Multivariate Unit Root Tests and Testing for Convergence

Andrew Harvey and David Bates
Faculty of Economics and Politics, University of Cambridge

December 9, 2002

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 presented and 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 carried out even if the number of series exceeds the number of time periods. 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 is concerned with testing for unit roots in cross-sections of time series. There are two contributions, one statistical, the other methodological. The statistical contribution revolves around multivariate unit root tests and their properties. A subsidiary theme is the effect of initial conditions on unit root tests. The methodological contribution tries to clarify the case for using various modifications of the Dickey-Fuller (DF) test to determine whether time series are converging. The typical applications are for a moderate number of series, perhaps countries or regions, observed over a reasonably long period of time. The example we give is for the per capita incomes of six US regions observed annually from 1950 to 1999.

One of the reasons why unit root tests fail to reject is because they lack power. More specifically, the 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 (2002). However, these studies assume that the units can be treated as though they were independent of each other. The independence assumption is often not a reasonable one. Ignoring such 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 test[1] 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 asymptotic properties of this Wald 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 suggest rules for setting the autoregressive parameter. Although we are primarily concerned with moderate size cross-sections, all the tests can be implemented when the number of units in the cross-section exceeds the length of the time series. Explanatory variables can also be introduced into the model. We note that the homogeneous model is an attractive one to adopt since it leads to simple

[1]A different approach, based on combining nonlinear IV estimators for each equation, is taken in Chang (2002). Phillips and Sul (2002) generalise the test proposed by Maddala and Wu (1999) to deal with certain types of cross-dependence.

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). However, our main focus is on converging economies.

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 that 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 or balanced growth. 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. Section 2 sets the scene by defining balanced growth and reviewing how the null hypothesis of balanced growth or stability can be tested by a Lagrange multiplier test. This test is derived under the assumption that there is only one extra parameter under the alternative of nonstationarity. A similar device is used to develop the multivariate homogeneous Dickey-Fuller test studied in section 3. 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 and the consequences of ignoring cross-correlation are explored. Section 4 deals with the tests when constants are included and sets out the ERS modification. Section 5 presents set of Monte Carlo experiments, one purpose of which is to compare the performance of the proposed tests with an unrestricted likelihood ratio test.

The effects of initial conditions on the size and power of various unit root tests is studied by Monte Carlo experiments in section 6. Section 7 discusses the tests in the context of convergence, starting with a model in levels and then deriving the implications for contrasts. The case for tests based on the homogenous multivariate model is then made, with data on regions of the United States used as an illustration. Section 8 reports a brief investigation into the properties of the tests when there is only partial convergence and section 9 concludes.

2 Stationarity tests and balanced growth

The multivariate stationarity test helps to set the scene for the statistical development of the multivariate unit root tests. In addition the way in which it is used to test for stability in the context of balanced growth is important for understanding the role played by unit root tests in testing for convergence.

2.1 Stationarity Tests

Consider a multivariate unobserved components model for N time series consisting of a random walk plus noise for a set of observations, $\mathbf{y}_t$:

$$\mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \boldsymbol{\eta}_t, \quad t = 1, ..., T, \tag{1}$$

where $\mu_0 = 0$ and the $\boldsymbol{\eta}_t's$ and $\boldsymbol{\varepsilon}_t's$ are mutually and serially uncorrelated Gaussian $N \times 1$ disturbance vectors with covariance matrices $\mathbf{\Sigma}_\eta$ and $\mathbf{\Sigma}_\varepsilon$ respectively. Nyblom and Harvey (2000) show that if $\mathbf{\Sigma}_\eta = q\mathbf{\Sigma}_\varepsilon$, where q is a scalar, the LM test of $q = 0$ against $q > 0$ has rejection region

$$\eta_1(N) = T^{-2} \sum_{i=1}^{T} \left[\sum_{t=1}^{i} \mathbf{e}_t \right]' \widehat{\mathbf{\Sigma}}_\varepsilon^{-1} \left[\sum_{t=1}^{i} \mathbf{e}_t \right] > c, \tag{2}$$

where $\mathbf{e}_t = \mathbf{y}_t - \overline{\mathbf{y}}$ and $\widehat{\mathbf{\Sigma}}_\varepsilon = T^{-1} \sum_{t=1}^{T} \mathbf{e}_t \mathbf{e}_t{}'$. Under the null hypothesis, the limiting distribution of $\eta(N)$ is (first level) Cramér-von Mises with N degrees of freedom, denoted $CvM_1(N)$. If there are no constant terms in the model, that is $\boldsymbol{\alpha} = \mathbf{0}$, $\mathbf{e}_t = \mathbf{y}_t$ and the distribution of $\eta_0(N)$ statistic is (zero level) Cramér-von Mises, $CvM_0(N)$. This distribution is such that

$$CvM_0(N) = \sum_{i=1}^{N} \int_0^1 W_i(r)^2 dr \tag{3}$$

where $W_i(r), i = 1, ..N$ are independent standard Wiener processes. In the first level distribution, $CvM_1(N)$, the Wiener processes are replaced by Brownian bridges, $B_i(r) = W_i(r) - rW_i(1)$, $i = 1, ..N$. Although the tests are derived under Gaussianity, it is sufficient for the observations to be martingale differences (with finite variance) to yield these asymptotic distributions.

If $\boldsymbol{\varepsilon}_t$ is serially correlated, a nonparametric estimator of the long-run covariance matrix can be used in place of $\widehat{\mathbf{\Sigma}}_\varepsilon$ without changing the asymptotic distribution. Parametric adjustments for serial correlation can also be made.

The *homogeneity* assumption, $\mathbf{\Sigma}_\eta = q\mathbf{\Sigma}_\varepsilon$, permits the derivation of a simple test that will nevertheless be powerful against a wide range of nonhomogeneous alternatives. Nyblom and Harvey (2000) show that if the series contain K unit roots, then $T^{-1}\eta(N)$ converges to a nondegenerate limiting distribution that depends only on K. Hence the test is consistent.

2.2 Balanced growth

The *balanced growth* UC model for $N + 1$ series is

$$\mathbf{y}_t^\dagger = \mathbf{i}_{N+1}\mu_t + \boldsymbol{\alpha}^\dagger + \boldsymbol{\varepsilon}_t^\dagger, \ , \qquad t = 1, ..., T, \tag{4}$$

where μ_t, is a univariate stochastic trend, such as a random walk plus drift,

$$\mu_t = \mu_{t-1} + \beta + \eta_t, \tag{5}$$

μ_0 is fixed, $\mathbf{i}_{N+1}$ is an $(N + 1) \times 1$ vector of ones, $\boldsymbol{\alpha}^\dagger$ is an $(N + 1) \times 1$ vector of constants, constrained so as to have only N free parameters, and $\boldsymbol{\varepsilon}_t^\dagger$ is an $(N+1) \times 1$ vector of jointly stationary processes. Note that although the levels may be different, all the series have the same slope, β.

