

Multivariate Unit Root Tests, Stability and Convergence

Andrew Harvey and David Bates

Faculty of Economics and Politics, University of Cambridge

April 7, 2004

Abstract

We examine the properties of a multivariate Dickey-Fuller t -statistic designed to test for a unit root in a panel while taking account of cross-correlations. The asymptotic distribution is given and asymptotic and finite sample critical values provided. When intercepts are present, a modification along the lines of Elliot, Rothenberg and Stock (Econometrica, 1996) can be implemented. The tests have invariance properties and can be used to test for stability or balanced growth. Non-zero initial conditions actually boost the power of the (unmodified) Dickey-Fuller tests confirming that they are useful for testing the hypothesis that the series are in the process of converging. Typical applications are for a moderate number of series observed over a reasonably long period of time. The example given is for the per capita incomes of six US regions observed annually from 1950 to 1999.

KEYWORDS: Balanced growth; cross-sectional dependence; Dickey-Fuller test, initial conditions, power envelope; stationarity tests.

JEL classification: C32, O40

1 Introduction

This paper discusses the role of unit root tests in assessing stability and convergence in cross-sections of time series. The typical applications are for a moderate number of

series, perhaps countries or regions, observed over a reasonably long period of time.

One of the reasons why unit root tests fail to reject is because they lack power. More specifically, the Dickey-Fuller (DF) test with constant included, which is the test most commonly applied, lacks power. This can be rectified, to some extent, by dropping the constant when it is appropriate to do so, or by making a modification along the lines suggested by Elliot, Rothenberg and Stock (1996), hereafter ERS. A further line of attack is to conduct a test on several series. There is a considerable literature on how to combine univariate tests for a large panel; see, for example, Bhargava (1986, p 378-9), Evans and Karras (1996), Maddala and Wu (1999), Levin, Lin and Lu (2002), Smith et al (2002) and Im, Pesaran and Shin (2003). These studies assume that the units can be treated as though they were independent of each other. Our concern is with situations where there is a moderate number of series and the independence assumption is not a reasonable one. Ignoring cross-sectional dependence can lead to considerable distortion in the size of tests as demonstrated in the context of purchasing power parity (PPP) by O'Connell (1998). O'Connell follows Abuaf and Jorion (1990) in proposing a unit root test¹ based on a homogeneous first-order autoregressive model that becomes a multivariate random walk under the null hypothesis and contains only one extra parameter under the stationary alternative. We study the properties of this Wald test, which we call the multivariate homogeneous Dickey-Fuller (MHDF) test, and introduce the corresponding 'LM type' test, thereby generalising the work of Sargan and Bhargava (1983). When a constant is present, we generalise the modified DF test proposed for univariate models by ERS. By analysing local power functions and investigating test sizes in finite samples we are able to reach some conclusions on how the autoregressive parameter should be set when residuals are obtained by generalised least squares detrending. We note that the homogeneous model is an attractive one to adopt since it leads to simple procedures that have certain invariance properties.

The tests described in the previous paragraph are relevant, not only for studies of PPP, but also for applications on the mean reversion of inflation rates and real wages, as in Culver and Papell (1997) and Lee and Wu (2001), and on the real interest rate parity hypothesis, as in Wu and Chen (1998). More generally the aim is to test for stability or balanced growth. Both unit root tests and stationarity tests may

¹Phillips and Sul (2002) generalise the test proposed by Maddala and Wu (1999) to deal with certain types of cross-dependence.

be relevant here. This situation should be carefully distinguished from one in which the series *are in the process of converging*. There is some confusion in the literature on the role of unit root tests in studies of the convergence of economic variables, such as income per capita, in different countries or regions. In contrast to cross-sectional results of the type reported in Barro and Sala-i-Martin (1995), the unit root tests used in time series studies often find little evidence for convergence. Bernard and Durlauf (1996), argue that this is because the forecast-convergence definition, by requiring output differences to be stationary, is more restrictive than the cross-sectional definition; see also Durlauf and Quah (1999, p 288). However, their argument is incorrect. While the forecast-convergence definition does indeed imply convergence to stationarity, this does not mean that convergence tests are necessarily checking for the compatibility of output differences with indeterministic stationary series. This is a matter of stability. The fact that DF tests are based on an error correction mechanism (ECM) makes them entirely appropriate for testing whether economies are in the process of converging. Although Bernard and Durlauf (1996, p 172) claim that ‘.. time series results accepting the no convergence null may be due to transitional dynamics..’, our investigation of the effects of initial conditions on power shows the opposite to be true.

The paper is organised as follows. The multivariate homogeneous Dickey-Fuller test is introduced in section 2. Its asymptotic properties, including local power, are investigated and critical values are given. An LM-type test generalising a test based on the work of Sargan and Bhargava(1983) is also considered. Section 3 deals with the tests when constants are included and sets out the ERS modification. Section 4 presents a set of Monte Carlo experiments, one purpose of which is to compare the performance of the proposed tests with an unrestricted likelihood ratio test.

Section 5 defines balanced growth and reviews how the null hypothesis of stability can be tested by the Lagrange multiplier test studied in Nyblom and Harvey (2000). Like the MHDF test, this test is derived under the assumption that there is only one extra parameter under the alternative of nonstationarity. The appropriateness of unit root and stationarity tests is discussed and the consequences of ignoring cross-correlation are explored.

The convergence section, section 6, begins with a discussion of the effects of initial conditions on the size and power of various unit root tests and the consequences for parameter values typically found in convergence studies is assessed by Monte Carlo

experiments. The case for tests based on the MHDF test is then made, with data on the per capita incomes of six United States regions observed annually from 1950 to 1999 used as an illustration. Section 7 concludes.

2 Multivariate unit root tests

Multivariate unit root tests will first be developed in the context of the model

$$\mathbf{y}_t = \Phi \mathbf{y}_{t-1} + \boldsymbol{\eta}_t, \quad t = 1, \dots, T, \quad (1)$$

where \mathbf{y}_0 is fixed, but unknown, Φ is an $N \times N$ matrix of autoregressive parameters and $\boldsymbol{\eta}_t$ is a serially uncorrelated Gaussian $N \times 1$ disturbance vector with positive definite covariance matrix Σ_η . The *homogeneous* model sets $\Phi = \phi \mathbf{I}_N$, where ϕ is a scalar, and the focus of attention is on a generalisation of the (augmented) Dickey-Fuller test, based on the t -statistic for the feasible GLS estimator of $\pi = \phi - 1$. The multivariate extension of a test derived from the work of Sargan and Bhargava (1983) is also considered.

If Φ is not restricted, a LR test of the hypothesis that Φ contains N unit roots against the alternative of stationarity can be constructed as in Johansen (1995). However, in many situations it makes little sense to let a series depend on lagged values other than its own. This has led to an interest in models in which Φ is diagonal; for example Taylor and Sarno (1998) investigate the properties of a Wald test of this hypothesis in the context of PPP. However, this heterogeneous model has some drawbacks, one of which is that the diagonality of Φ is lost when \mathbf{y}_t is pre-multiplied by a nonsingular $N \times N$ matrix. The homogeneous model, on the other hand, retains its structure and the test statistics are invariant.

In what follows all asymptotic results are for fixed N , unless explicitly stated otherwise. Although many of the tests are developed under the assumption of serially independent Gaussian disturbances, this can be weakened to martingale differences; see, for example, Stock (1994).

2.1 Multivariate homogeneous Dickey-Fuller test

The maximum likelihood estimator of the parameter π is a feasible GLS estimator

$$\tilde{\pi} = \sum_{t=2}^T \mathbf{y}'_{t-1} \tilde{\Sigma}_\eta^{-1} \Delta \mathbf{y}_t / \left[\sum_{t=2}^T \mathbf{y}'_{t-1} \tilde{\Sigma}_\eta^{-1} \mathbf{y}_{t-1} \right] \quad (2)$$

where $\tilde{\Sigma}_\eta = T^{-1} \sum_{t=2}^T (\Delta \mathbf{y}_t - \tilde{\pi} \mathbf{y}_{t-1})(\Delta \mathbf{y}_t - \tilde{\pi} \mathbf{y}_{t-1})'$. The Wald test of the null hypothesis of that $\pi = 0$ or, equivalently, $\phi = 1$, against the alternative that $\pi < 0$ is based on the ' t -statistic'

$$\tau_0(N) = \sum_{t=2}^T \mathbf{y}'_{t-1} \tilde{\Sigma}_\eta^{-1} \Delta \mathbf{y}_t / \left[\sum_{t=2}^T \mathbf{y}'_{t-1} \tilde{\Sigma}_\eta^{-1} \mathbf{y}_{t-1} \right]^{1/2}. \quad (3)$$

We will refer to this as the *multivariate homogeneous Dickey-Fuller* (MHDF) statistic. The test rejects for $\tau_0(N)$ less than a given critical value.

If desired, the estimate of π and its t -statistic can be calculated by transforming the observations in each time period by the inverse of the Cholesky decomposition of $\tilde{\Sigma}_\eta$ - that is, premultiplying \mathbf{y}_t by $\tilde{\Sigma}_\eta^{-1/2}$ - and then simply applying OLS to the pooled observations. Maximum likelihood estimation requires iterating to convergence starting with $\tilde{\pi} = 0$. However, the estimator of π will have the same asymptotic distribution if Σ_η is estimated with $\tilde{\pi}$ replaced by a consistent estimator, for example as obtained from the first iteration. We will refer to this as the *two-step* estimator.²

As in the univariate case, the unit root test can be based directly on the coefficient $\tilde{\pi}$ in (2), since the statistic $\pi_0(N) = T\tilde{\pi}$ has a known limiting distribution. The potential gain in power for a multivariate test is immediately apparent from the fact that, for a stationary model, the asymptotic variance of $\tilde{\pi}$ is the asymptotic variance in a univariate model divided by N .

