. Apart from this case, the convergence rate of $O_p(T^{-3/2})$ is reasonably reflected in the RMSE of both estimators: doubling of the sample size leads to a decrease with a factor approximately equal to $2^{-3/2}$ ($\approx$0.35).

TABLE 2 : Bias and RMSE of $\hat{\theta}_T$ and $\tilde{\theta}_T$, $\lambda = 0.5$

	T	$\hat{\theta}_T$ bias	$\hat{\theta}_T$ RMSE	$\tilde{\theta}_T$ bias	$\tilde{\theta}_T$ RMSE
$\gamma = 0$,	24	-0.0497	0.1065	-0.0015	0.0774
$\sigma_1 = 0.01$	48	-0.0132	0.0376	-0.0004	0.0244
	100	-0.0033	0.0108	-0.0003	0.0077
	200	-0.0008	0.0036	-0.0001	0.0027
$\gamma = 0.1$,	24	-0.0553	0.1108	-0.0053	0.0833
$\sigma_1 = 0.01$	48	-0.0150	0.0384	-0.0007	0.0268
	100	-0.0037	0.0109	-0.0006	0.0084
	200	-0.0009	0.0037	-0.0002	0.0024
$\gamma = 0$,	24	-0.0306	0.3265	0.0131	0.3857
$\sigma_1 = 0.05$	48	-0.0090	0.1143	0.0052	0.1193
	100	-0.0044	0.0375	-0.0011	0.0378
	200	-0.0012	0.0130	-0.0004	0.0128
$\gamma = 0.1$,	24	-0.0778	0.3242	-0.0626	0.7927
$\sigma_1 = 0.05$	48	-0.0229	0.1222	-0.0257	0.1425
	100	-0.0071	0.0389	-0.0080	0.0433
	200	-0.0020	0.0135	-0.0023	0.0145

To base a recommendation on these results is not easy. On the one hand, $\tilde{\theta}_T$ has a smaller bias in most cases. However, if the adjustment speed of the system is low, and only a small sample is available, there is a serious risk of obtaining a nonsensical value. Nevertheless, in this case cointegration will show up very weakly in the data, so an inaccurate estimate is to be expected with both procedures; moreover, the estimated bias and mean squared error may be especially unfavorable for $\tilde{\theta}_T$, because this estimator has no finite moments; other measures of location and dispersion may be more opportune.

Testing

In order to obtain 5% critical values for the Durbin-Watson statistic, we interpolate the upper and lower values from SARGAN AND BHARGAVA (1983) linearly in k (the number of regressors except for the constant term) and with a quadratic in 1/T. Note that for DW^* the number of regressors is inclusive of the trend term. With these, we construct critical values as:

$$d = 0.25\, d_L + 0.75\, d_U + k/2T , \tag{38}$$

which is suggested by the authors as a crude approximation. For the ADF test, we use the 5% critical value for $\hat{\tau}_\tau$ from FULLER (1976, p.373), which is appropriate because of theorem 2(a). For ADF^*, the 5% critical value of table 2c of PHILLIPS AND OULIARIS (1988a) is applicable, with n(=N-1)=1.

For the Wald test statistics F and F^*, we take critical values from tables B2 and B3, respectively. Because $\mu \neq 0$, it follows from theorem 3 that the asymptotic distribution of F depends upon nuisance parameters; because of this problem we use table B2, which would be applicable if there were no deterministic trends. Note that for both cases, the critical values are larger than the 5% critical value of the $\chi^2(2)$ distribution, which is (asymptotically) appropriate for a Wald test on the significance of two stationary variables.

Next, we establish the actual rejection frequency under the null hypothesis $\lambda=0$ and $\gamma=0$. Because p, the order of the system is equal to 1, the distributions of all test statistics do not depend on σ_1.

TABLE 3 : Rejection frequencies of the DW, ADF and F tests under H_0

T	DW	DW^*	ADF	ADF^*	F	F^*
24	0.068	0.060	0.094	0.104	0.107	0.088
48	0.075	0.088	0.083	0.095	0.091	0.065
100	0.064	0.049	0.061	0.057	0.082	0.052
200	0.058	0.048	0.062	0.067	0.081	0.052

For the Durbin-Watson test, the critical values for the detrended version seem to perform a little better; with the Dickey-Fuller test the converse is true. However, in both cases the difference is not large. The standard version of the Wald test has a rejection frequency which fails to come near the nominal significance level of 5%. This should come as no surprise, because the critical values used in this case are actually only applicable if the variables have no deterministic trends. The detrended test F^* performs much better. Note that for the ADF and F tests, the empirical significance level approaches the nominal size from above; this indicates that 'classical' critical values for these tests, taken from the N(0,1) and $\chi^2(2)$ distributions respectively, would perform even worse in small samples than asymptotically.

On the basis of these results, we recommend the use of the standard versions of the residual based tests DW and ADF, because they have an empirical significance level that is not too far from the nominal size. Furthermore, as we have shown in the previous section, the standard versions have the advantage that they test simultaneously whether there is a cointegrating vector and whether this vector also eliminates the deterministic trend form the variables.