A balanced growth model implies that the series have a stable relationship over time. This means that there is a full rank $N \times (N+1)$ matrix, $\mathbf{D}$, with no null columns and the property that $\mathbf{Di} = \mathbf{0}$, thereby eliminating the common trend and rendering the N series in $\mathbf{Dy}_t^\dagger$ jointly stationary. The rows of $\mathbf{D}$ are therefore balanced growth co-integrating vectors, while the elements of $\mathbf{Dy}_t^\dagger$ 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 $\mathbf{M} = \mathbf{I}_{N+1} - (N+1)^{-1}\mathbf{i}_{N+1}\mathbf{i}_{N+1}{}'$.

The test of stability is carried out by applying the multivariate stationarity test to $\mathbf{Dy}_t^\dagger = \mathbf{y}_t$. The choice of $\mathbf{D}$ does not matter since pre-multiplication of $\mathbf{Dy}_t^\dagger$ by a non-singular $N \times N$ matrix leaves the test statistic unchanged; Kuo and Mikkola (2001) note the importance of this property in testing the hypothesis of PPP. Under the null hypothesis, the limiting distribution of $\eta(N)$ is Cramér-von Mises with N degrees of freedom, $CvM(N)$. If there are no constant terms in (4), that is $\boldsymbol{\alpha}^\dagger = \mathbf{0}$, the series contain an *identical* common trend and the stability test statistic is $\eta_0(N)$; see Hobijn and Franses (2000).

3 Multivariate unit root tests

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

$$\mathbf{y}_t = \mathbf{\Phi y}_{t-1} + \boldsymbol{\eta}_t, \quad t = 1, ..., T, \tag{6}$$

where $\mathbf{y}_0$ is fixed, but unknown, $\mathbf{\Phi}$ 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 $\mathbf{\Sigma}_\eta$. The *homogeneous* model sets $\boldsymbol{\phi} = \phi\mathbf{I}_N$, where ϕ is a scalar. As will be seen, this restriction has a number of statistical attractions.

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. All of these test statistics are invariant to pre-multiplication of $\mathbf{y}_t$ by a nonsingular $N \times N$ matrix. The extension to cases where $\boldsymbol{\eta}_t$ is serially correlated is straightforward.

If $\mathbf{\Phi}$ is not restricted, a LR test of the hypothesis that $\mathbf{\Phi}$ contains N unit roots against the alternative of stationarity can be constructed as in Johansen (1988, 1995). This test is contrasted with the proposed tests in a series of Monte Carlo experiments reported later. Another possibility is to have $\mathbf{\Phi}$ diagonal. Taylor and Sarno (1998) investigate the properties of a Wald test of this hypothesis in the context of PPP, while Phillips and Sul (2002) derive the asymptotic distribution. However, the heterogeneous model has some drawbacks, one of which is that the diagonality of $\mathbf{\Phi}$ is lost when the observations are pre-multiplicatied by a nonsingular $N \times N$ matrix. The homogeneous model, on the other hand, retains its structure.

3.1 Multivariate homogeneous Dickey-Fuller test

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

$$\widetilde{\pi} = \sum_{t=2}^{T} \mathbf{y}_{t-1}' \widetilde{\mathbf{\Sigma}}_\eta^{-1} \Delta \mathbf{y}_t / \left[\sum_{t=2}^{T} \mathbf{y}_{t-1}' \widetilde{\mathbf{\Sigma}}_\eta^{-1} \mathbf{y}_{t-1} \right] \tag{7}$$

where $\widetilde{\mathbf{\Sigma}}_\eta = T^{-1} \sum_{t=2}^{T} (\Delta \mathbf{y}_t - \widetilde{\pi} \mathbf{y}_{t-1})(\Delta \mathbf{y}_t - \widetilde{\pi} \mathbf{y}_{t-1})'$. 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'

$$\tau_0(N) = \sum_{t=2}^{T} \mathbf{y}_{t-1}' \widetilde{\mathbf{\Sigma}}_\eta^{-1} \Delta \mathbf{y}_t / \left[\sum_{t=2}^{T} \mathbf{y}_{t-1}' \widetilde{\mathbf{\Sigma}}_\eta^{-1} \mathbf{y}_{t-1} \right]^{1/2}. \tag{8}$$

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 $\widetilde{\mathbf{\Sigma}}_\eta$ - that is, premultiplying $\mathbf{y}_t$ by $\widetilde{\mathbf{\Sigma}}_\eta^{-1/2}$– and then simply applying OLS to the pooled observations. Maximum likelihood estimation requires iterating to convergence starting with $\widetilde{\pi} = 0$. However, the estimator of π will have the same asymptotic distribution if $\mathbf{\Sigma}_\eta$ is estimated with $\widetilde{\pi}$ replaced by a consistent estimator, for example as obtained from the first iteration. We will refer to this as the *two-step* estimator. O'Connell (1998) estimates $\mathbf{\Sigma}_\eta$ from first differences, that is $\widetilde{\pi} = 0$; 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 $\widetilde{\pi}$ and $\tau_0(N)$ respectively under the null hypothesis.

As in the univariate case, the unit root test can be based directly on the coefficient $\widetilde{\pi}$ in (7), since the statistic $\pi_0(N) = T\widetilde{\pi}$ has a known limiting distribution. The potential gain in power is immediately apparent from the fact that, for a stationary model, the asymptotic variance of $\widetilde{\pi}$ is the asymptotic variance in a univariate model divided by N.

3.2 Asymptotic distribution under the null hypothesis

The invariance of the test statistics, together with the fact that $\mathbf{\Sigma}_\eta$ is estimated consistently, means that their asymptotic properties under the null hypothesis can be derived by letting $\mathbf{\Sigma}_\eta$ 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. They correspond to statistics obtained from a pooling of the data as in Levin, Lin and Lu (2002).

The asymptotic distribution of $\tau_0(N)$ depends only on N since

$$\begin{aligned}\tau_0(N) &\rightarrow \frac{\sum_{i=1}^N \int_0^1 W_i(r)dW_i(r)}{\left[\sum_{i=1}^N \int_0^1 W_i(r)^2dr\right]^{1/2}} = \frac{(1/2)\sum_{i=1}^N (W_i(1)^2-1)}{\left[\sum_{i=1}^N \int_0^1 W_i(r)^2dr\right]^{1/2}} \\ &= \frac{(1/2)(\chi^2_N - N)}{\{CvM_0(N)\}^{1/2}}\end{aligned} \tag{9}$$

where $W_i(r)$ and $CvM_0(N)$ are as defined in sub-section 2.1. (In writing down the last representation it must be remembered that the distributions depend on the same Wiener processes and so are not independent.) The simplicity of (9) is attractive. By contrast, the asymptotic distribution of the Wald test in the heterogeneous model depends on $\mathbf{\Sigma}_\eta$; see Phillips and Sul (2002).