2.2 Asymptotic distribution under the null hypothesis

The invariance of the test statistics, together with the fact that Σ_η is estimated consistently, means that their asymptotic properties under the null hypothesis can be

²O'Connell (1998) estimates Σ_η from first differences, that is $\tilde{\pi} = 0$; this estimator is consistent under the null and so the asymptotic distributions of the (*one-step*) estimator of π and its t -statistic will be the same as those of $\tilde{\pi}$ and $\tau_0(N)$ respectively under the null hypothesis.

derived by letting Σ_η be an identity matrix. Thus the numerator and denominator of both the coefficient and t -statistics are the sums of independent terms each of which has exactly the same form and distribution as in the univariate case. Hence the asymptotic distribution of $\tau_0(N)$ depends only on N since

$$\begin{aligned} \tau_0(N) &\rightarrow \frac{\sum_{i=1}^N \int_0^1 W_i(r) dW_i(r)}{\left[\sum_{i=1}^N \int_0^1 W_i(r)^2 dr \right]^{1/2}} = \frac{(1/2) \sum_{i=1}^N (W_i(1)^2 - 1)}{\left[\sum_{i=1}^N \int_0^1 W_i(r)^2 dr \right]^{1/2}} \\ &= \frac{(1/2)(\chi_N^2 - N)}{\{CvM_0(N)\}^{1/2}} \end{aligned} \quad (4)$$

where $W_i(r)$, $i = 1, \dots, N$ are independent standard Wiener processes and $CvM_0(N)$ is a (zero level) Cramér-von Mises with N degrees of freedom, that is

$$CvM_0(N) = \sum_{i=1}^N \int_0^1 W_i(r)^2 dr . \quad (5)$$

The simplicity of (4) is attractive. By contrast, the asymptotic distribution of the Wald test in the heterogeneous model depends on Σ_η ; see Phillips and Sul (2002).

The distribution of $\pi(N)$ is the same as in (4) except that there is no square root in the denominator.

2.3 Critical values

Table 1a shows asymptotic critical values obtained by simulating directly from (4), using 50,000 replications and approximating W_c by a discrete realisation from a sample of size 500. Simulating directly from a model in which $\Sigma_\eta = \mathbf{I}$ and $T = 500$ and then calculating $\tau_0(N)$ with $\tilde{\Sigma}_\eta$ set to \mathbf{I} in the test statistic gives virtually the same results. There is clear evidence that the distribution is approaching normality as N is increasing ; this accords with the analysis³ of Levin, Lin and Lu (2002).

³Levin, Lin and Lu (2002) prove that $\tau_0(N) \rightarrow N(0,1)$ by letting N be a function of T , such that $\sqrt{N}/T \rightarrow 0$ as $T \rightarrow \infty$.

Table 1a Asymptotic critical values for MHDF test without constant

Prob.	1	2	3	4	5	6	7	8	9	10	15	20
0.01	-2.56	-2.52	-2.55	-2.52	-2.50	-2.50	-2.50	-2.50	-2.49	-2.49	-2.47	-2.45
0.05	-1.93	-1.93	-1.91	-1.88	-1.87	-1.85	-1.82	-1.80	-1.80	-1.81	-1.77	-1.75
0.1	-1.61	-1.58	-1.57	-1.53	-1.52	-1.48	-1.46	-1.46	-1.44	-1.44	-1.41	-1.39
0.9	0.90	0.99	1.04	1.07	1.09	1.10	1.12	1.12	1.12	1.15	1.17	1.16
0.95	1.30	1.37	1.42	1.45	1.46	1.47	1.49	1.49	1.50	1.52	1.53	1.53
0.99	2.00	2.08	2.15	2.14	2.16	2.16	2.17	2.18	2.19	2.19	2.22	2.24

Table 2 shows 5% critical values for the MHDF t -test for $T = 100$ and 500 with $\Sigma_\eta = \mathbf{I}$ but $\tilde{\Sigma}_\eta$ obtained by iterating to convergence as indicated below (3). The number of replications was 300,000. The corresponding asymptotic critical values are given for comparison. As can be seen, the asymptotic critical values are larger, but not by a great deal. However, the discrepancies become larger as N increases. Thus the use of asymptotic critical values means that the tests will be slightly oversized: hence the reason for giving the table 1b, showing the critical values for $T = 100$.

Table 2 5% critical values for $\tau_0(N)$ test

N	2	5	10	20
$T = 100$	-1.94	-1.89	-1.85	-1.83
$T = 500$	-1.93	-1.87	-1.82	-1.77
Asymptotic	-1.93	-1.87	-1.81	-1.75
$T = 100, 0.9$ Off Diagonals	-1.94	-1.90	-1.85	-1.87
$T = 100, \text{Two Step}$	-1.94	-1.88	-1.86	-1.88

Table 2 also gives the critical values for the t -statistic based on the two-step estimator and for a data generation process in which the off-diagonals are set equal to 0.9. The critical values for the two-step test statistic are very similar to those for the iterated statistic shown in the first row, while the results for the non-zero off-diagonals indicates that the critical values are essentially unaffected by the form of Σ_η .

Table 1b Critical values for MHDF test for $T = 100$

Prob.	1	2	3	4	5	6	7	10	15
0.01	-2.59	-2.59	-2.59	-2.58	-2.57	-2.55	-2.55	-2.54	-2.52
0.05	-1.96	-1.94	-1.92	-1.91	-1.89	-1.87	-1.87	-1.85	-1.83
0.1	-1.63	-1.60	-1.57	-1.55	-1.53	-1.51	-1.50	-1.47	-1.45
0.9	0.91	1.03	1.09	1.13	1.17	1.19	1.21	1.28	1.35
0.95	1.31	1.42	1.48	1.53	1.57	1.60	1.62	1.70	1.80
0.99	2.06	2.18	2.22	2.27	2.35	2.38	2.40	2.53	2.69

2.4 Multivariate Sargan-Bhargava test

Following Sargan and Bhargava (1983), we consider a test in which we are led to reject the null hypothesis of $\pi = 0$ by small values of the statistic

$$\zeta_0(N) = \frac{1}{T^2} \sum_{t=1}^T \mathbf{y}'_t \widehat{\Sigma}_\eta^{-1} \mathbf{y}_t, \quad (6)$$

where $\widehat{\Sigma}_\eta = T^{-1} \sum_{t=2}^T \Delta \mathbf{y}_t \Delta \mathbf{y}'_t$. Like $\tau_0(N)$, $\zeta_0(N)$ is invariant to pre-multiplication of \mathbf{y}_t by a nonsingular $N \times N$ matrix, but it is even easier to calculate as no iterations are needed. In the univariate case, the rationale for basing a test on this statistic can be found in Stock (1994) and Tanaka (1996). Schmidt and Phillips (1992) observe that if a time trend is included a Lagrange multiplier test is obtained and appendix A shows how this applies in the multivariate case as well. With no trend, the Sargan-Bhargava (SB) test may be referred to as ‘LM type’.

Under the null hypothesis, the $\zeta_0(N)$ statistic is asymptotically distributed as $CvM_0(N)$. The lower tail now defines the critical region. The local asymptotic distribution is easily found from results for the univariate case as given in Stock (1994, p 2772). Figure 1 in ERS (1996) indicates that there is little to choose between the DF and Sargan-Bhargava tests, with both of them having local power close to the power envelope.

2.5 Serial correlation

In a univariate series, the Dickey-Fuller test extends to a more general autoregressive model in which the disturbance η_t is a $p - th$ order autoregressive process by means of a regression of Δy_t on $y_{t-1}, \Delta y_{t-j}, j = 1, \dots, p$. This is the augmented Dickey-Fuller (ADF) test. The t -test simply uses the t -statistic of y_{t-1} while the direct coefficient test is based on $T(\hat{\phi} - 1)/(1 - \sum_1^p \hat{\phi}_j^*)$ where $\hat{\phi}_j^*$ is the estimated coefficient of Δy_{t-j} . Methods of choosing the lag length, p , are discussed in Ng and Perron (1995). For the Sargan- Bhargava ζ test one option would be to base a test on the coefficient of y_{t-1} from an augmented Dickey-Fuller regression as in Oya and Toda (1998); see also Stock (1994). Another nonparametric amendment is given in Schmidt and Phillips (1992).

Serial correlation may be removed from the multivariate t -test, in the spirit of the ADF test, by first regressing $\Delta \mathbf{y}_t$ and \mathbf{y}_{t-1} on lagged values of $\Delta \mathbf{y}_t$ and then inserting the residuals into (3). A similar device is used by Johansen (1995) to remove serial correlation before constructing a likelihood ratio test for co-integration. To simplify we might just use the lags for the equation in question.⁴

3 Constants and time trends

If the mean of \mathbf{y}_t is the $N \times 1$ vector $\boldsymbol{\alpha}$, the model in (1) becomes

$$\mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\mu}_t, \quad \boldsymbol{\mu}_t = \boldsymbol{\Phi} \boldsymbol{\mu}_{t-1} + \boldsymbol{\eta}_t, \quad t = 1, \dots, T. \quad (7)$$

Again the prime concern is with the homogeneous model in which $\boldsymbol{\Phi} = \phi \mathbf{I}$. All of the tests proposed here are invariant to affine transformations of the data⁵.