TABLE 4 : Test rejection frequencies of the DW, ADF and F^* tests under H_1 (cointegration)

		DW		ADF		F^*	
Case	T	λ = 0.1	λ = 0.5	λ = 0.1	λ = 0.5	λ = 0.1	λ = 0.5
A	24	0.074	0.412	0.068	0.387	0.094	0.635
	48	0.090	0.950	0.105	0.907	0.131	0.989
	100	0.229	1.000	0.258	1.000	0.411	1.000
	200	0.675	1.000	0.702	1.000	0.916	1.000
B	24	0.080	0.485	0.087	0.459	0.095	0.592
	48	0.183	0.980	0.131	0.947	0.112	0.982
	100	0.529	1.000	0.413	1.000	0.259	1.000
	200	0.977	1.000	0.952	1.000	0.722	1.000
C	24	0.089	0.475	0.112	0.457	0.093	0.345
	48	0.152	0.954	0.120	0.915	0.083	0.813
	100	0.305	1.000	0.223	1.000	0.156	1.000
	200	0.733	1.000	0.649	1.000	0.497	1.000
D	24	0.094	0.549	0.112	0.530	0.115	0.312
	48	0.211	0.981	0.156	0.956	0.108	0.731
	100	0.533	1.000	0.396	1.000	0.183	0.997
	200	0.977	1.000	0.946	1.000	0.389	1.000

Cases: A: γ=0 , σ_1=0.01 , B: γ=0.1 , σ_1=0.01 ,
C: γ=0 , σ_1=0.05 , D: γ=0.1 , σ_1=0.05 .

For the Wald test, we recommend the detrended version F^*, because the standard version does not have asymptotically appropriate critical values, independent of nuisance parameters, a problem which is also reflected in the small sample significance levels.

Next, we compare the power of the DW, ADF and F^* test for the same cases that we considered for comparison of the estimators, listed in table 4. First the two residual based tests are compared. Although the DW test on average has higher power than the ADF test, in most cases the difference between them is not very large. The higher divergence rate of T DW (see theorem 4) does not lead to a sharper increase in power of this test.

An increase in λ has a positive effect on the power of all tests. This could be expected, because a faster adjustment rate implies that y_t and x_t move together more closely. In this case, a sample size of 100 is sufficient for all testing procedures to attain a 100% rejection frequency. The parameters γ and σ_1 again have an asymmetrical effect on the performance of the static and the dynamic regression procedure. First, the presence of an error correction term in the equation for Δx_t raises the power the DW and ADF tests, whereas it lowers the power of the F test. Second, an increase of the standard error σ_1 doesn't effect the residual based tests very much, but again leads to a worse performance of the Wald test.

Because of these effects, the F test only has higher power than the other tests if σ_1=0.01 and γ=0; in the other cases the residual based tests are superior. We argued before that the lower value of the standard error may be more relevant for log-linear models; whether or not x_t may be considered weakly exogenous (so that γ=0) depends entirely on the variables being modelled.

6. *Conclusions*

In this paper we have compared the asymptotic and small sample behavior of two long run multiplier estimators and three tests on cointegration. The asymptotic properties of the two estimators do not differ very much; both are consistent and have the same convergence rate. In small samples, the static regression estimator often has a higher bias; in terms of root mean squared error however, it often is superior to the dynamic regression estimator. This measure may be especially unfavorable for the latter estimator because it has no finite moments.

For the test statistics, we have argued that the use of detrended data is appropriate for the Wald test in the dynamic regression equation, because it leads to a parameter-free asymptotic distribution; for the residual based tests it is not recommended. In a small sample power comparison the DW and ADF tests generally perform better. If we restrict ourselves to the cases that are regarded most relevant, the decision which one of the testing procedures is superior depend on the question whether or not the regressor x_t is weakly exogenous for the long run parameter. If it isn't, the static regression procedure generally performs better. However, in this case full information procedures seem to be more appropriate.

References

BANERJEE, A., J. J. DOLADO, D. F. HENDRY AND G. W. SMITH (1986), "Exploring Equilibrium Relationships in Econometrics through Static Models: Some Monte Carlo Evidence", *Oxford Bulletin of Economics and Statistics*, 48, 253-277.

BEWLEY, R. A. (1979), "The direct Estimation of the Equilibrium Response in a Linear Dynamic Model", *Economics Letters*, 3, 357-362.

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

DAVIDSON, J. E. H., D. F. HENDRY, F. SRBA AND S. YEO (1978), "Econometric Modelling of the Aggregate Time Series Relationship Between Consumer's Expenditure and Income in the United Kingdom", *Economic Journal*, 88, 661-692.

DICKEY, D. A. AND W. A. FULLER (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", *Journal of the American Statistical Association*, 74, 427-431.

ENGLE, R. F. AND C. W. J. GRANGER (1987), "Cointegration and Error Correction: Representation, Estimation, and Testing", *Econometrica*, 55, 251-276.

ENGLE, R. F. AND B. S. YOO (1987), "Forecasting and Testing in Cointegrated Systems", *Journal of Econometrics*, 35, 143-159.

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

GRANGER, C. W. J. (1981), Some Properties of Time Series Data and Their Use in Econometric Model Specification", *Journal of Econometrics*, 16, 121-130.