The distribution of $\pi(N)$ is the same as in (9) except that there is no square root in the denominator. As with the multivariate stationarity test, the condition of independent Gaussian disturbances can be weakened to that of martingale differences; see, for example, Stock (1994).

The $\tau_0(N)$ statistic converges to a standard normal as N increases, that is $\tau_0(N) \rightarrow N(0,1)$. Levin, Lin and Lu (2002) prove the result by letting N be a function of T, such that $\sqrt{N}/T \rightarrow 0$ as $T \rightarrow \infty$; see also appendix B.

3.3 Critical values

Table 1a shows asymptotic critical values obtained by simulating directly from (9), 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 $\mathbf{\Sigma}_\eta = \mathbf{I}$ and $T = 500$ and then calculating $\tau_0(N)$ with $\widetilde{\mathbf{\Sigma}}_\eta$ set to $\mathbf{I}$ in the test statistic for gives virtually the same results. There is clear evidence that the distribution is approaching normality as N is increasing.

Table 2 shows 5% critical values for the MHDF t-test for $T = 100$ and 500 with $\mathbf{\Sigma}_\eta = \mathbf{I}$ but $\widetilde{\mathbf{\Sigma}}_\eta$ obtained by iterating to convergence as indicated below (8). 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 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 $\mathbf{\Sigma}_\eta$.

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
$T = 100$,Two Step	-1.94	-1.88	-1.86	-1.88

3.4 Local power

Premultiplying a homogeneous model through by $\mathbf{\Sigma}_\eta^{-1/2}$ 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 (9). Thus

$$\tau_0(N) \rightarrow \frac{\sum_{i=1}^{N} \int_0^1 W_{ci}(r) dW_i(r)}{\left[\sum_{i=1}^{N} \int_0^1 W_{ci}(r)^2 dr\right]^{1/2}} = \frac{(1/2) \sum_{i=1}^{N} (W_{ci}(1)^2 - 1)}{\left[\sum_{i=1}^{N} \int_0^1 W_{ci}(r)^2 dr\right]^{1/2}} \tag{10}$$

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

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

each satisfying the stochastic differential equation

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

with $W_{ci}(0) = 0$; note that $W_{0i}(.) = W_i(.)$. Following ERS(1996, theorem 1) the local asymptotic power function[2] for the Neyman-Pearson most powerful test against the alternative $c = \overline{c}$ and having significance level is ϵ is given by

$$P(c, \overline{c}) = \Pr\left[\overline{c}^2 \sum_{i=1}^{N} \int W_{ci}(r)^2 dr - \overline{c} \sum_{i=1}^{N} W_{ci}(1)^2 < b(\overline{c})\right] \tag{11}$$

where $b(\overline{c})$ satisfies $\Pr\left[\overline{c}^2 \sum_{i=1}^{N} \int W_i(r)^2 dt - \overline{c} \sum_{i=1}^{N} W_i(1)^2 < b(\overline{c})\right] = \epsilon$. 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 $\mathbf{\Sigma}_\eta$ 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).

3.5 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

$$\zeta_0(N) = \frac{1}{T^2} \sum_{t=1}^{T} \mathbf{y}_t' \widehat{\mathbf{\Sigma}}_\eta^{-1} \mathbf{y}_t, \tag{12}$$

[2]The theory follows from appendix B of ERS except that there are determinants of estimators of $\mathbf{\Sigma}_\eta$ present; these cancel under the local alternative.

where $\widehat{\mathbf{\Sigma}}_\eta = T^{-1}\sum_{t=2}^{T}\Delta\mathbf{y}_t\Delta\mathbf{y}_t'$. 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 test may be referred to as 'LM type'.

The $\zeta_0(N)$ statistic is asymptotically distributed as $CvM_0(N)$ under the null hypothesis. This is the same distribution as for the stationarity test statistic, $\eta_0(N)$, but it is now the lower tail that 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.

3.6 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 $y_{t-1}, \Delta y_{t-j}$, $j = 1, .., p$. 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 $T(\widehat{\phi}-1)/(1-\sum_1^p \widehat{\phi}_j^*)$ where $\widehat{\phi}_j^*$ 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 (8). A similar device is used by Johansen (1988, 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.[3] Correcting the direct

[3]Introduction 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

coefficient test is less straightforward[4] and is not something we have investigated.

Serial correlation might arise from $\boldsymbol{\eta}_t$ or from the presence of a stationary component added to (6). In the latter case the autoregressive approximation may not be satisfactory; indirect evidence on this matter may be found in the simulations reported by Stock (1994, table 1) for moving average processes. Again this is not an issue that will be pursued here.

3.7 Singular covariance matrices and large cross-sections

If, in any of the above test statistics, the estimate of $\mathbf{\Sigma}_\eta$ is singular, the statistic cannot be computed. This will happen if $N > T$ or if some series are perfectly correlated. Suppose, more generally, that the rank of the estimate of $\mathbf{\Sigma}_\eta$ is M. Let $\mathbf{\Lambda}$ denote the $M \times M$ diagonal matrix containing the non-zero eigenvalues of this estimate and let $\mathbf{P}$ be the corresponding $N \times M$ matrix of eigenvectors. Define the $M \times 1$ vector of transformed observations $\mathbf{z}_t = \mathbf{\Lambda}^{-1/2}\mathbf{P}'\mathbf{y}_t, t = 1, ..., T$. The test statistics are then formed by dropping $\widetilde{\mathbf{\Sigma}}_\eta$, or $\widehat{\mathbf{\Sigma}}_\eta$, and replacing $\mathbf{y}_t$ by $\mathbf{z}_t$; in other word we simply pool the transformed observations. (If $\widetilde{\mathbf{\Sigma}}_\eta$ were of full rank we would be inverting it by the singular value decomposition). The test statistics then have the same properties as the corresponding test statistics indexed M, for example $\tau_0(N)$ is distributed as $\tau_0(M)$. The same device may be used in the multivariate stationarity test of sub-section 2.1.

More generally we might wish to avoid numerical instability by setting very small eigenvalues to zero.

3.8 Effect of ignoring cross-correlations

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 wil be distorted. To analyse the effect of cross-correlations on the asymptotic distribution of the pooled statistic,

of the covariance matrix. The statistics reported for US regions in section 7 were computed in this way.

[4]There are other reasons, concerned with initial conditions and discussed in section 6, as to why the direct coefficient test is not practical for testing for convergence.