Univariate DF tests are carried out simply by regressing Δy_t on y_{t-1} and a constant. For a given estimate of $\boldsymbol{\Sigma}_\eta$, the multivariate Wald test statistics corresponding to $\pi_0(N)$ and $\tau_0(N)$ in the previous section, can be computed, as in O'Connell (1998), by transforming using the Cholesky decomposition and then applying OLS to the

⁴Introduction of p lags reduces the sample size from $T - 1$ to $T - p - 1$. Compatibility with the (univariate) ADF test in small samples requires the use of $T - 2p - 1$ as a divisor in the estimate of the covariance matrix. The statistics reported for US regions in section 6 were computed in this way.

⁵That is $\mathbf{a} + \mathbf{B}\mathbf{y}$, where \mathbf{a} and \mathbf{B} are, respectively, a fixed vector and matrix. The tests of section 2 are not, of course, invariant to shifts brought about by \mathbf{a} .

pooled observations with dummy variables used to give a different constant for each series. This can be embedded in an iterative loop if the ML estimate of π is to be used. Alternatively one might just de-mean once and for all and then proceed as in the no constant case. This may be attractive if an augmented DF test statistic is to be formed by regressing Δy_{it} and $y_{i,t-1}$ on lagged differences and then working with the residuals; the constant is just included in these regressions.

3.1 Asymptotic distribution and critical values

Under the null hypothesis that $\pi = 0$, the asymptotic distribution of the MHDF t -statistic is

$$\tau_1(N) \rightarrow \frac{\sum_{i=1}^N \int_0^1 W_i(r) dW_i(r) - \sum_{i=1}^N W_i(1) \int_0^1 W_i(r) dr}{\left[\sum_{i=1}^N \int_0^1 W_i(r)^2 dr - \sum_{i=1}^N \left(\int_0^1 W_i(r) dr \right)^2 \right]^{1/2}} \quad (8)$$

As in the no constant case this is obtained straightforwardly from the asymptotics in the univariate case.

Table 3 gives asymptotic critical values for different N calculated in the same way as in table 1. O'Connell (1998, table 3) gives some small sample critical values for various combinations of N and T . For example, with $N = 10$, the 5% critical value is -5.31 for $T = 60$ and -5.34 for $T = 100$ against our -5.43. In contrast to the no constant case, the t -statistic does not converge to a standard normal as N increases; see Levin, Lin and Lu (2002, p8).

Table 3 Asymptotic critical values for MHDF test with constant

	1	2	3	4	5	6	7	8	9	10	15	20
0.01	-3.43	-3.89	-4.29	-4.58	-4.88	-5.13	-5.38	-5.62	-5.81	-6.01	-6.87	-7.62
0.05	-2.85	-3.33	-3.73	-4.02	-4.32	-4.56	-4.81	-5.03	-5.23	-5.43	-6.29	-7.01
0.1	-2.56	-3.03	-3.42	-3.73	-4.02	-4.26	-4.50	-4.73	-4.91	-5.11	-5.98	-6.69
0.9	-0.45	-0.84	-1.19	-1.50	-1.75	-2.01	-2.23	-2.45	-2.65	-2.83	-3.69	-4.42
0.95	-0.09	-0.49	-0.85	-1.16	-1.42	-1.66	-1.89	-2.10	-2.31	-2.50	-3.36	-4.09
0.99	0.61	0.15	-0.21	-0.52	-0.79	-1.03	-1.26	-1.45	-1.67	-1.87	-2.74	-3.46

Table 1:

Local asymptotic theory goes through as before. As in the no constant case, the practical implication is that so long as Σ_η is positive definite the local asymptotic power does not depend on it. The small sample simulations in O'Connell (1998, table 3) show that the form of Σ_η has little effect in small samples.

3.2 Multivariate Sargan-Bhargava test

The multivariate homogeneous Sargan-Bhargava test statistic is

$$\zeta_1(N) = \frac{1}{T^2} \sum_{t=1}^T \tilde{\boldsymbol{\mu}}_t' \hat{\Sigma}_\eta^{-1} \tilde{\boldsymbol{\mu}}_t, \quad (9)$$

where $\tilde{\boldsymbol{\mu}}_t = \mathbf{y}_t - \mathbf{y}_1$ for $t = 1, \dots, T$, and $\hat{\Sigma}_\eta = T^{-1} \sum_{t=2}^T \Delta \tilde{\boldsymbol{\mu}}_t \Delta \tilde{\boldsymbol{\mu}}_t'$. Like $\zeta_0(N)$, the asymptotic distribution of $\zeta_1(N)$ is $CvM_0(N)$ under the null hypothesis.

3.3 The MHDF-GLS test

The asymptotic distribution against local alternatives can be obtained, for both multivariate DF and SB tests, just as in the no constant case, using the relevant univariate formulae; see the expressions in Stock (1994, p2772-3). It is on the basis of this local analysis that Elliott, Rothenberg and Stock (1996) suggest a (univariate) unit root test that has power close to the power envelope. As is clear from their figure 2, this promises considerable gains over the Dickey-Fuller test, particularly in its t-statistic form, and also over the ζ_1 test, albeit to a lesser extent.

The most convenient way of implementing the Elliott, Rothenberg and Stock (1996) - ERS- suggestion is by what they call the DF-GLS statistic. In the multivariate context this means replacing \mathbf{y}_t in (3) by $\mathbf{y}_t - \hat{\boldsymbol{\alpha}}_c$, where

$$\hat{\boldsymbol{\alpha}}_c = \left[\mathbf{y}_1 + (1 - \bar{\phi}) \sum_{t=2}^T (\mathbf{y}_t - \bar{\phi} \mathbf{y}_{t-1}) \right] / [1 + (T - 1)(1 - \bar{\phi})^2] \quad (10)$$

and $\bar{\phi} = 1 + \bar{c}/T$. This de-meaning is based on GLS estimation⁶, assuming that $\boldsymbol{\mu}_0 = \mathbf{0}$. For reasons given below, the covariance matrix, $\boldsymbol{\Sigma}_\eta$, is best estimated using the ML estimator of ϕ rather than setting it to $\bar{\phi}$.

In the univariate case, ERS show that the asymptotic distribution of the DF-GLS statistic under the null hypothesis is the same as that of τ_0 , so the same critical values can be used. Furthermore the local asymptotic distribution is the same.⁷ These results carry over to the multivariate test which we denote as *MHDF - GLS*. Thus the asymptotic critical values for the test statistic, $\tau - GLS_c(N)$, are the same as those for $\tau_0(N)$ and the local asymptotic distribution is as in (16) in appendix B. Similarly, the local asymptotic representation of $\tau_1(N)$ is obtained by generalising (3.23b) in Stock along the lines of (16). In the univariate case, the local asymptotic power of $\tau_1(N)$ is well below that of the power envelope and this inefficiency can be expected to carry over to the multivariate case.

Generalising theorem 1 in ERS(1996, p 818) shows that the power envelope in the multivariate case is given by $P(c, c)$, where $P(c, \bar{c})$ is as in (17). As in ERS, p821 we estimated the power envelope by 50,000 Monte Carlo replications, approximating W_c by a discrete realisation from a sample of size 500. The result is shown in figure 1. There is a clear increase in power as N gets bigger.

ERS suggest setting $\bar{c} = -7$ as this corresponds to 0.5 on the power envelope. The table below shows the values of c giving a power of 0.5 for different N . A plot of $\log c$ on $\log N$ is almost linear and it seems that $c = -6.9N^{-0.85}$ corresponds to a power close to 0.5, at least within the range in the table. For $N = 20$, we found $c = -0.7$, while the equation predicted -0.54 .

Values of c corresponding to a local power of 0.5

N	1	2	3	4	5	10
$-c$	6.9	3.6	2.6	2.0	1.7	1.0

⁶Note that the covariance matrix is irrelevant as it cancels. Of course it is still needed in (3).

⁷(3.23k) in Stock (1994, p2772-3) is same as (3.23b) in the no constant case.