HENDRY, D. F. AND J.-F. RICHARD (1983), "The Econometric Analysis of Economic Time Series", *International Statistical Review*, 51, 111-163.

JOHANSEN, S. (1988), "Statistical Analysis of Cointegration Vectors", *Journal of Economic Dynamics and Control*, 12, 231-254.

KINAL, T. W. (1980), "The Existence of Moments of k-class Estimators", *Econometrica*, 48, 241-249.

PARK, J. Y. AND P. C. B. PHILLIPS (1988), "Statistical Inference in Regressions with Integrated Processes: Part 1", *Econometric Theory*, 4, 468-497.

PARK, J. Y. AND P. C. B. PHILLIPS (1989), "Statistical Inference in Regressions with Integrated Processes: Part 2", *Econometric Theory*, 5, 95-131.

PHILLIPS, P. C. B. (1986), "Understanding Spurious Regressions in Econometrics", *Journal of Econometrics*, 33, 311-340.

PHILLIPS, P. C. B. (1988), "Weak Convergence to the Matrix Stochastic Integral $\int_0^1 BdB'$ ", *Journal of Multivariate Analysis*, 24, 252-264.

PHILLIPS, P. C. B. AND S. N. DURLAUF (1986), "Multiple Time Series Regression with Integrated Processes", *Review of Economic Studies*, 53, 473-495.

PHILLIPS, P. C. B. AND S. OULIARIS (1988a), "Asymptotic Properties of Residual Based Tests for Cointegration", Cowles Foundation Discussion Paper No. 847-R, forthcoming in *Econometrica*.

PHILLIPS, P. C. B. AND S. OULIARIS (1988b), "Testing for Cointegration using Principal Components Methods", *Journal of Economic Dynamics and Control*, 12, 205-230.

SAID, S. E. AND D. A. DICKEY (1984), "Testing for Unit Roots in Autoregressive Moving Average Models of Unknown Order", *Biometrika*, 71, 599-607.

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

STOCK, J. H. (1987), "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors", *Econometrica*, 55, 1035-1056.

STOCK, J. H. AND M. W. WATSON (1988), "Testing for Common Trends", *Journal of the American Statistical Association*, 83, 1097-1107.

WHITE, H. AND I. DOMOWITZ (1984), "Nonlinear Regression with Dependent Observations", *Econometrica*, 52, 643-661.

WICKENS, M. R. AND T. S. BREUSCH (1988), "Dynamic Specification, the Long-Run and the Estimation of Transformed Regression Models", *The Economic Journal*, 98, 189-205.

Appendix A

Proof of lemma 1 :

Follows from corollary 2.2 of PHILLIPS AND DURLAUF (1986). Because u_t is a stationary Gaussian process, the condition on the higher moments is satisfied; moreover, for such a process the autocovariances decrease exponentially, which implies that the condition on the mixing numbers is satisfied.

Proof of lemma 2 :

See BILLINGSLEY (1968), Corollary 1, p.31 .

Proof of lemma 3 :

Statements (a), (b) and (c) are given in Lemma 3.1 of PHILLIPS AND DURLAUF (1986); part (c) is proved more rigorously in PHILLIPS (1988). Parts (d) and (e) are given in Lemma 2.1 in PHILLIPS AND PARK (1988).

Proof of theorem 1 :

The results for the static regression estimator $\hat{\theta}_T$ follow from theorems 3.2 and 3.6 of PARK AND PHILLIPS (1988).

We shall now prove the results for the dynamic regression estimator. Let $q_t = (\Delta y_{t-1},...,\Delta y_{t-p+1}, \Delta x'_t,...,\Delta x'_{t-p+1})'$ and $\psi = (\varphi^*_1,...,\varphi^*_{p-1}, \beta'_0, \beta^{*\prime}_1,...,\beta^*_{p-1}{}')'$. Because $GG' = g_1 g'_1 + G_2 G'_2 = I_N$, we can write equation (16) as

$$\Delta y_t = \kappa + \lambda' z_{t-1} + \psi' q_t + \eta_t ,$$

$$= \kappa + (\lambda' g_1)(g'_1 z_{t-1}) + (\lambda' G_2)(G'_2 z_{t-1}) + \psi' q_t + \eta_t . \qquad (A1)$$

We shall first derive the asymptotic properties of the transformed estimators $g'_1\hat{\lambda}_T$ and $G'_2\hat{\lambda}_T$. Because g_1 is the normalized cointegrating vector, the transformed regressor variable $g'_1 z_{t-1}$ is stationary. The regressor vector q_t is also stationary, because its components are all first differences of I(1) processes. The regressor $G'_2 z_{t-1}$ however is an N-1 dimensional integrated process (with drift if $\mu_G \neq 0$). In such a regression we can apply theorems 3.1 and 3.2 of PARK AND PHILLIPS (1989), which tell us that for $g'_1\hat{\lambda}_T$ and $\hat{\psi}_T$, the classical properties apply. In particular, we have:

$$g'_1\hat{\lambda}_T \xrightarrow{p} g'_1\lambda . \qquad (A2)$$

From these theorems it also follows that the estimator $G_2'\hat{\lambda}_T$ has asymptotic properties, similar to $\hat{\theta}_T$: if $\mu \neq 0$ and N=2, then $G_2'z_{t-1}$ asymptotically behaves as the linear trend $\mu_G t$, where μ_G is in this case equal to $(1+\theta^2)^{1/2}\mu_x$, because cointegration requires $\mu_y = \theta'\mu_x$; this implies

$$T^{3/2}G_2'(\hat{\lambda}_T - \lambda) \quad \Rightarrow \quad \frac{12\sigma}{\mu_G}\int_0^1 (r-\tfrac{1}{2})dW(r) \quad \sim \quad N(0,12\sigma^2/(1+\theta^2)\mu_x^2) \,. \tag{A3}$$