τ_0^*, under the null hypothesis, we make the transformation $\mathbf{y}_t^* = \boldsymbol{\Sigma}_\eta^{-1/2}\mathbf{y}_t$ and write

$$\tau_0^* = \sum_{t=2}^{T} \mathbf{y}_{t-1}^{*\prime} \boldsymbol{\Sigma}_\eta \Delta \mathbf{y}_t^* / \left[\sum_{t=2}^{T} \mathbf{y}_{t-1}^{*\prime} \boldsymbol{\Sigma}_\eta \mathbf{y}_{t-1}^* \right]^{1/2} \tag{13}$$

Given the independence of the elements of $\mathbf{y}_t^*$ the asymptotic distribution of τ_0^* can be found for a particular structure of $\boldsymbol{\Sigma}_\eta$ by the spectral decomposition. If $\lambda_i, i = 1, .., N$ denote the eigenvalues of $\boldsymbol{\Sigma}_\eta$, then

$$T^{-2} \sum_{t=2}^{T} \mathbf{y}_{t-1}^{*\prime} \boldsymbol{\Sigma}_\eta \mathbf{y}_{t-1}^* \rightarrow \sum_{i=1}^{N} \lambda_i \int_0^1 W_i(r)^2 dr = \sum_{i=1}^{N} \lambda_i CvM_{i0}(1)$$

and

$$T^{-2} \sum_{t=2}^{T} \mathbf{y}_{t-1}^{*\prime} \boldsymbol{\Sigma}_\eta \Delta \mathbf{y}_{t-1}^* \rightarrow \sum_{i=1}^{N} \lambda_i \int_0^1 W_i(r) dW_i(r) \ = (1/2) \sum_{i=1}^{N} \lambda_i (\chi_{1i}^2 - 1)$$

Thus the asymptotic distribution of τ_0^* depends only on the eigenvalues of $\boldsymbol{\Sigma}_\eta$. The same is true of the distributions of the direct coefficient test statistic, $\pi_0^*(N)$, and the Sargan-Bhargava statistic. Local power can also be obtained by a corresponding weighting in (10).

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

$$\boldsymbol{\Sigma} = (1-\omega)\mathbf{I}_N + \omega \mathbf{i}_N \mathbf{i}_N' = \mathbf{E}_N(\omega) \tag{14}$$

where ω is a parameter that gives the correlation. In this case one eigenvalue is $N\omega + 1$ while the rest are $1 - \omega$. As a result the $N\omega + 1$ eigenvalue dominates. It is shown in appendix B that $\tau_0(N)$ needs to be divided by $N^{1/2}$ if it is to have a limiting distribution as $N \rightarrow \infty$.

In his paper on PPP, O'Connell (1998) starts off by assuming a covariance matrix of the form $\mathbf{E}_N(\omega)$. However, if the N series in $\mathbf{y}_t$ are a set of balanced growth contrasts as in sub-section 2.2, it may be better 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. This may be motivated by a two-factor disturbance structure in which each disturbance has a common and a specific part[5]. The implied covariance matrices for benchmark and deviation from the mean contrasts are as follows.

Benchmark- For any benchmark set of contrasts

$$\mathbf{DE}_{N+1}(\omega)\mathbf{D}' = (1-\omega)(\mathbf{I}_N + \mathbf{i}_{N+1}\mathbf{i}'_{N+1}) = 2(1-\omega)\mathbf{E}_N(0.5) \tag{15}$$

Thus whatever the original correlation between the series, the correlation between the benchmarked contrasts is 0.5. The eigenvalues are all 0.5 except for one which is $1 + N/2$. The simulations in O'Connell (1998, table 1) imply that for a test at the nominal 5% level of significance, the true size is around 0.17 for $N = 10$ and 0.38 for $N = 50$.

Deviation from the mean -Since

$$\mathbf{ME}_{N+1}(\omega)\mathbf{M} = (1-\omega)\mathbf{M}$$

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 $\mathbf{\Sigma} = \mathbf{I}_N - (N+1)^{-\mathbf{1}}\mathbf{i}_N\mathbf{i}'_N$. Unlike $\mathbf{M}$, this matrix is no longer singular. Its eigenvalues are all unity except for one which is $1/(N+1)$. Thus for moderate size N, the distribution of the τ_0^* statistic will be better approximated by the distribution of a $\tau_0(N-1)$ statistic than by that of a $\tau_0(N)$ statistic. Since both converge to the same normal distribution as $N \to \infty$, the difference will not be great unless N is small. However, once we move away from the equal correlation case, the eigenvalues can be very different and these may lead to considerable size distortion.

O'Connell notes that the mean is sometimes removed from exchange rate data, this being the same as including time dummies in the model. Exchange rate data typically uses one currency as a benchmark and so the structure might be expected to be as in (15) above. Removing the means from N such series creates a covariance matrix with all correlations equal to $-1/(N-1)$.

[5]More generally the weights assigned to the common part may differ across units as in Barro and Sala-i-Martin (1992) and Phillips and Sul (2002).

4 Constant

If the mean of $\mathbf{y}_t$ is the $N \times 1$ vector $\boldsymbol{\alpha}$, the model in (6) becomes

$$\mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\mu}_t, \qquad \boldsymbol{\mu}_t = \boldsymbol{\Phi}\boldsymbol{\mu}_{t-1} + \boldsymbol{\eta}_t, \quad t = 1, ..., T. \tag{16}$$

Again the prime concern is with the homogeneous model in which $\boldsymbol{\Phi} = \phi\mathbf{I}$.

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 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.

All of the tests below are invariant to linear transformations of the data[6].

4.1 Asymptotic distribution and critical values

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

$$\tau_1(N) \to \frac{\sum_{i=1}^N \int_0^1 W_i(r)dW_i(r) \; - \; \sum_{i=1}^N W_i(1) \int_0^1 W_i(r)dr}{\left[\sum_{i=1}^N \int_0^1 W_i(r)^2 dr - \sum_{i=1}^N \left(\int_0^1 W_i(r)dr \right)^2 \right]^{1/2}} \tag{17}$$

As in the no constant case this is obtained straightforwardly from the asymptotics in the univariate case. However, in contrast to the no constant case, the $t-$statistic does not converge to a standard normal; Levin, Lin and Lu (2002, p8) obtain a modified statistic that does.

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

[6] That is $\mathbf{a} + \mathbf{B}\mathbf{y}$, where $\mathbf{a}$ and $\mathbf{B}$ are, respectively, a fixed vector and matrix. The tests of section 3 are not, of course, invariant to shifts brought about by $\mathbf{a}$.

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.

Local asymptotic theory goes through as before. As in the no constant case, the practical implication is that so long as $\mathbf{\Sigma}_\eta$ 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 $\mathbf{\Sigma}_\eta$ has little effect in small samples.