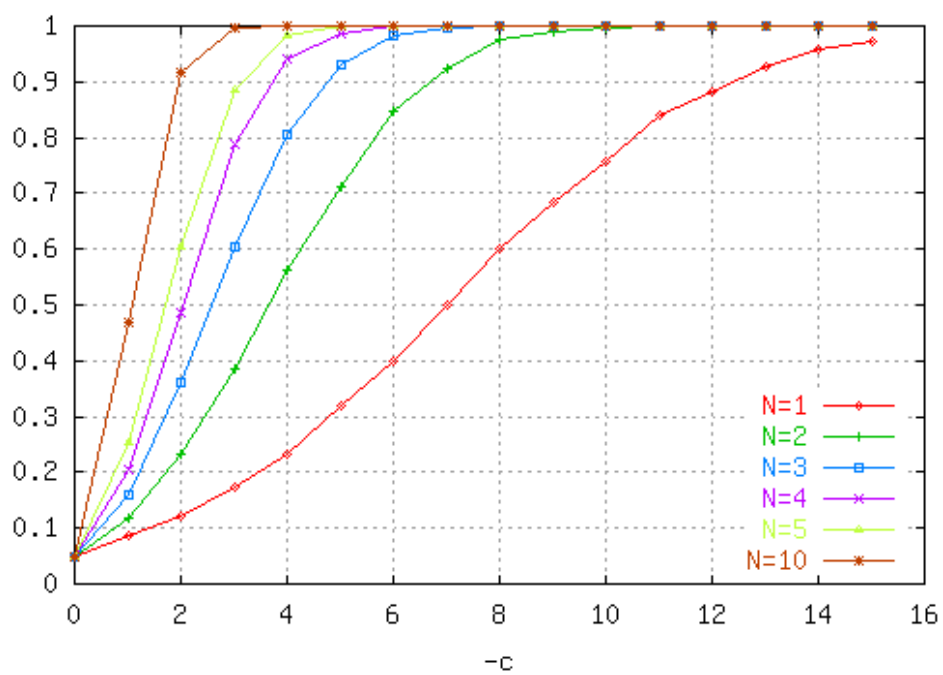


Figure 1: Power envelopes for different N

Setting $\bar{c} = 0$ in (10) gives $\hat{\alpha}_c = \mathbf{y}_1$. The square of the one-step version of the $\tau - GLS_0(N)$ statistic, that is with Σ_η estimated from first differences, is actually the LM statistic; see appendix A. Apart from a term involving \mathbf{y}_T , this is a monotonic transformation of the SB $\zeta_1(N)$ statistic. That the asymptotic distribution of $\tau - GLS_0(N)$ is the same as that of $\tau_0(N)$ is clear from (15). At the other extreme, setting $\bar{c} = -T$ simply de-means, resulting in a statistic that is approximately the same as DF and has the same asymptotic distribution. This implies that \bar{c} should not be too far from zero if the size of the test is required to be close to the nominal. Table 4 shows the results of a set of Monte Carlo experiments designed to investigate this matter. As can be seen, the values of \bar{c} that give a local power of 0.5 also give test sizes a little above the nominal. There is a slight reduction if the finite sample critical values are used, but not by much. The tests for higher values of N , namely 5 and 10, have sizes less than 0.06 when \bar{c} is set to give a local power of 0.5.

Table 4 Sizes of tests at nominal 5% level of significance

(a) T=100. Asymptotic critical values

N	$\tau_0(N)$	$\tau_1(N)$	$\tau - GLS_c(N)$					<i>Values of $-c$</i>	
			15	7	3.5	2	1	0	
1	0.053	0.055	0.170	0.081	0.061	0.056	0.055	0.054	
2	0.051	0.061	0.328	0.109	0.066	0.055	0.053	0.052	
5	0.054	0.071	0.680	0.203	0.087	0.063	0.055	0.054	
10	0.053	0.097	0.951	0.353	0.112	0.072	0.059	0.053	

(b) T=100. Critical values for T=100

N	$\tau_0(N)$	$\tau_1(N)$	$\tau - GLS_c(N)$					<i>Values of $-c$</i>	
			15	7	3.5	2	1	0	
1	0.050	0.050	0.159	0.076	0.057	0.052	0.051	0.050	
2	0.050	0.050	0.319	0.105	0.064	0.053	0.051	0.050	
5	0.050	0.050	0.662	0.191	0.081	0.058	0.050	0.050	
10	0.050	0.050	0.947	0.344	0.108	0.070	0.057	0.050	

(c) T=200. Asymptotic critical values

N	$\tau_0(N)$	$\tau_1(N)$	$\tau - GLS_c(N)$					Values of $-c$	
			15	7	3.5	2	1	0	
1	0.053	0.055	0.109	0.067	0.056	0.054	0.053	0.053	
2	0.051	0.057	0.182	0.077	0.059	0.053	0.052	0.050	
5	0.050	0.059	0.391	0.110	0.065	0.057	0.054	0.050	
10	0.052	0.070	0.669	0.158	0.076	0.060	0.055	0.053	

We experimented with estimating Σ_η , using $\bar{\phi}$ rather than the ML estimator of ϕ . This made little difference to the size for the low values of N , but, using the T=100 critical value, it increased the size to 0.24 in the case of $N = 10$.

If the slight inflation in size is a concern, the best option is to set \bar{c} to zero, which is essentially the same as doing the multivariate SB test. Alternatively \bar{c} can be set to give a local power of 0.5 and the exact critical value simulated. This matter is investigated further in the next section.

Note that the $\tau_1(N)$ test suffers a more serious size inflation than the $\tau_0(N)$ test when asymptotic critical values are used. This is worse for high N . For example with $T = 100$ and $N = 10$, the size of the $\tau_1(N)$ test is 0.097 as opposed to 0.053 for $\tau_0(N)$.

3.4 Time trend

Jorion and Sweeney (1996) extend the MHDF test by letting (7) include (heterogeneous) time trends. They find the empirical distribution by simulation and use the statistic to test for PPP. Using the methods described above the asymptotic distribution of the test statistic can be written down and a MHDF-GLS test can be set up. However, as a comparison of figures 2 and 3 in ERS makes clear, a high price is paid, in terms of power, by unnecessarily allowing for a time trend, even if the test is based on GLS detrending.

The extension of the SB test to the time trend model is discussed in appendix A.

4 Power in small samples

A series of Monte Carlo experiments were carried out to examine the power of the DF tests, and the ERS modification, in small samples and to compare them with a

likelihood ratio (LR) test in which no restrictions are put on Φ under the alternative hypothesis of stationarity. The LR test is easily carried out by two unrestricted vector autoregressions, one of $\Delta \mathbf{y}_t$ on \mathbf{y}_{t-1} and lagged values of $\Delta \mathbf{y}_t$ and another without \mathbf{y}_{t-1} . The theory of the LR test is set out in Johansen (1988,1995) where it is shown that because the test is of the null hypothesis of N unit roots ($\Phi = \mathbf{I}$), the asymptotic distribution depends only on N . Critical values are tabulated in Johansen (1995, p 214-5) up to $N = 12$. However, in the simulations below we used exact critical values obtained by simulation for both LR and DF tests. Thus the powers are directly comparable. Note that for $N = 1$ with no constant, the asymptotic distribution of the LR test statistic, LR_0 , is the same as that of τ_0^2 , but there appears to be little advantage from the one-sidedness of the DF test.

The simulation programs were written in the Ox language; see Doornik (1999). Table 5a shows powers for $T=100, 200$ and 500 and $N=2$ and 5 based on 50,000 replications. The DF and ERS tests were computed by iterating to convergence, though this makes little difference as compared with a two-step procedure. When a constant is included in the LR test, denoted LR_1 , it is estimated unrestrictedly. Table 5b shows powers for some heterogeneous models.

Table 5a Powers of *DF*, *ERS* and *LR* tests for homogeneous models

		$T = 100$					
N	ϕ	$\tau_0(N)$	$\tau_1(N)$	$\tau - GLS_c$	$\tau - GLS_0$	LR_0	LR_1
	.98	.226	.086	.227	.225	.063	.068
2	.95	.721	.204	.718	.718	.177	.120
	.90	.995	.637	.995	.995	.600	.366
	.98	.605	.143	.613	.613	.029	.057
5	.95	.998	.475	.998	.998	.051	.071
	.90	1	.981	1	1	.235	.207

N	ϕ	$T = 200$				$T = 500$			
		$\tau_0(N)$	$\tau_1(N)$	LR_0	LR_1	$\tau_0(N)$	$\tau_1(N)$	LR_0	LR_1
	.98	.553	.147	.131	.095	.996	.646	.596	.357
2	.95	.995	.631	.596	.349	1	1	1	.995
	.90	1	1	.996	.937	1	1	1	1
	.98	.981	.343	.041	.060	1	.985	.256	.214
5	.95	1	.984	.244	.208	1	1	.990	.992
	.90	1	1	.959	.889	1	1	1	1

Table 5b Powers of DF and LR tests for some heterogeneous models

N	ϕ_1	$\phi_2 = \dots = \phi_N$	$T = 100$				$T = 200$			
			$\tau_0(N)$	$\tau_1(N)$	LR_0	LR_1	$\tau_0(N)$	$\tau_1(N)$	LR_0	LR_1
2	.98	.95	.429	.156	.106	.090	.819	.343	.321	.201
	.98	.90	.574	.275	.243	.168	.878	.548	.772	.531
5	.98	.95	.987	.459	.044	.065	1	.910	.174	.162
	.98	.90	1	.846	.150	.152	1	.989	.828	.721

The main conclusions are as follows:

1) The power of the MHDF test increases with N against the homogeneous alternative. Thus for $\phi = .95$ and $T = 100$, the powers for $N = 1, 2$ and 5 (the $N = 1$ case being extracted from table 7) are 0.321, 0.721 and 0.998 respectively. These figures are close to the power envelope of figure 1.

2) When the test statistic allows for a constant, the powers for $N = 1, 2$ and 5 are 0.123, 0.204 and 0.475 respectively. The fact that these are much lower than the corresponding powers for the no constant test is to be expected from the results in the univariate case; see ERS (1996, figure 2).

3) The power increases as T increases in the homogeneous model confirming the consistency of the tests.

4) In the homogeneous model, the LR test suffers a loss in power as compared with the $\tau(N)$ tests. This is not surprising as it is designed for a more general hypothesis, and indeed its relative performance worsens as N increases. For $N = 5$ it is sometimes biased.