If μ=0 or N>2, then

$$G_2'(\hat{\lambda}_T - \lambda) \;=\; O_p(T^{-1}) \,. \tag{A4}$$

In both cases $G_2'\hat{\lambda}_T$ converges in probability to $G_2'\lambda$; combining this with (A2), and applying the continuous mapping theorem, we have

$$\hat{\lambda}_T \;=\; G\begin{bmatrix} g_1'\hat{\lambda}_T \\ G_2'\hat{\lambda}_T \end{bmatrix} \quad \xrightarrow{p} \quad \lambda \,. \tag{A5}$$

With these results, we can derive the asymptotic properties of $\tilde{\theta}_T$ as follows:

$$\begin{aligned} \tilde{\theta}_T - \theta \quad &= (-\hat{\lambda}_{2T}/\hat{\lambda}_{1T} - \theta) \\ &= -(\hat{\lambda}_{2T} + \theta\hat{\lambda}_{1T})/\hat{\lambda}_{1T} \\ &= -(\theta : I_{N-1})\hat{\lambda}_T/\hat{\lambda}_{1T} \quad = \; - (I_{N-1} + \theta\theta')^{1/2}(G_2'\hat{\lambda}_T) \;/\; \hat{\lambda}_{1T} \,, \end{aligned} \tag{A6}$$

which leads, by substitution of the limits of $G_2'\hat{\lambda}_T$ and $\hat{\lambda}_{1T}$, to the required results. □

Proof of lemma 4 :
Statement (c) is given by PHILLIPS (1986), theorem 2(a). Statement (d) is a straightforward extension of this result; the main effect of a trend term included in the regression is the elimination of possible trends in the regressor x_t; as a side effect, it leads to expressions in terms of detrended Brownian motions.
For part (a), note that y_t and x_t are scalar processes, consisting of a linear trend, a random walk and a stationary part. Because the linear trend dominates the rest we have

$$T^{-3}\sum_{t=1}^{T}(x_t-\bar{x})^2 \;=\; \mu_x^2/12 \;+\; o_p(1) \,, \tag{A7}$$

$$T^{-3}\sum_{t=1}^{T}(x_t-\bar{x})y_t \;=\; \mu_x\mu_y/12 \;+\; o_p(1) \,, \tag{A8}$$

so $\hat{\theta}_T$ converges to μ_y/μ_x $(=\pi)$. Now let $y_t^* = y_t - \pi' x_t$, a driftless I(1) process. For this process, the assumptions of lemma 1 apply, so that

$$T^{-3/2} \sum_{t=1}^{T} (x_t - \bar{x}) y_t^* = T^{-3/2} \mu_x \sum_{t=1}^{T} (t - \tfrac{1}{2}T) y_t^* + o_p(1)$$

$$\Rightarrow \mu_x \int_0^1 (r - \tfrac{1}{2}) B_{21}(r) dr , \tag{A9}$$

which leads, together with (A7), to part (a).

For statement (b), we rewrite the static regression as

$$y_t^* = \nu + (\theta - \pi)' x_t + u_t$$

$$= \nu + (\theta - \pi)' H H' x_t + u_t$$

$$= \nu + ((\theta - \pi)' h_1)(h_1' x_t) + ((\theta - \pi)' H_2)(H_2' x_t) + u_t . \tag{A10}$$

Now $(y_t^*, x_t' H_2)' = R' z_t$ is an (N-1) dimensional integrated process with no trend, satisfying the assumptions of lemma 1. For the regressor $h_1' x_t$, the linear trend dominates the other elements. Using the techniques of partitioned regression and the continuous mapping theorem, it is not hard to show that

$$\sqrt{T}\, h_1'(\hat{\theta}_T - \pi) \Rightarrow (\mu_x' \mu_x)^{-1/2} \left(\int_0^1 P(r)^2 dr \right)^{-1} \int_0^1 P(r) B_{21}^*(r) dr , \tag{A11}$$

and

$$H_2'(\hat{\theta}_T - \pi) \Rightarrow \left(\int_0^1 B_{22}^{**}(r) B_{22}^{**}(r)' dr \right)^{-1} \int_0^1 B_{22}^{**}(r) B_{21}^{**}(r) dr , \tag{A12}$$

where $P(r)$ is the limiting process for $h_1' x_t$, corrected for $H_2' x_t$ and the constant term, and $B_{22}^{**}(r)$ is the limiting process for $H_2' x_t$, corrected for $h_1' x_t$ and the constant term. The second part of statement (b) is just (A11) with both sides multiplied by $(\mu_x' \mu_x)^{1/2}$, and the first part follows from