4.2 Multivariate Sargan-Bhargava test

The multivariate homogeneous Sargan-Bhargava test statistic is

$$\zeta_1(N) = \frac{1}{T^2}\sum_{t=1}^{T}\widetilde{\boldsymbol{\mu}}_t'\widehat{\mathbf{\Sigma}}_\eta^{-1}\widetilde{\boldsymbol{\mu}}_t, \tag{18}$$

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

4.3 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 (8) by $\mathbf{y}_t - \widehat{\boldsymbol{\alpha}}_c$, where

$$\widehat{\boldsymbol{\alpha}}_c = \left[\mathbf{y}_1 + (1-\overline{\phi})\sum_{t=2}^{T}(\mathbf{y}_t - \overline{\phi}\mathbf{y}_{t-1})\right] / [1 + (T-1)(1-\overline{\phi})^2] \tag{19}$$

and $\overline{\phi} = 1 + \overline{c}/T$. This detrending is based on GLS estimation[7], assuming that $\boldsymbol{\mu}_0 = \mathbf{0}$.

[7]Note that the covariance matrix is irrelevant as it cancels. Of course it is still needed in in (8).

For reasons given below, the covariance matrix, $\mathbf{\Sigma}_\eta$, is best estimated using the ML estimator of ϕ rather than setting it to $\overline{\phi}$.

In the univariate case, ERS show that the local 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. [(3.23k) in Stock (1994, p2772-3) is same as 3.23b in the no constant case]. 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* (10). Similarly, the local asymptotic representation of $\tau_1(N)$ is obtained by generalising (3.23b) in Stock along the lines of (10). 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 $P(c, \overline{c})$ is as in (11). 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 $\overline{c} = -7$ 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

Setting $\overline{c} = 0$ in (19) gives $\widehat{\boldsymbol{\alpha}}_c = \mathbf{y}_1$. The square of the one-step version of the $\tau - GLS_0(N)$ statistic, that is with $\mathbf{\Sigma}_\eta$ 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 (28). At the other extreme, setting $\overline{c} = -T$ simply de-means, resulting in a statistic that is approximately the same as DF and has the same asymptotic distribution. This implies that $\overline{c}$ should not be too far from zero if the size of the test is required to be close to the nominal.

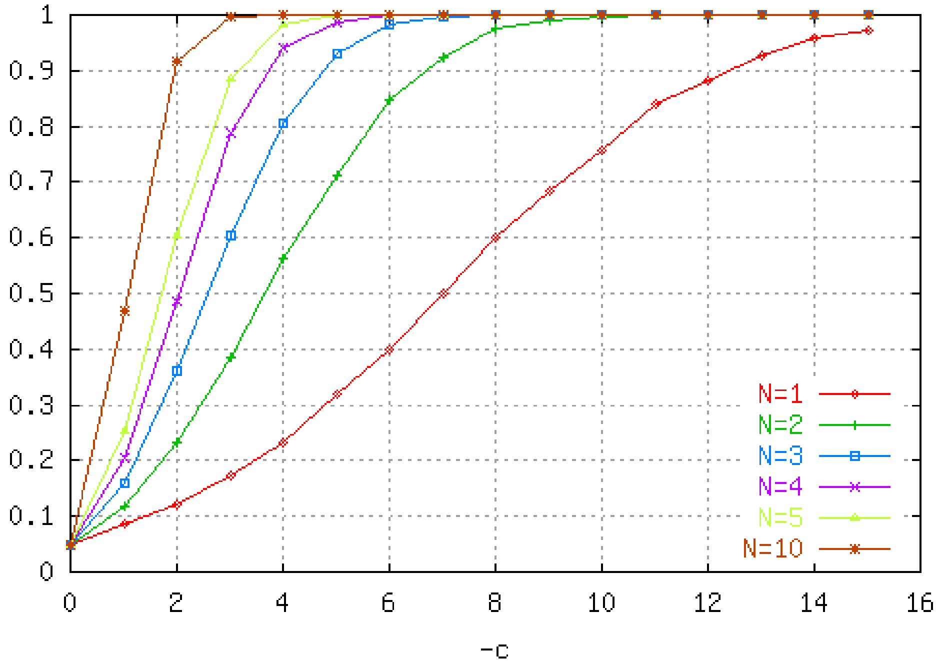

Figure 1: Power envelopes for different N

Table 4 shows the results of a set of Monte Carlo experiments designed to investigate this matter. As can be seen, the values of $\overline{c}$ 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 $\overline{c}$ is set give a local power of 0.5.

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

(a) T=100. Asymptotic critical values

					$\tau - GLS_c(N)$	$Values\ of - c$		
N	$\tau_0(N)$	$\tau_1(N)$	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

					$\tau - GLS_c(N)$	$Values\ of - c$		
N	$\tau_0(N)$	$\tau_1(N)$	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

					$\tau - GLS_c(N)$	$Values\ of - c$		
N	$\tau_0(N)$	$\tau_1(N)$	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 $\mathbf{\Sigma}_\eta$, using $\overline{\phi}$ 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 $\overline{c}$ to zero, which is essentially the same as doing the multivariate SB test. Alternatively $\overline{c}$ 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 far 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 $T = 200$ and $N = 10$, the size of the $\tau_1(N)$ test is 0.070 as opposed to 0.052 for $\tau_0(N)$.

4.4 Time trend

Jorion and Sweeney (1996) extend the MHDF test by letting (16) 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 B.

4.5 Explanatory variables

The homogeneous model, (16), may be extended so as to include explanatory variables that represent (permanent) characteristics of different units and attempt to explain the differences between the series, that is

$$\mathbf{y}_t = \boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}_t, \qquad t = 1, ..., T, \tag{20}$$

where $\mathbf{X}$ is an $N \times k$ matrix and $\boldsymbol{\beta}$ is the corresponding $k \times 1$ vector of coefficients. (The variables may originally be defined for the series before a balanced growth transformation, so that $\mathbf{X} = \mathbf{D}\mathbf{X}^{\dagger}$). Estimation of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and π can again be carried out by applying OLS to the pooled observations after the transformation by $\widetilde{\boldsymbol{\Sigma}}_{\eta}^{-1/2}$. The MHDF-GLS test may be implemented by estimating $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ from pooled observations that have been subjected to the transformation $(1-\overline{\phi})\mathbf{y}_t$, $(1-\overline{\phi})\mathbf{i}$ and $(1-\overline{\phi})\mathbf{X}$, for $t = 2, .., T$, together with the original observations at $t = 1$, that is $\mathbf{y}_1$, $\mathbf{i}$ and $\mathbf{X}$.

Note that if $N > T$, estimation may still be carried out by using the eigenvector transformation of sub-section 3.7.