5) The MHDF still performs relatively well in the heterogeneous case. A test specifically designed for heterogeneity would obviously do better, but it is interesting that the LR test still has lower power than MHDF in the mixed case when ϕ_1 is 0.98 and the other ϕ'_i s are 0.95 or 0.90. As expected, its performance is relatively better for mixed cases; for example, compare the relative performance when all ϕ'_i s are 0.95 with what happens when ϕ_1 is 0.98 and the other ϕ'_i s are 0.90.

6) The power of the MHDF-GLS test, with \bar{c} set as in table 4, is similar to that of $\tau_0(N)$. If it is not size corrected, the rejection probability is higher but, as table 4 showed, this is at the cost of a higher actual size. For example with $N=2$, the rejection probabilities for $\phi = 1, .98, .95$ and 0.9 are 0.066, 0.284, 0.790 and 0.997 respectively.

7) Further simulations show that power of the MHDF-GLS test is not very sensitive to the choice of \bar{c} . Of particular importance is the case of $\bar{c} = 0$, since then the critical values in table 1 give tests with a size close to the nominal. As can be seen from table 5, there is little, if any, loss in power from having $\bar{c} = 0$. Thus there is a strong case for basing a test on $\bar{c} = 0$. Alternatively, the SB test might be used since this behaves in a very similar fashion.

5 Balanced growth and stability

Two time series exhibit balanced growth if the difference between them follows a stationary process. More generally, $N + 1$ series in a vector \mathbf{y}_t^\dagger exhibit balanced growth if there is a full rank $N \times (N + 1)$ matrix, \mathbf{D} , with the property that $\mathbf{D}\mathbf{i} = \mathbf{0}$, thereby rendering the N series in $\mathbf{y}_t = \mathbf{D}\mathbf{y}_t^\dagger$ jointly stationary. The rows of \mathbf{D} are therefore balanced growth co-integrating vectors, while the elements of \mathbf{y}_t may be termed *balanced growth contrasts*. Typically each row will contain a one, a minus one and zeroes elsewhere. For example, one country may be used as a benchmark or numeraire; if it is the $(N + 1)$ -th, then $\mathbf{D} = [\mathbf{I}_N, -\mathbf{i}_N]$. Alternatively the mean may be subtracted, so that \mathbf{D} consists of (any) N rows of $\mathbf{M} = \mathbf{I}_{N+1} - (N + 1)^{-1}\mathbf{i}_{N+1}\mathbf{i}_{N+1}'$.

Tests of balanced growth are based on the contrasts. In some situations, the data are initially in the form of what are effectively contrasts, an example being real exchange rates used in the study of PPP. It is of some importance that tests are invariant to the choice of \mathbf{D} , in other words pre-multiplication of \mathbf{y}_t by a non-singular $N \times N$ matrix should leave the test statistic unchanged. Kuo and Mikkola (2001) note the importance of this property in testing the hypothesis of PPP.

5.1 The null of instability

Since, in many situations, economic theory tends to suggest a stable relationship between series, perhaps emanating from balanced growth, this is the null hypothesis that should be tested. Hence the use of stationarity tests, multivariate versions of which are given in Nyblom and Harvey (2000) and Hobijn and Franses (2000). However, attention sometimes focuses on autoregressive mechanisms and in such circumstances unit root tests are an appropriate way of testing the null hypothesis of instability. Examples might include the mean reversion of inflation rates and the real interest rate parity hypothesis. The earlier discussion suggest that the SB test may be attractive in such cases.

What happens to multivariate test on benchmark contrasts when only some of the series are in a state of balanced growth with the benchmark? Nyblom and Harvey (2000) show that if a set of series contains at least one unit root, then the multivariate stationarity test is consistent. An interesting issue is whether a corresponding result holds for unit root tests. Suppose that $\Phi = \text{diag} \{ \phi_1, \dots, \phi_N \}$. Then if $\phi_i = 1$, $i = 1, \dots, K$ while $\phi_i = \phi$, $|\phi| < 1$, $i = K + 1, \dots, N$, there is only partial convergence if K is positive. In order to see the implications for the $\tau_0(N)$ test, we make the simplifying assumption that the disturbances η_{it} , $i = 1, \dots, N$, are distributed independently of each other with the same variance. Then

$$\tau_0(N) \xrightarrow{L} \frac{(1/2) (\chi_K^2 - K) - (N - K)/(1 + \phi)}{\{CvM_0(K)\}^{1/2}}, \quad K = 1, \dots, N.$$

While this result is somewhat limited, it does indicate that $\tau_0(N)$ test is not consistent against alternatives in which only some of the series are stationary. In other words there is a finite probability that the test will not reject. (Showing that the test is consistent when $K = 0$ is straightforward). This probability becomes smaller as $N - K$ increases. A similar analysis can be carried out for the $\tau_1(N)$, SB and $\tau - GLS_c$ tests.

5.2 Cross-correlations and contrasts

The usual approach to testing for a unit root in panel data is to pool the observations and estimate π by OLS, usually after making a correction for heteroscedasticity across units. If there are cross-correlations the size of the DF test will be distorted.

One structure that is often used is a covariance matrix in which the correlations between all pairs are the same, that is

$$\boldsymbol{\Sigma} = (1 - \omega)\mathbf{I}_N + \omega\mathbf{i}_N\mathbf{i}'_N = \mathbf{E}_N(\omega) \quad (11)$$

where ω is a parameter that gives the correlation. This may be motivated by a disturbance structure in which each disturbance has a common factor and a specific part. If the N series in \mathbf{y}_t are a set of balanced growth contrasts, it is more appropriate to begin by positing an $\mathbf{E}_{N+1}(\omega)$ matrix for the $N + 1$ series, \mathbf{y}_t^\dagger , from which the contrasts are obtained. For any benchmark set of contrasts, the implied covariance matrix is then

$$\mathbf{D}\mathbf{E}_{N+1}(\omega)\mathbf{D}' = (1 - \omega)(\mathbf{I}_N + \mathbf{i}_{N+1}\mathbf{i}'_{N+1}) = 2(1 - \omega)\mathbf{E}_N(0.5) \quad (12)$$

Thus whatever the original correlation between the series, the correlation between the benchmarked contrasts is 0.5. The simulations in O'Connell (1998, table 1) imply that for a τ_1 test based on simple pooling, the true size for a nominal 5% level of significance is around 0.17 for $N = 10$ and 0.38 for $N = 50$.

For deviations from the mean, $\mathbf{M}\mathbf{E}_{N+1}(\omega)\mathbf{M} = (1 - \omega)\mathbf{M}$, so the correlation is again independent of ω , being equal to $-1/N$. Note that one row of \mathbf{M} is removed if a set of balanced growth co-integrating vectors are to be formed and so the covariance matrix of the contrasts will have one row and one column removed to make it of dimension $N \times N$, that is, it is proportional to $\mathbf{I}_N - (N + 1)^{-1}\mathbf{i}_N\mathbf{i}'_N$. Unlike \mathbf{M} , this matrix is no longer singular.

6 Testing convergence

Let y_t denote the difference between two regions or countries in terms of, say, income per capita. The error correction mechanism

$$y_t = \alpha + \mu_t, \quad \Delta\mu_t = \pi\mu_{t-1} + \eta_t, \quad t = 1, \dots, T, \quad (13)$$

where $\pi = \phi - 1$, captures the process of convergence. If the mean, α , is zero the countries are in a state of *absolute convergence*. If not, we have *conditional* or *relative convergence*. The regression of the difference in the growth rate of the two series, Δy_t , on the lagged difference, y_{t-1} , directly estimates the rate of convergence, π , and

the t -statistic tests the null hypothesis that no convergence is taking place. If initial conditions are unimportant, stability implies that y_t is stationary for virtually the whole period and the levels of the series exhibit balanced growth. In other words the series *have converged*. This is to be contrasted with a situation in which the initial value of μ_t is some way from zero, so the situation is one in which the series *are converging*. The claim by Bernard and Durlauf (1996 p 170) that the time series approaches to convergence check for the compatibility of the difference in (log) output with an indeterministic stationary series misses the point since it confuses stability with convergence.

Writing the model in EC form, (13), accords with the notion of beta convergence in the cross-sectional literature, as expounded by Barro and Sala-i-Martin (1995) and others, except that there the growth rate is taken to be a linear function of an arbitrarily chosen initial value; see Evans and Karras (1996, p 253).

The purpose of this section is to explore various aspects of testing for convergence. Multivariate methods are then used to investigate convergence of US regions.

6.1 Initial conditions

We now consider the implications that plausible values for the rate of convergence and initial conditions have for the power of convergence tests. Table 2 reports some Monte Carlo experiments estimating the probability of rejecting the null hypothesis of a unit root for the univariate model in (13) with $T = 100$ and a range of starting values, μ_0 . Asymptotic critical values are used and the level parameter, α , is set to zero in the data generating process.

When convergence is supposed to be absolute, rates of convergence of around 0.02 seem to be the norm; see, for example Sala-i-Martin (1996). The annual data for US regions from 1950 to 1999, analysed later, fits into this pattern as it shows an estimate of a little over 0.02. As regards initial values, the log difference between the richest and poorest US regions, north-east and south-east, in 1950 was about 0.5. With a standard deviation, σ_η , of 0.025, 0.5 corresponds to an initial value of 20 and 40.