$$\hat{\theta}_T - \pi = h_1 h_1'(\hat{\theta}_T - \pi) + H_2 H_2'(\hat{\theta}_T - \pi)$$

$$= H_2 H_2'(\hat{\theta}_T - \pi) + o_p(1) , \tag{A13}$$

which concludes the proof. □

Proof of theorem 2 :
The first part of (c), concerning the Durbin-Watson statistic, is from PHILLIPS (1986), theorem 2(f). The first part of statement (d) follows directly from (c); the only difference is that the functional is now in terms of a detrended vector Brownian motion, as a consequence of the trend term included in the regression.
The second part of (c) and (d), concerning the Augmented Dickey-Fuller test statistic, has been proved by PHILLIPS AND OULIARIS (1988a). Note that the asymptotic distribution is free of nuisance parameters, because the lag length q is asymptotically sufficient to correct for the serial correlation of $\Delta\hat{u}_t$.
The results of part (a) and (b) are not straightforward extensions, because the linear trend forces certain moment matrices to be singular, which leads to the reduction of the dimension of the relevant vector Brownian motions. We define the partial regressor vector $w_t = (\Delta\hat{u}_{t-1},\ldots,\Delta\hat{u}_{t-q})$, and the following sequences (summations are over the relevant sample sizes):

$$a_T = T^{-1}\sum_t \Delta\hat{u}_t^2 = \hat{\alpha}_T'\,(\,T^{-1}\sum_t (\Delta z_t-\Delta\bar{z})(\Delta z_t-\Delta\bar{z})'\,)\,\hat{\alpha}_T\,, \tag{A14}$$

$$b_T = T^{-2}\sum_t \hat{u}_t^2 = \hat{\alpha}_T'\,(\,T^{-2}\sum_t (z_t-\bar{z})(z_t-\bar{z})'\,)\,\hat{\alpha}_T\,, \tag{A15}$$

$$c_T = T^{-1}\sum_t \hat{u}_{t-1}\Delta\hat{u}_t - T^{-1}\sum_t \hat{u}_{t-1}w_t'\,(\sum_t w_t w_t')^{-1}\sum_t w_t\Delta\hat{u}_t\,, \tag{A16}$$

$$d_T = T^{-2}\sum_t \hat{u}_{t-1}^2 - T^{-1}\sum_t \hat{u}_{t-1}w_t'\,(\sum_t w_t w_t')^{-1}\sum_t w_t\hat{u}_{t-1}\,, \tag{A17}$$

With these, we can express the test statistics as:

$$T\,DW = a_T/b_T\,, \tag{A18}$$

$$ADF = c_T/(\hat{\sigma}_{eT}\sqrt{d_T})\,, \tag{A19}$$

where $\hat{\sigma}_{eT}$ is the square root of the residual variance of the auxiliary regression (19). For the limit of a_T, we use the law of large numbers for mixing sequences (see e.g. WHITE AND DOMOWITZ (1984)) to establish that

$$T^{-1}\sum_t (\Delta z_t-\Delta\bar{z})(\Delta z_t-\Delta\bar{z})' \xrightarrow{p} \Sigma_0\,, \tag{A20}$$

which leads, by application of the continuous mapping theorem, to

$$a_T \Rightarrow \zeta_i'\Sigma_0\zeta_i\,, \tag{A21}$$

where i=1 for case (a) and i=2 for (b). In order to establish the limits for the sequences b_T, c_T, and d_T, we consider the asymptotic behavior of the following process on [0,1]:

$$U_T(r) = \frac{1}{\sqrt{T}}\hat{u}_{[rT]} \,. \tag{A22}$$

Because $z_t = z_0 + \mu_t + S_t$ (see equation (25)), we have

$$\hat{u}_t = \hat{\alpha}_T'(z_t-\bar{z}) = \hat{\alpha}_T'\mu(t-\tfrac{1}{2}T) + \hat{\alpha}_T'(S_t - \bar{S}) \,. \tag{A23}$$

Next, note that $\mu'\hat{\alpha}_T = \mu_y - \mu_x'\hat{\theta}_T = -\mu_x'(\hat{\theta}_T-\pi)$, and that $T^{-1/2}S_{[rT]} \Rightarrow B_1(r)$. Therefore, we have

$$U_T(r) \Rightarrow -\xi_3(r-\tfrac{1}{2}) + \zeta_2'B_1^*(r) = U(r) \,. \tag{A24}$$

Because of the definition of ξ_3 and ζ_2, this is equal to

$$\begin{aligned} U(r) &= (1,-\pi)'B_1^*(r) - (0, \xi_2')B_1^*(r) - \xi_3(r-\tfrac{1}{2}) \\ &= B_{21}^*(r) - B_{22}^*(r)' \left(\int_0^1 B_{22}^{**}(s)B_{22}^{**}(s)'ds\right)^{-1} \int_0^1 B_{22}^{**}(s)B_{21}^{**}(s)ds \\ &\quad - (r-\tfrac{1}{2}) \int_0^1 P(s)^2ds \int_0^1 P(s)B_{21}(s)ds \,. \end{aligned} \tag{A25}$$