5 Power in small samples

A series of Monte Carlo experiments were carried out to examine the power of the DF tests, and the ERS modifiction, 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

				$T = 100$			
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	$\tau_0(N)$	$T=200$ $\tau_1(N)$	LR_0	LR_1	$\tau_0(N)$	$T=500$ $\tau_1(N)$	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	$\phi_2 = .. = \phi_N$	$\tau_0(N)$	$T=100$ $\tau_1(N)$	LR_0	LR_1	$\tau_0(N)$	$T=200$ $\tau_1(N)$	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 6 in section 6) 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 $\phi_i's$ are 0.95 or 0.90. As expected, its performance is relatively better for mixed cases; for example, compare the relative performance when all $\phi_i's$ are 0.95 with what happens when ϕ_1 is 0.98 and the other $\phi_i's$ are 0.90.

6) The power of the MHDF-GLS test, with $\overline{c}$ 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 $\overline{c}$. Of particular importance is the case of $\overline{c} = 0$, 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 $\overline{c} = 0$. Thus there is a strong case for basing a test on $\overline{c} = 0$. Alternatively, the SB test might be used since this behaves in a very similar fashion.

6 Initial conditions

This section examines the impact of initial conditions in the univariate model

$$y_t = \alpha + \mu_t, \qquad \mu_t = \phi\mu_{t-1} + \eta_t, \quad t = 1, ..., T. \tag{21}$$

Rather than assuming that, when $|\phi| < 1$, μ_0 is drawn from a distribution with mean zero and variance $\sigma_\eta^2/(1 - \phi^2)$, we suppose that μ_0 is fixed. This makes no difference to the asymptotic distribution under the alternative, but it does affect small sample behaviour and it is of crucial importance in the context of convergence. Reference should be made to Tanaka (1996, p91) and ERS, p181-20 for a discussion of the assumptions on initial conditions needed for the DF-GLS asymptotic theory to go through.

6.1 Monte Carlo experiments

Table 6 shows the probability of rejecting the null hypothesis of a unit root for $T = 100$ and a range of starting values, μ_0. Asymptotic critical values are used. The level parameter, α, is set to zero in the data generating process. 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. This 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).

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

Table 6 Estimated rejection probabilities for different initial conditions

.

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.

The transitional dynamics actually help in that the powers of the τ_0 and τ_1 increase the further is the starting value from the final equilibrium, that is the unconditional mean.

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 $\overline{c}$ 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 5, 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.

6.2 Test on the constant

The estimator of α obtained by regressing Δy_t on a constant and y_{t-1} is

$$\widetilde{\alpha} = \frac{y_T - y_1}{(T-1)(1-\phi)} + \frac{\sum_{t=1}^{T-1} y_t}{T-1}.$$

Given ϕ, this is an unbiased estimator of α. On the other hand, the sample mean is biased and the bias could be quite big if T is small and $|\mu_0|$ is large. Note that like ϕ, α is more accurately estimated with big initial conditions; see Hendry (2001).

Similar considerations apply in the multivariate case. If the null of a unit root is rejected, we might wish to test the joint significance of the constant terms. This can be conveniently done if estimation is carried out by pooling the data using the standard 'F-statistic'. This will be asymptotically χ^2_N when multiplied by N. Alternatively the LR statistic can be used as this can easily be constructed from the estimates of $\mathbf{\Sigma}_\eta$ obtained when the $\tau_0(N)$ and $\tau_1(N)$ statistics are calculated.

7 Testing convergence

Two series *have converged* if the difference between them, y_t, is stable. 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. If the mean of y_t is zero the countries are in a state of *absolute convergence.* If the mean, α, is not zero we have *conditional* or *relative convergence.*

The error correction mechanism

$$y_t = \alpha + \mu_t, \qquad \Delta\mu_t = \pi\mu_{t-1} + \eta_t, \quad t = 1, ..., T, \tag{22}$$

captures the process of 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. High initial values of μ_t actually help to estimate π more accurately and, as was shown in the previous section, the corresponding DF t-tests have more power. 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. Indeed if testing stability is the aim, stationarity tests will often be more appropriate[8].

The purpose of this section is to explore various aspects of the convergence literature in the light of the statistical results of the previous section. Multivariate methods are then used to investigate convergence of US regions.

[8] Hobijn and Franses (2000) use stationarity tests but then add to the confusion by saying that they are testing whether the countries "*are converging*"

7.1 Beta and sigma convergence

Writing the model in EC form, (22), accords with the notion of convergence in the cross-sectional literature, as expounded by Barro and Sala-i-Martin (1992, 1995) and others, except that there the growth rate is taken to be a linear function of an arbitrarily chosen initial value, giving a model which is internally inconsistent over time since the growth rate doesn't change as the gap narrows; see also Evans and Karras (1996, p 253). More specifically Barro and Sala-i-Martin average growth rates over time and then regress on initial values. Averaging growth rates in (22) on the other hand gives

$$(T-1)^{-1}\sum_{t=2}^{T}\Delta y_t = -\pi\alpha + \pi(T-1)^{-1}\sum_{t=2}^{T} y_{t-1} + (T-1)^{-1}\sum_{t=1}^{T}\eta_t$$

so that the average $(T-1)^{-1}\sum_{t=2}^{T} y_{t-1}$ appears instead of y_1. The approximation may not matter much for small T. Indeed for $T = 2$, the coefficient of y_1 in a cross-sectional regression, normally referred to as β, corresponds exactly to $-\pi$ in the multivariate homogeneous model of (26). In contrast to the GLS estimation of π in (7), cross-sectional estimation of β is carried out by OLS. *Unless the series are mutually independent, the t-test on the OLS estimate of β will be invalid.*

Sigma convergence is based on tracking the variance of a cross-section over time. If $\mathbf{y}_t$ is a set of $N+1$ contrasts in deviation from the mean form and α is zero, then

$$\sigma_t^2 = \gamma + \phi^2\sigma_{t-1}^2 + \zeta_t, \qquad t = 1, ..., T, \tag{23}$$

where $\sigma_t^2 = \sum_{i=1}^{N+1} y_{it}^2/(N+1)$, $\gamma = \sum_{i=1}^{N+1}\sigma_{\eta i}^2/(N+1)$ and $\zeta_t = \sum_{i=1}^{N+1}\eta_{it}^2/(N+1) + 2\phi\sum_{i=1}^{N+1} y_{i,t-1}\eta_{it}/(N+1) - \sum_{i=1}^{N+1}\sigma_{\eta i}^2/(N+1)$; compare Sala-i-Martin (1996, p 1329). The disturbance has been standardised so as to have a mean of zero. A valid unit root test can still be based on (23), since, although ζ_t will be non-Gaussian, it is a martingale difference. However, a great deal of information is sacrificed by using only one series. Furthermore, power is lost by the need to include a constant term.