The DF t -test and the SB test are shown with and without the inclusion of a constant. The DF-GLS test of ERS, $\tau - GLS_c$, is also included. The results for the tests with a constant included are invariant to α , so the fact that it is set to zero is irrelevant. These tests are also invariant to μ_0 under the null hypothesis. Although

the τ_0 test is not invariant to μ_0 it appears to be quite robust in this respect.⁸

⁸This is not true, incidentally, for the direct coefficient test, π_0 . Although its limiting distribution under the null hypothesis does not depend on μ_0 , its finite sample distribution becomes more concentrated as μ_0/σ_η increases and so using asymptotic critical values will tend to give a conservative test; see Evans and Savin (1981, p764).

Table 7 Estimated rejection probabilities for different initial conditions

ϕ	μ_0	τ_1	τ_0	$\tau - GLS_7$	$\tau - GLS_0$	ζ_1	ζ_0
1	0	0.053	0.051	0.081	.054	0.046	0.046
	2	"	0.051	"	"	"	0.042
	5	"	0.051	"	"	"	0.024
	10	"	0.050	"	"	"	0.004
	20	"	0.050	"	"	"	0.000
0.98	0	0.072	0.119	0.188	0.128	0.106	0.105
0.98	2	0.073	0.123	0.177	0.117	0.098	0.100
0.98	5	0.072	0.143	0.124	0.080	0.069	0.070
0.98	10	0.076	0.242	0.036	0.021	0.019	0.023
0.98	20	0.106	0.748	0.000	0.000	0.000	0.000
0.95	0	0.123	0.321	0.447	0.336	0.281	0.281
0.95	2	0.128	0.338	0.346	0.229	0.195	0.273
0.95	5	0.149	0.433	0.080	0.033	0.025	0.238
0.95	10	0.235	0.767	0.000	0.000	0.000	0.150
0.95	20	0.685	1.000	0.000	0.000	0.000	0.031
0.90	0	0.328	0.765	0.857	0.782	0.720	0.716
0.90	2	0.342	0.794	0.646	0.362	0.302	0.713
0.90	5	0.442	0.903	0.031	0.001	0.001	0.712
0.90	10	0.765	0.998	0.000	0.000	0.000	0.719
0.90	20	1.000	1.000	0.000	0.000	0.000	0.584

The results for zero initial conditions are in accordance with what is already known and summarised by Stock (1994, 2773-80). The DF t -test has much higher power as a test for absolute convergence when no constant is included. Since its size appears to be unaffected by non-zero initial conditions, no price is paid for the higher rejection probabilities obtained in these cases. For non-zero initial conditions, the results are consistent with the findings of Muller and Elliot (2003). In particular, the transitional dynamics actually help in that the powers of the τ_0 and τ_1 increase the further is the starting value from zero.

Although the size of the ζ_1 test is not affected by different initial conditions its power under the alternative falls as μ_0 increases. This is because the process tends to resemble a nonstationary process and the distribution shifts to the right. The ζ_0 test

is undersized, but as ϕ moves away from one, the probability of rejection increases. In fact with $\phi = 0.9$, the rejection probability is well above that of τ_1 and not too far below that of τ_0 . On the other hand, the power falls as μ_0 increases and it starts to fall quite rapidly beyond a certain point. The same tendency can be observed in the results reported by Schmidt and Lee (1991), though they include a time trend.

The $\tau - GLS_0$ test statistic is not very different from the ζ_1 statistic and it displays the same pattern as ζ_1 as μ_0 increases. As we have seen in the previous section, the power of the DF-GLS test is relatively insensitive to the choice of \bar{c} and this is reflected in the fact that $\tau - GLS_7$ also shows diminution of power as μ_0 increases.

The general thrust of the above conclusions carries over to multivariate tests. The LR test, examined in section 4, was not included in any simulations here, but, like the DF tests, it will show an increase in power as the initial conditions move away from zero; when a constant is included it is important that it be estimated unrestrictedly.

Although table 7 shows that the power of τ_1 increases with non-zero initial conditions, its power is still very low. When $\phi = .98$, the rejection probability is 0.07 for initial values of 10 or less, rising to a mere 0.11 when it is 20. On the other hand τ_0 performs much better, rejecting with probability 0.24 and 0.75 for initial values of 10 and 20, as opposed to only 0.12 when the initial value is zero. Nevertheless a power of .24 is still rather low. This provides an incentive for testing the convergence hypothesis using several series. The benefit of pooling series is clear from the Monte Carlo experiments reported in table 4; for five series the power of τ_0 is 0.61 and this will rapidly approach one as the initial conditions move away from zero.

6.2 Invariance and heterogeneity

Tests of convergence are based on balanced growth contrasts, $\mathbf{D}\mathbf{y}_t^\dagger = \mathbf{y}_t$. An immediate attraction of the tests in sections 2 and 3 is that they are invariant to the choice of \mathbf{D} . Furthermore if the model for one set of contrasts is homogeneous then the model for any other set of contrasts is homogeneous. This property does not carry over to heterogeneous models in which Φ is diagonal rather than scalar. Heterogeneity is specific to a particular contrast. Similarly there is no particular reason to suppose that a diagonal structure is more appropriate for one benchmark economy rather than another.⁹ As regards tests based on deviations from the mean, having equal weights

⁹The US is typically taken to be the benchmark; see, for example, Linden (2000).

in constructing the mean is arbitrary and is particularly hard to justify if the units are of different sizes, as would be the case with different countries.

6.3 Time trends

Converging economies eventually end up on a balanced growth path as in (??). The \mathbf{D} matrix eliminates the time trend from the contrasts, but if the time trend had a different coefficient in each series, it would remain in the contrasts. One could certainly test for convergence using the DF test with a constant and a time trend included and many researchers do. However, such a test will tend to have very low power; see the discussion in sub-section 3.4. The question to be asked is whether testing for convergence to diverging growth paths makes any economic sense in the first place.

6.4 US regions

The issues involved in testing for convergence are nicely illustrated by data on the logarithms of annual real per capita incomes in US census regions from 1950 to 1999. The data are obtained from the Bureau of Economic Analysis and deflated by the US implicit price deflator (1996=100). The regions are: New England (NE), Mid East (ME), Great Lakes (GL), Plains (PL), South East (SE), South West (SW), Rocky Mountains (RM) and Far West (FW). An inspection of the series indicates that there is a strong case for absolute convergence amongst all regions apart from NE and ME. Figure 2 shows the series for PL, SE, SW, RM and FW relative to GL, which is arbitrarily chosen as the benchmark.

If there are $N + 1$ economies then there are $N(N + 1)/2$ pairwise comparisons. However, all the information for a test of overall convergence is based on a set of N contrasts. Most researchers choose the set of contrasts based on deviations from the mean, although the fact that there are $N + 1$ such contrasts but one is redundant is not always recognized when the information in the tests is combined. Evans and Karras (1996) note that a necessary and sufficient condition for convergence across $N + 1$ economies is that each of the deviation from the mean contrasts should converge. Unless all pairs of economies converge, this convergence condition will be violated for all $N + 1$ because the mean depends on all economies. Evans and Karras (1996)

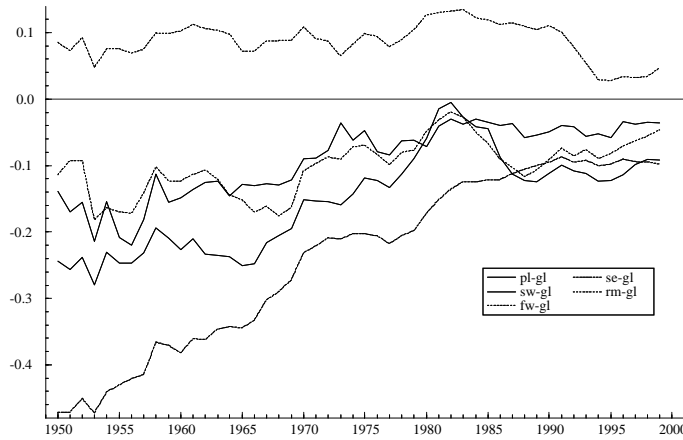


Figure 2: US regions relative to Great Lakes

therefore argue that we should reject convergence if just one unit root is not rejected. The problem with such a strategy is that the overall test has very little power. The results in table 8 illustrate the dilemma. There are no rejections with the ADF test when a constant is included and without the constant only SE is significant at the 5% level. The test statistics don't change much if the number, p , of lagged differences is varied. Adding a time trend only makes matters worse; for example $SE - GL$ gives $t = -.942$ as opposed to -1.876 with a constant.¹⁰

Table 8 ADF τ tests on deviations from the mean ($p = 4$)

Region	PL	GL	SE	SW	RM	FW
No constant	-1.807*	-1.753*	-2.014**	-0.722	-1.201	-1.538
Constant	-2.193	-1.418	-1.734	-1.761	-1.330	-0.553

* denotes significant at 10% level, ** denotes significant at 5% level.

Table 9 shows the full set of pairwise DF tests without the constant. If a constant is included, nothing is rejected at the 10% level of significance. Without the constant, all pairs are linked either directly or indirectly by significant test statistics indicating some degree of support for overall convergence.

¹⁰Turning to sigma convergence, a plot of the cross-sectional variance shows a sharp fall from a value of .032 in 1950 to one of .003 in 1999.