If N=2, then the dimension of $B_{22}(r)$ is zero, so that the second term vanishes and P(r) simplifies to $(r-\frac{1}{2})$. In that case U(r) is equal to the Brownian motion $B_{21}(r)$, corrected for a mean and a trend, or $B_{21}^{**}(r)$. If N>2, U(r) is equal to $B_{21}(r)$, corrected for a mean, a trend, and for $B_{22}(r)$, which we call $B_{21\bullet 2}^{**}(r)$. Having established this, we can follow the line of the argument in PHILLIPS AND OULIARIS (1988a) to prove the results of theorem 2. First, we have

$$b_T \Rightarrow \int_0^1 U(r)^2dr \,, \tag{A26}$$

where U(r) is either $B_{21}^{**}(r)$ (in case (a)) or $B_{21\bullet 2}^{**}(r)$ (in case (b)); together with (A21) this gives us the required results for the DW statistic. Next, note that π in the auxiliary regression (19) is equal to 0, so that

$$c_T = T^{-1} \sum_t \hat{u}_{t-1}e_t - T^{-1}\sum_t \hat{u}_{t-1}w_t' \left(\sum_t w_tw_t'\right)^{-1} \sum_t w_te_t \,, \tag{A27}$$

where e_t is a white noise process, provided that q is sufficiently large:

$$e_t = \Delta\hat{u}_t - \sum_{j=1}^{q} \delta_j\Delta\hat{u}_{t-j} = \delta(L)\Delta\hat{u}_t \,. \tag{A28}$$

Therefore we have

$$c_T \quad \Rightarrow \quad \int_0^1 U(r)\, dU(r)\, \delta(1)\ , \tag{A29}$$

$$\hat{\sigma}_{eT} \quad \Rightarrow \quad (\varsigma_i'\Sigma\varsigma_i)^{1/2}\delta(1)\ , \tag{A30}$$

where $\varsigma_i'\Sigma\varsigma_i$, i=1,2, is the variance of U(r), conditional upon ς_i (PHILLIPS AND OULIARIS (1988a) discuss the conditions under which such conditioning is possible). Finally note that

$$d_T \quad = \quad b_T + o_p(1) \quad \Rightarrow \quad \int_0^1 U(r)^2 dr\ , \tag{A31}$$

and the rest of the proof is analogous to PHILLIPS AND OULIARIS (1988a). □

Proof of theorem 3 :
We define the partial regressor vectors $q_t^* = (1,\ q_t')'$ and $q_t^{**} = (1,\ t,\ q_t')'$. Next, we use the observations matrices $Q_1 = (q_1^*,...q_T^*)'$, $Q_2 = (q_1^{**},...q_T^{**})'$, $Z_{-1} = (z_0,...,z_{T-1})'$ and $\eta = (\eta_1,...,\eta_T)'$. We also define the projection matrices $M_1 = (I_T - Q_1(Q_1'Q_1)^{-1}Q_1')$ and $M_2 = (I_T - Q_2(Q_2'Q_2)^{-1}Q_2')$. Then we can express the Wald test statistics (under the null hypothesis) as:

$$F \ = \ \eta' M_1 Z_{-1}(Z_{-1}'M_1Z_{-1})^{-1}Z_{-1}'M_1\eta \ /\ \hat{\sigma}^2\ , \tag{A32}$$

$$F^* \ = \ \eta' M_2 Z_{-1}(Z_{-1}'M_2Z_{-1})^{-1}Z_{-1}'M_2\eta \ /\ \hat{\sigma}^{*2}\ , \tag{A33}$$

where $\hat{\sigma}^2$ and $\hat{\sigma}^{*2}$ are the estimated variances of η_t in (16) and (16) with a trend term added, respectively.
For part (a), we use the matrix $K = (T^{-1}k_1,\ T^{-1/2}K_2)$, where k_1 is the normalized drift vector $(\mu'\mu)^{-1/2}\mu$ and K_2 is orthogonal to μ. Note that K is non-singular, so that we can express F as

$$F \ = \ \eta' M_1 Z_{-1}K(K'Z_{-1}'M_1Z_{-1}K)^{-1}K'Z_{-1}'M_1\eta \ /\ \hat{\sigma}^2\ . \tag{A34}$$

Now $k_1'z_t$ is a scalar integrated process with drift, divided by T :

$$k_1'z_t \ = \ (\mu'\mu)^{1/2}t/T \ + \ k_1'S_t/T \ + \ k_1'z_0/T\ , \tag{A35}$$

so that the trend term dominates the stochastic part. $K_2'z_t$ on the other hand is an (N-1) dimensional integrated process with no drift, devided by $\sqrt{T}$, satisfying the assumptions of lemma 1, so

$$Z_T(r) \;=\; K_2' z_{[rT]} \;\;\Rightarrow\;\; K_2' B_1(r) \,. \qquad \text{(A36)}$$