7.2 Rates of convergence and initial values in practice

We now consider the implications that plausible values for the rate of convergence and initial conditions have for the power of convergence tests.

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 0.98. Given an initial condition gap of μ_0, the expected gap after τ periods is $\phi^{\tau}\mu_0$. Thus with ϕ =0.98, the half life is 34, while the expected gap is one tenth of the original after 114 periods.

Within the context of convergence, the range of initial values used in the Monte Carlo experiments of section 6 is not unreasonable. For example, the log difference between the richest and poorest US regions, north-east and south-east, in 1930 was about one, while in 1950 it was about 0.5. With a standard deviation, σ_η, of 0.025, one and 0.5 correspond to initial values of 20 and 40. In fact, σ_η is a little below 0.025, so the initial values are higher.

Although table 6 shows that the power of the Dickey-Fuller $t-$test with constant,τ_1, increases with non-zero initial conditions, its power is still very low. When $\phi = .98$, 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 5; 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.

7.3 Invariance, heterogeneity and a model in levels

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 3 and 4 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.

We now elaborate a little on the case for the homogeneity assumption by considering a model in levels. Let

$$\mathbf{y}_t^{\dagger} = \boldsymbol{\alpha}^{\dagger}+\beta\mathbf{i}t + \boldsymbol{\mu}_t{}^{\dagger},$$

with

$$\boldsymbol{\mu}_t{}^\dagger = \boldsymbol{\Psi}\boldsymbol{\mu}_{t-1}^\dagger + \boldsymbol{\eta}_t^\dagger, \qquad Var(\boldsymbol{\eta}_t) = \boldsymbol{\Sigma}_\eta^\dagger \tag{25}$$

The homogeneous model has $\boldsymbol{\Psi} = \phi\mathbf{I} + (1-\phi)\mathbf{i}_{N+1}\boldsymbol{\lambda}'$,where $\boldsymbol{\lambda}$ is an $(N+1)\times 1$ vector of parameters such that $\boldsymbol{\lambda}'\mathbf{i}_{N+1}$=0. The Jacobian of the matrix $(\boldsymbol{\lambda}, \mathbf{D}')'$ is independent of any parameters and pre-multiplying the system by $(\boldsymbol{\lambda}, \mathbf{D}')'$ yields an equation for $\boldsymbol{\lambda}'\mathbf{y}_t$, that is a random walk plus drift common trend as in (5), together with a set of equations for N contrasts:

$$\begin{aligned} \mathbf{D}\mathbf{y}_t^\dagger &= \mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\mu}_t, \\ \boldsymbol{\mu}_t &= \phi\boldsymbol{\mu}_{t-1} + \boldsymbol{\eta}_t \end{aligned} \tag{26}$$

where $\boldsymbol{\mu}_t = \mathbf{D}\boldsymbol{\mu}_t{}^\dagger$ and $\boldsymbol{\alpha}$ is unrestricted. There is only one autoregressive parameter, ϕ, as $\mathbf{D}\boldsymbol{\Psi} = \phi\mathbf{D}\mathbf{I} + (\mathbf{1}-\phi)\mathbf{D}\mathbf{i}\boldsymbol{\lambda}' = \phi\mathbf{D}$. Further discussion on this model can be found in Harvey and Carvalho (2002). *The implication of (26) is that a test of convergence carried out on any set of contrasts is independent of the common trend to which the economies are converging.*

In a heterogenous model, $\mathbf{D}\boldsymbol{\Psi} = \boldsymbol{\phi}\mathbf{D}$, where $\boldsymbol{\phi}$ is diagonal. *Heterogeneity is specific to a particular contrast*[9]. Most tests in the literature are based on a set of deviation from the mean contrasts. Having equal weights 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[10].

Benchmark contrasts also suffer some from the problem of heterogeneity being specific to the constrast. The US is typically taken to be the benchmark; see, for example, Linden (2000). A different benchmark may lead to very different conclusions being drawn.

7.4 Time trends

The multivariate levels model for convergence, (25), has series with a common time trend, as does the balanced growth model (4). The $\mathbf{D}$ matrix eliminates the time trend from the contrasts, but if the time trend had a different coefficient in each series, this would carry over to the constrasts. One could certainly test for convergence using

[9]In general one cannot find a matrix $\mathbf{A}$ such that $\mathbf{AD} = \boldsymbol{\phi}^*\mathbf{AD}$, where $\boldsymbol{\phi}^*$ is also diagonal.

[10]Note that the common trend is not, in general, as implied by the selected contrast, but is a weighted average which depends on $\boldsymbol{\Psi}$; see Harvey and Carvalho (2002).

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 subsection 4.4. The question to be asked is whether testing for convergence to diverging growth paths makes any economic sense in the first place.

7.5 Testing for divergence

The MHDF test with no constant can be used for testing against divergence. The upper tail critical values in table 1 are then appropriate.

7.6 US regions

Carvalho and Harvey (2002) analyse trend and cyclical dynamics in the logarithms of annual real per capita incomes in US census regions from 1950 to 1999. 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). The data were obtained from the Bureau of Economic Analysis and deflated by the US implicit price deflator (1996=100). Carvalho and Harvey (2002) conclude that there is a strong case for absolute convergence amongst all regions apart from NE and ME. Figure 2 shows the series for five regions 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) 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

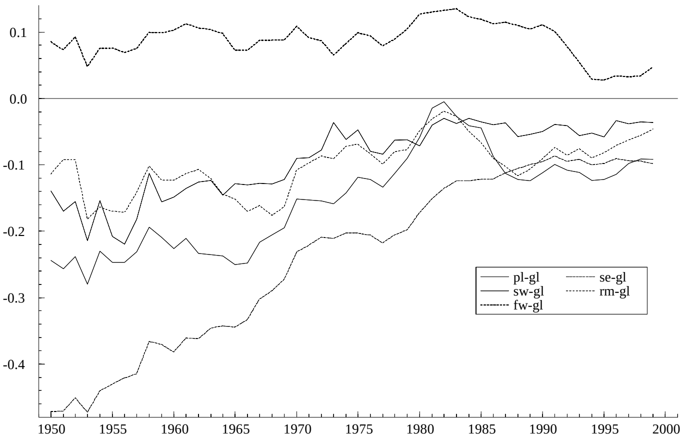

Figure 2: US regions relative to Great Lakes

varied. Adding a time trend only makes matters worse; for example $SE - GL$ gives $t = -.942$ as opposed to -1.876 with a constant.

Table 7 ADF τ_0 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 8 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.