Table 9 Pairwise ADF τ_0 tests ($p=4$)

	<i>GL</i>	<i>PL</i>	<i>SE</i>	<i>SW</i>	<i>RM</i>
<i>PL</i>	-1.833*				
<i>SE</i>	-2.526**	-2.462**			
<i>SW</i>	-1.144	-1.114	-2.381**		
<i>RM</i>	-1.300	-1.398	-1.967**	-0.924	
<i>FW</i>	-0.748	-1.454	-2.417**	-1.069	-1.181

Table 9 allows us to examine the inferences that might be drawn from taking different regions as the benchmark. For example, if SE is a benchmark all the statistics reject, while with RM or FW there is only one rejection (the SE of course). Thus although we get a valid inference from any one pair and although all the information for a joint test is contained in N contrasts with a particular benchmark, trying to simply combine the univariate tests can be highly misleading. Different benchmarks can apparently give contradictory results.

Fitting the multivariate model, also with four lags, gives $\phi = 0.982$ and $\tau_0(5) = -3.65$. The null of no convergence is rejected as the 1% critical value is -2.50. The LR statistic, 86.06, also rejects. Remember that these results are unambiguous as they do not depend on which benchmark is adopted or whether (five) deviations from the mean are used.

With a constant included, $\tau_1(5) = -1.32$, so the null of no convergence cannot be rejected at any reasonable level of significance; the 10% critical value is -4.02. This is again symptomatic of the low power that results when a constant is included. The ERS test similarly fails to reject; a range of values of \bar{c} from -7 to zero were tried and in all cases the $\tau - GLS_c(5)$ statistics were actually positive. The LR test for the joint significance of the intercepts is 4.29; this is not significant at the 10% level since the critical value for χ_5^2 is 9.24.

As well as giving different results with different benchmarks, failing to take account of cross-correlations can result in considerable size distortions. The cross-correlation matrix of first differences of the six regions has fairly high values ranging from .46 to .85. However, they are far from being the same and the correlations for deviations from the mean are some way from the value of -0.17 that would have been expected with equal correlations in the original series. The correlations for the series with GL as a benchmark, as in figure 2, ranged between 0.42 and 0.78.

7 Conclusions

A joint test of the null hypothesis that a set of series do not have a stable relationship, against the alternative that they do, can be carried out by generalising the Dickey-Fuller Wald test under the assumption of homogeneity. An allowance is made for cross-dependence between the series. A multivariate homogeneous Sargan-Bhargava test, closely related to the Lagrange multiplier test, can be similarly derived. Because these tests have only one extra parameter under the stationary alternative, they are very simple, with asymptotic distributions under the null hypothesis that depend only on the number of series. The critical values for the multivariate Dickey-Fuller test, with and without constant terms, are tabulated. We argue that the power properties of tests based on the homogeneity assumption are good and note that tests allowing for heterogeneity are not invariant to the set of contrasts on which they are based. Indeed the heterogeneous model itself suffers from the drawback that its structure is specific to a particular set of contrasts.

An analysis of local power suggests suitable ways of implementing the multivariate version of the Elliot, Rothenberg and Stock (1996) modification and highlights the gain in power as the number of series increases. The asymptotic critical values are the same as those tabulated for the multivariate Dickey-Fuller test without a constant. The power of the ERS test is relatively insensitive to the choice of \bar{c} , the parameter that fixes the autoregressive coefficient for the purposes of detrending. There is a good case for letting $\bar{c} = 0$, since the tabulated critical values give sizes close to the nominal. The Sargan-Bhargava test has very similar properties, and has tabulated critical values from the Cramér-von Mises distribution. Hence there is much to recommend it in studies involving PPP and reversion to the mean of interest rates and real wages.

If two series are in the process of converging, the dynamics can be captured by a simple error correction model in which the difference in the growth rate depends on the gap in the previous period. A test of whether the ECM parameter is significantly different from zero is then a test of the null hypothesis of no convergence. This is just the Dickey-Fuller t -test. Our simulation results show that the power of the DF test is higher, the bigger is the gap at the start of the series. However, when a constant is included the power of the test is relatively low so if absolute convergence is being entertained as the main possibility, a constant should not be included unnecessarily. The ERS and Sargan-Bhargava tests are effectively ruled out by the adverse effect of

non-zero initial conditions. If a test of relative, as opposed to absolute, convergence is required, it should be based on the estimated constant in the DF regression since the sample mean is distorted by the initial conditions.

The augmented Dickey-Fuller test, based on the inclusion of lagged first differences in the DF regression, offers a suitable way of handling serial correlation and this extends to the proposed multivariate tests. A joint test against the alternative hypothesis that several series are converging absolutely can be carried out by the multivariate homogeneous augmented Dickey-Fuller test. This will typically have high power. Testing for relative convergence is more problematic.

Acknowledgements

The work was supported by the Economic and Social Research Council (ESRC) as part of a project on Dynamic Common Factor Models for Regional Time Series, grant number L138 25 1008. We are grateful to Fabio Buseti, Gernot Doppelhofer, Bernard Fingleton, Petra Geraats, David Hendry, Soren Johansen, Jukka Nyblom, Vanessa Smith and Melvyn Weeks for helpful comments. All errors are our responsibility.

APPENDICES

A Multivariate LM and Sargan-Bhargava tests

Consider the model

$$\mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}t + \boldsymbol{\mu}_t, \quad \boldsymbol{\mu}_t = \boldsymbol{\Phi}\boldsymbol{\mu}_{t-1} + \boldsymbol{\eta}_t, \quad t = 1, \dots, T. \quad (14)$$

with $\boldsymbol{\Phi} = \phi\mathbf{I}$. Let $\tilde{\boldsymbol{\mu}}_t$ denote \mathbf{y}_t detrended under the null hypothesis, that is $\tilde{\boldsymbol{\mu}}_t = \mathbf{y}_t - \tilde{\boldsymbol{\alpha}}_0 - \tilde{\boldsymbol{\beta}}t$, where $\tilde{\boldsymbol{\beta}} = (\mathbf{y}_T - \mathbf{y}_1)/(T - 1)$ and $\tilde{\boldsymbol{\alpha}}_0 = \mathbf{y}_1 - \tilde{\boldsymbol{\beta}}$, where $\boldsymbol{\alpha}_0 = \boldsymbol{\alpha} + \boldsymbol{\mu}_0$. Note that $\tilde{\boldsymbol{\mu}}_1 = \mathbf{0}$, as a consequence of fitting the constant, while $\tilde{\boldsymbol{\mu}}_T = \mathbf{0}$ because of the slope; $\tilde{\boldsymbol{\mu}}_0$ is taken to be zero in all cases. The LM test of the null hypothesis that $\pi = 0$ ($\phi = 1$) is found by evaluating the first derivative of the log-likelihood function at $\phi = 1$ to yield

$$\frac{\partial \log L}{\partial \phi} = \sum_{t=1}^T (\tilde{\boldsymbol{\mu}}_t' - \phi \tilde{\boldsymbol{\mu}}_{t-1}') \hat{\boldsymbol{\Sigma}}_{\eta}^{-1} \tilde{\boldsymbol{\mu}}_{t-1} = \sum_{t=1}^T \Delta \tilde{\boldsymbol{\mu}}_t' \hat{\boldsymbol{\Sigma}}_{\eta}^{-1} \tilde{\boldsymbol{\mu}}_{t-1}.$$

On evaluating the second derivative, we find that

$$LM = \left(\sum_{t=1}^T \Delta \tilde{\boldsymbol{\mu}}_t' \hat{\boldsymbol{\Sigma}}_\eta^{-1} \tilde{\boldsymbol{\mu}}_{t-1} \right)^2 / \sum_{t=1}^T \tilde{\boldsymbol{\mu}}_{t-1}' \hat{\boldsymbol{\Sigma}}_\eta^{-1} \tilde{\boldsymbol{\mu}}_{t-1}.$$

This is the square of the one-step $\tau - GLS_0(N)$ statistic.

Now, since $\tilde{\boldsymbol{\mu}}_0$ is always zero, we find that

$$\sum_{t=1}^T \Delta \tilde{\boldsymbol{\mu}}_t' \boldsymbol{\Sigma}_\eta^{-1} \Delta \tilde{\boldsymbol{\mu}}_t = -2 \sum_{t=1}^T \Delta \tilde{\boldsymbol{\mu}}_t' \boldsymbol{\Sigma}_\eta^{-1} \tilde{\boldsymbol{\mu}}_{t-1} + \tilde{\boldsymbol{\mu}}_T' \boldsymbol{\Sigma}_\eta^{-1} \tilde{\boldsymbol{\mu}}_T.$$

If $\boldsymbol{\Sigma}_\eta$ is estimated by $T^{-1} \sum_{t=1}^T \Delta \tilde{\boldsymbol{\mu}}_t \Delta \tilde{\boldsymbol{\mu}}_t'$, the left hand side of the above expression reduces to NT because $\sum_{t=1}^T \Delta \tilde{\boldsymbol{\mu}}_t' \hat{\boldsymbol{\Sigma}}_\eta^{-1} \Delta \tilde{\boldsymbol{\mu}}_t = tr \left[\hat{\boldsymbol{\Sigma}}_\eta^{-1} \sum_{t=1}^T \Delta \tilde{\boldsymbol{\mu}}_t \Delta \tilde{\boldsymbol{\mu}}_t' \right]$ and so, provided the slope is estimated, $\tilde{\boldsymbol{\mu}}_T = 0$ and it follows that $\sum_{t=1}^T \Delta \tilde{\boldsymbol{\mu}}_t' \hat{\boldsymbol{\Sigma}}_\eta^{-1} \tilde{\boldsymbol{\mu}}_{t-1} = -NT/2$. Thus the LM statistic is a monotonic transformation of $\zeta_2(N)$, the Sargan-Bhargava statistic constructed from detrended observations, being equal to $N^2/4 \zeta_2(N)$.