Now it can be shown, using the convergence rates of lemma 3, that, asymptotically, the only effect of M_1 on Z_{-1} is that the limiting processes will be in deviations from their mean, while the asymptotical effect of M_2 is that the limiting process will be demeaned and detrended. If we define $\underline{z}_{t-1}'$ as the t-th row of M_1Z_{-1}, we have

$$M_T(r) \;=\; K'\underline{z}_{[rT]} \;\;\Rightarrow\;\; M(r) \,. \qquad \text{(A37)}$$

Next, we consider the partial sum process P_t, defined by

$$P_t \;=\; \sum_{s=1}^{t} \eta_s \;=\; (1,-\omega_{21}'\Omega_{22}^{-1}) \sum_{s=1}^{t} \varepsilon_s \,, \qquad \text{(A38)}$$

where the second equality follows from the definition of η_t :

$$\eta_t \;=\; y_t \;-\; E\left[\; y_t \mid x_t, F_{t-1} \;\right] , \qquad \text{(A39)}$$

together with the definition of ε_t (see equation (3)) and the normality of z_t. If we compare P_t with S_t, the stochastic part of z_t (see (9)):

$$S_t \;=\; C(1) \sum_{s=1}^{t} \varepsilon_s \;+\; C^*(L)\varepsilon_t \,, \qquad \text{(A40)}$$

it is clear that the limiting processes of P_t and S_t will be defined from the same vector Brownian motion $B_1(r) = \Sigma^{1/2}W_1(r)$. In particular, we have

$$H_T(r) \;=\; T^{-1/2}\, P_{[rT]} \;\;\Rightarrow\;\; (1,-\omega_{21}'\Omega_{22}^{-1})\, C(1)^{-1}B_1(r)$$

$$=\;\; (1,-\omega_{21}'\Omega_{22}^{-1})\Omega^{1/2}W_1(r) \;\;=\;\; \sigma\, s'W_1(r) \,, \qquad \text{(A41)}$$

where the vector s is defined in theorem 3.
Finally, it is easily seen that

$$\hat{\sigma}^2 \;=\; T^{-1}\eta'\eta \;+\; o_p(1) \;\;\xrightarrow{p}\;\; \sigma^2 \,, \qquad \text{(A42)}$$

$$\hat{\sigma}^{*2} \;\;\xrightarrow{p}\;\; \sigma^2 \,. \qquad \text{(A43)}$$

With these results, we can apply the continuous mapping theorem to show that, if $\mu \neq 0$,

$$T^{-1} K'Z_{-1}'M_1Z_{-1}K \Rightarrow \int_0^1 M(r)M(r)'dr , \quad (A44)$$

$$T^{-1/2} K'Z_{-1}'M_1\eta \Rightarrow \int_0^1 M(r)\, dW_1(r)'s\, \sigma . \quad (A45)$$

If $\mu=0$, then

$$T^{-2} Z_{-1}'M_1Z_{-1} \Rightarrow \int_0^1 B_1^*(r)B_1^*(r)'dr , \quad (A46)$$

$$T^{-1} Z_{-1}'M_1\eta \Rightarrow \int_0^1 B_1^*(r)\, dW_1(r)'s\, \sigma , \quad (A47)$$

whereas, irrespective of μ, we have

$$T^{-2} Z_{-1}'M_2Z_{-1} \Rightarrow \int_0^1 B_1^{**}(r)B_1^{**}(r)'dr , \quad (A48)$$

$$T^{-1} Z_{-1}'M_2\eta \Rightarrow \int_0^1 B_1^{**}(r)\, dW_1(r)'s\, \sigma . \quad (A49)$$

Part (a) is now easily proved, using (A34), (A42), (A44) and (A45). For part (b), we now have

$$F \Rightarrow s' \int_0^1 dW_1^*(r)\, W_1^*(r)'\Sigma^{1/2} \left(\Sigma^{1/2} \int_0^1 W_1^*(r)W_1^*(r)'dr\, \Sigma^{1/2} \right)^{-1} \Sigma^{1/2} \int_0^1 W_1^*(r)\, dW_1^*(r)'s$$

$$= s' \int_0^1 dW_1^*(r)\, W_1^*(r)' \left(\int_0^1 W_1^*(r)W_1^*(r)'dr \right)^{-1} \int_0^1 W_1^*(r)\, dW_1^*(r)'s , \quad (A50)$$

because Σ is non-singular. From the definition of s and σ, it is clear that $s's = 1$. We define an orthogonal matrix S which has s as its first column, so $s = Se_1$, where e_j is an N-vector with N-1 components equal to 0 and the j-th component equal to 1. We also define the vector Brownian motion $V(r) = S'W_1(r)$ and note that its covariance matrix is equal to $S'I_NS = I_N$, because S is orthogonal. From (A50), we can now proceed to obtain

$$F \Rightarrow e_1'S' \int_0^1 dW_1^*(r)\, W_1^*(r)'S\left(S' \int_0^1 W_1^*(r)W_1^*(r)'drS \right)^{-1}S' \int_0^1 W_1^*(r)\, dW_1^*(r)'Se_1$$

$$= e_1' \int_0^1 dV^*(r)\, V^*(r)' \left(\int_0^1 V^*(r)V^*(r)'dr \right)^{-1} \int_0^1 V^*(r)\, dV^*(r)'\, e_1 . \quad (A51)$$

Because V(r) is an N dimensional Brownian motion with covariance matrix I_N, it has the same distributional properties as $W_1(r)$, so as far as the distribution of F is concerned we can replace $V^*(r)$ in (A51) by $W_1^*(r)$, which results in statement (b).
The proof of part (c) is entirely analogous, with demeaned Brownian motion replaced by their detrended versions. □

Proof of theorem 4 :
For part (a) and (b), we use the superconsistency of $\hat{\theta}_T$ to establish that

$$m_j = T^{-1}\sum_t \hat{u}_t\hat{u}_{t-j} = T^{-1}\sum_t u_t u_{t-j} + o_p(1) , \qquad \text{(A52)}$$

$$m_j^* = T^{-1}\sum_t \hat{u}_t^*\hat{u}_{t-j}^* = T^{-1}\sum_t u_t u_{t-j} + o_p(1) , \qquad \text{(A53)}$$

for j=0,1,... . Because u_t is a stationary process, all sample moments m_j and m_j^* are $O_p(1)$ and converge in probability to Eu_tu_{t-j} . For the Durbin-Watson statistic, we have

$$DW = 2(1-m_1/m_0) \xrightarrow{p} 2(1-\rho) , \qquad \text{(A54)}$$

where $\rho = (Eu_tu_{t-1})/(Eu_t^2)$; for DW^*, the same convergence applies.
The divergence rate of the ADF statistic is proved in PHILLIPS AND OULIARIS (1988a), theorem 5.1. The proof for the ADF^* statistic is completely analogous.
For the Wald test statistic, we use the same notation as in the proof of theorem 3 to express F and F^* as

$$F = \hat{\lambda}_T'(Z_{-1}'M_1Z_{-1})\hat{\lambda}_T/\hat{\sigma}^2 , \qquad \text{(A55)}$$

$$F^* = \hat{\lambda}_T^{*\prime}(Z_{-1}'M_2Z_{-1})\hat{\lambda}_T^*/\hat{\sigma}^{*2} . \qquad \text{(A56)}$$

In theorem 1 we proved that $-\hat{\lambda}_T/\hat{\lambda}_{1T} = -(1,-\hat{\theta}_T')'$ is a superconsistent estimator of the cointegrating vector. Using the same methods this can be proved for the detrended version $\hat{\lambda}_T^*$. Therefore $\hat{\lambda}_T'z_{t-1}$ and $\hat{\lambda}_T^{*\prime}z_{t-1}$ are asymptotically proportional to u_{t-1}. The consistency of $\hat{\sigma}^2$ and $\hat{\sigma}^{*2}$ is also straightforward to prove. If we let U_{-1} denote $(u_0,...,u_{T-1})'$ then we have, because u_t is stationary,

$$F = U_{-1}'M_1U_{-1}\lambda_1^2/\sigma^2 + o_p(1) = O_p(T), \qquad \text{(A57)}$$

$$F_* = U_{-1}'M_2U_{-1}\lambda_1^2/\sigma^2 + o_p(1) = O_p(T), \qquad \text{(A58)}$$

which ends the proof. □

Appendix B

In tables B1-B3 the estimated critical values for the Wald test on cointegration are reported. Table B1 corresponds to a Wald test in a regression without a constant term; tables B2 and B3 apply to the test statistics F and F^*, respectively. The critical values of B1 and B2 can only be used if the integrated processes do not have deterministic trends. The tables allow for 2 to 5 variables in the cointegrating relationship. N denotes the number of variables in the system, and α is the significance level. For N=1, the distribution of the square of $\hat{\tau}$, $\hat{\tau}_\mu$ and $\hat{\tau}_\tau$ is tabulated, see FULLER (1976, p.373). Critical values were estimated using the Monte Carlo method, with 10000 replications and a sample size (T) of 500.

Table B1 : Asymptotic critical values for the Wald (F) test; standard

N \ α	0.15	0.10	0.075	0.05	0.025	0.01
1	2.35	3.00	3.42	4.09	5.24	7.03
2	5.44	6.38	7.04	7.93	9.60	11.67
3	8.22	9.32	10.17	11.41	13.17	15.23
4	10.84	12.16	12.95	14.12	16.04	18.63
5	13.30	14.72	15.68	16.93	19.03	21.52

Table B2 : Asymptotic critical values for the Wald (F) test; demeaned

N \ α	0.15	0.10	0.075	0.05	0.025	0.01
1	5.56	6.49	7.20	8.09	9.44	11.53
2	8.30	9.38	10.10	11.20	12.85	15.24
3	10.82	12.19	13.15	14.24	16.07	18.31
4	13.37	14.83	15.73	16.92	18.98	21.48
5	15.79	17.25	18.18	19.53	21.68	24.14

Table B3 : Asymptotic critical values for the Wald (F) test; demeaned and detrended

N \ α	0.15	0.10	0.075	0.05	0.025	0.01
1	8.58	9.73	10.55	11.57	13.29	15.55
2	11.12	12.36	13.22	14.45	16.39	18.51
3	13.49	14.91	16.01	17.28	19.07	21.39
4	15.88	17.39	18.45	19.67	21.60	24.40
5	18.18	19.65	20.67	22.15	24.53	27.54