Without the slope,

$$\begin{aligned} \tau - GLS_0(N) &= \sqrt{LM} = \frac{-NT + \tilde{\boldsymbol{\mu}}_T' \hat{\boldsymbol{\Sigma}}_\eta^{-1} \tilde{\boldsymbol{\mu}}_T}{2 \left(\sum_{t=1}^T \tilde{\boldsymbol{\mu}}_{t-1}' \hat{\boldsymbol{\Sigma}}_\eta^{-1} \tilde{\boldsymbol{\mu}}_{t-1} \right)^{1/2}} \\ &= \frac{(\mathbf{y}_T - \mathbf{y}_1)' \hat{\boldsymbol{\Sigma}}_\eta^{-1} (\mathbf{y}_T - \mathbf{y}_1) - NT}{2 \left(\sum_{t=1}^T (\mathbf{y}_t - \mathbf{y}_1)' \hat{\boldsymbol{\Sigma}}_\eta^{-1} (\mathbf{y}_t - \mathbf{y}_1) \right)^{1/2}}. \end{aligned} \quad (15)$$

The fact that this has the distribution in (4) is immediately apparent.

B Local power

Premultiplying a homogeneous model through by $\boldsymbol{\Sigma}_\eta^{-1/2}$ yields a homogeneous model with a disturbance covariance matrix equal to the identity matrix. The invariance of the $\tau(N)$ and $\pi(N)$ statistics means that their asymptotic distributions against the local alternative, $\phi = 1 + c/T$, can be obtained using the relevant univariate formulae, as given in Stock (1994, p2772-3). The formula for the t -statistic is constructed in an analogous fashion to (4). Thus

$$\tau_0(N) \rightarrow \frac{\sum_{i=1}^N \int_0^1 W_{ci}(r) dW_i(r)}{\left[\sum_{i=1}^N \int_0^1 W_{ci}(r)^2 dr \right]^{1/2}} = \frac{(1/2) \sum_{i=1}^N (W_{ci}(1)^2 - 1)}{\left[\sum_{i=1}^N \int_0^1 W_{ci}(r)^2 dr \right]^{1/2}} \quad (16)$$

where $W_{ci}(r), i = 1, \dots, N$ are independent diffusion processes defined by

$$W_{ci}(r) = \int_0^r \exp\{c(r-t)\} dW_i(t), \quad i = 1, \dots, N,$$

each satisfying the stochastic differential equation

$$dW_{ci}(t) = cW_{ci}(t)dt + dW_i(t),$$

with $W_{ci}(0) = 0$; note that $W_{0i}(\cdot) = W_i(\cdot)$. Following ERS(1996, theorem 1) the local asymptotic power function¹¹ for the Neyman-Pearson most powerful test against the alternative $c = \bar{c}$ and having significance level is ϵ is given by

$$P(c, \bar{c}) = \Pr \left[\bar{c}^2 \sum_{i=1}^N \int W_{ci}(r)^2 dr - \bar{c} \sum_{i=1}^N W_{ci}(1)^2 < b(\bar{c}) \right] \quad (17)$$

where $b(\bar{c})$ satisfies $\Pr \left[\bar{c}^2 \sum_{i=1}^N \int W_i(r)^2 dt - \bar{c} \sum_{i=1}^N W_i(1)^2 < b(\bar{c}) \right] = \epsilon$. The envelope for the family of point optimal tests is given by $P(c, c)$. The test statistics depend on the ratio of the determinants of the estimators of Σ_η under the alternative and null hypotheses; compare a similar result for the multivariate stationarity test in Nyblom and Harvey (2000, theorem A.1). Although $\tau_0(N)$ is not of this form, the analysis in ERS suggests that its power will be close.

Simulations can be carried out from these expressions and the local powers estimated. In the univariate case, both the direct coefficient test and τ_0 are very close to the power envelope; see figure 1 in ERS (1996) and Stock(1994).

¹¹The theory follows from appendix B of ERS except that there are determinants of estimators of Σ_η present; these cancel under the local alternative.

References

- [1] Abuaf, N. and P. Jorion., 1990. Purchasing power parity in the long run. *Journal of Finance* 45, 157-74.
- [2] Barro, R.J. and Sala-i-Martin.,1992. Convergence. *Journal of Political Economy*, 100, 223-51.
- [3] Barro, R.J. and X. Sala-i-Martin., 1995. *Economic Growth*. McGraw-Hill, Boston, MA.
- [4] Bhargava, A., 1986. On the theory of testing for unit roots in observed time series, *Review of Economic Studies* 53, 36-84.
- [5] Bernard, A. and S. Durlauf., 1995. Convergence in international output. *Journal of Applied Econometrics* 10, 97-108.
- [6] Bernard, A. and S. Durlauf., 1996. Interpreting tests of the convergence hypothesis. *Journal of Econometrics*, 71, 161-173.
- [7] Culver S.E. and D.H. Papell., 1997. Is there a unit root in the inflation rate? Evidence from sequential break and panel data models. *Journal of Applied Econometrics*, **12**: 435-444.
- [8] Doornik, J. A., 1999. *Ox: An Object-Oriented Matrix Language*. 3rd edition. Timberlake Consultants Press: London.
- [9] Durlauf, S. and D. Quah., 1999. The new empirics of economic growth. In J.B. Taylor and M. Woodford (eds.). *Handbook of Macroeconomics*, Volume 1. Chapter 4, 235-308. Amsterdam: Elsevier Science.
- [10] Elliot, G., Rothenberg, T.J. and J.H.Stock., 1996. Efficient tests for an autoregressive unit root. *Econometrica*, 64, 813-36.
- [11] Evans, P. and G. Karras., 1996. Convergence revisited. *Journal of Monetary Economics* 37, 249-265.
- [12] Hobijn, B. and P.H. Franses., 2000. Asymptotically perfect and relative convergence of productivity. *Journal of Applied Econometrics* 15, 59-81.

- [13] Im, K.S., Pesaran, M H and Y Shin., 2003. Testing for unit roots in heterogeneous panels. *Journal of Econometrics*, 115, 53-74.
- [14] Johansen, S., 1995. *Likelihood-Based Inference in Co-Integrated Vector Autoregressive Models*. Oxford. Oxford University Press.
- [15] Jorion, P. and R. Sweeney., 1996. Mean reversion in real exchange rates: evidence and implications for forecasting. *Journal of International Money and Finance*, 15, 535-50.
- [16] Kuo, B-S. and A. Mikkola., 2001. How sure are we about purchasing power parity? Panel evidence with the null of stationary exchange rates. *Journal of Money Credit and Banking* 33, 767-89.
- [17] Lee, H.Y. and J.L. Wu., 2001. Mean reversion of inflation rates: Evidence from 13 OECD countries. *Journal of Macroeconomics* 23, 477-487.
- [18] Levin, A., Lin, C-F. and C.-S. Lu., 2002. Unit root tests in panel data: asymptotic and finite sample properties. *Journal of Econometrics* 108, 1-24.
- [19] Linden, M., 2000. Testing growth convergence with time series data - a nonparametric approach. *International Review of Applied Econometrics* 14, 361-70.
- [20] Maddala, G.S. and S.Wu., 1999. A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and Statistics*, 61, 631-52.
- [21] Muller, U.K. and G. Elliot, 2003. Tests for unit roots and the initial condition. *Econometrica*, 71, 1269-1286.
- [22] Ng, S. and P. Perron., 1995. Unit root tests in ARMA models with data dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association* 90, 268-81.
- [23] Nyblom, J. and A.C. Harvey., 2000. Tests of Common Stochastic Trends. *Econometric Theory* 16, 176-199.
- [24] O'Connell, P., 1998. The overvaluation of purchasing power parity. *Journal of International Economics* 44, 1-19.

- [25] Oya, K. and H.Y.Toda.,1998. Dickey-Fuller, Lagrange multiplier and combined tests for a unit root in autoregressive time series. *Journal of Time Series Analysis* 19, 325-47.
- [26] Phillips, P.C.B and D.Sul., 2002. Dynamic panel estimation and homogeneity testing under cross section dependence. Cowles Foundation discussion paper no. 1362.
- [27] Sargan, J.D. and A. Bhargava., 1983. Testing residuals from least squares regression for being generated by the Gaussian random walk. *Econometrica* 51, 153-74.
- [28] Schmidt, P. and P.C.B. Phillips., 1992. LM Tests for a Unit Root in the Presence of Deterministic Trends. *Oxford Bulletin of Economics and Statistics* 54, 257-287.
- [29] Schmidt, P. and J. Lee., 1991. A modification of the Schmidt-Phillips unit root test. *Economics Letters* 36, 285-89.
- [30] Smith, V., Leybourne, S., Kim, T-H., and P. Newbold., 2002. More powerful panel data unit root tests with an application to mean reversion in real exchange rates. Mimeo, University of Nottingham.
- [31] Stock, J.H., 1994. Unit roots, structural breaks and trends, in R.F. Engle and D.L. McFadden (eds.), *Handbook of Econometrics* 4, Elsevier Science, 2739-2840.
- [32] Taylor, M.P. and L. Sarno., 1998. The behaviour of real exchange rates during the post-Bretton Woods period. *Journal of International Economics* 46, 281-312.