Table 8 Pairwise ADF τ_0 tests (p=4)

	GL	PL	SE	SW	RM
PL	-1.833^{*}				
SE	-2.526^{**}	-2.462^{**}			
SW	-1.144	-1.114	-2.381^{**}		
RM	-1.300	-1.398	-1.967^{**}	-0.924	
FW	-0.748	-1.454	-2.417^{**}	-1.069	-1.181

Table 8 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 $\overline{c}$ 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^2_5 is 9.24.

As well as giving different results with different benchmarks, failing to take account of cross-correlations can results in considerable size distortions; see the analysis in sub-section 3.7. In order to obtain some idea of potential distortions in practice, it is worth examining the cross-correlations in the present example. Of course estimates of the covariance matrix, $\mathbf{\Sigma}_{\eta}$, will depend on the model fitted. However, we found that the cross-correlation matrices of residuals from models with lags were not very different from the cross-correlation matrices computed from raw data. We therefore concentrate on the latter.

The cross-correlation matrix of first differences of the six regions and for the deviations from the mean are respectively :

$$\begin{array}{c} \\ GL \\ PL \\ SE \\ SW \\ RM \\ FW \end{array}\begin{array}{c} \begin{array}{cccccc} GL & PL & SE & SW & RM & FW \end{array} \\ \left[\begin{array}{cccccc} 1 & .50 & .85 & .60 & .56 & .81 \\ & 1 & .56 & .55 & .56 & .46 \\ & & 1 & .73 & .65 & .84 \\ & & & 1 & .68 & .72 \\ & & & & 1 & .64 \\ & & & & & 1 \end{array}\right]\end{array} \quad \begin{array}{c} \begin{array}{cccccc} GL & PL & SE & SW & RM & FW \end{array} \\ \left[\begin{array}{cccccc} 1 & -.37 & .28 & -.48 & -.46 & .16 \\ & 1 & -.42 & -.17 & -.09 & -.56 \\ & & 1 & -.15 & -.37 & .16 \\ & & & 1 & .02 & -.05 \\ & & & & 1 & -.23 \\ & & & & & 1 \end{array}\right]\end{array}$$

.

The correlations for the first differences are quite high ranging from .46 to .85. However, they are far from being the same. Correspondingly, 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. We also examined the covariance matrix for the series with GL as a benchmark, as in figure 2. These ranged between 0.42 and 0.78.

Finally, turning to sigma convergence, a plot shows a sharp fall from a value of .032 in 1950 to one of .003 in 1999. Nevertheless, the DF test with constant[11] is unable to reject at the 10% level of significance: with $p = 0$, the t-statistic was -2.40, while with $p = 4$, it fell to -1.63.

8 Partial convergence and convergence clubs

One notion of convergence clubs is that of absolute convergence within clubs with the clubs themselves displaying relative convergence. Thus overall there is relative convergence, but within the clubs the constants are all zero. Such clubs may, in principle, be detected by pairwise comparisons of estimated constants.

The more usual idea of convergence clubs has absolute or relative convergence within clubs but with no (relative) convergence between the clubs. Relative convergence of the full set of series does not hold, though the trace LR test of Johansen (1988) may be used, as in Bernard and Durlauf (1995), to try to detect the number of

[11] The coefficient of σ^2_{t-1} was -0.034 which corresponds $\phi = .983$. This is very close to the estimate reported for the multivariate homogeneous model.

co-integrating relationships and hence the number of clubs. However, the low power of the LR test reported in section 5 is worrying. The trace test is similarly likely to have low power and hence to indicate too many clubs. For example with the six US regions, the trace test based on four lags cannot reject the null hypothesis of two clubs, that is two common trends, against the alternative that there are fewer at the 5% level of significance and it is barely able to reject three. With two lags it cannot even reject five common trends at the 5% level.[12] The composition of the clubs may, in principle, be detected by pairwise unit root tests, but again the low power of univariate tests, particularly when a constant is included, means that this may not be very effective. Inspecting graphs of long-term movements extracted by unobserved component models, as in Carvalho and Harvey (2002), can play an important role.

What are the likely results of multivariate unit root tests when some of the (subsets of) series do not converge? The results below attempt to throw some light on this question.

8.1 Asymptotic distribution of test statistics for partial convergence

Suppose that

$$\Phi = diag\left\{\phi_1, ..., \phi_N\right\},$$

then if

$$\phi_i = 1, \quad i = 1, ..., K \text{ while } \phi_i = \phi, |\phi| < 1, \quad i = K+1, ..., N,$$

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 $\eta_{it}, i = 1, ..., N$, are distributed independently of each other with the same variance. Then

$$\tau_0(N) \xrightarrow{L} \frac{(1/2)\left(\chi^2_K - K\right) - (N-K)/(1+\phi)}{\left\{CvM_0\left(K\right)\right\}^{1/2}}, \qquad K = 1, ..., N.$$

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

[12]With four lags the trace statistic for the null of two common trends is 20.75, for three it is 43.26. The corresponding 5% critical values for the restricted trend model are 25.3 and 42.4.

is consistent when $K = 0$ is straightforward). However, this probability becomes smaller as $N - K$ increases. A similar analysis could be carried out for the $\tau_1(N)$ and $\tau - GLS_c$ tests.

8.2 Monte Carlo results for multivariate tests

The tables 9a and 9b below are constructed on the same basis as table 5, except that ϕ_1 is unity. The power of the $\tau_0(N)$ test drops considerably when there is one unit root. For example in the bivariate model with $\phi_2 = .95$, the power falls from 0.721 to 0.198 for $T = 100$ and from .995 to .273 for $T = 200$. However, it still quite high for T=500. For $N = 5$, the rejection probabilities are much higher though in the empirically relevant case when there is one unit root and the remaining $\phi's$ are 0.98, the power falls from .605 to .379.

Table 9a Rejection probabilities for N=2

		$T = 100$		T=200		$T = 500$	
ϕ_1	ϕ_2	$\tau_0(N)$	LR	$\tau_0(N)$	LR	$\tau_0(N)$	LR
1	.98	.103	.051	.170	.072	.838	.201
1	.95	.198	.082	.273	.198	.841	.838
1	.90	.283	.200	.318	.643	.840	1

Table 9b Rejection probabilities for N=5

		$T = 100$		T=200		$T = 500$	
ϕ_1	$\phi_2 = .. = \phi_5$	$\tau_0(N)$	LR	$\tau_0(N)$	LR	$\tau_0(N)$	LR
1	.98	.379	.030	.647	.035	.972	.152
1	.95	.692	.040	.763	.143	.973	.968
1	.90	.759	.141	.792	.789	.974	1

9 Conclusions

A joint test of the null hypothesis that a set of series do not have a stable relationship against the alternative that they have converged to a steady state 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. The consequences of ignoring cross-correlations are analysed and it is shown that it can lead to considerable size distortion.

An analysis of local power suggests suitable ways of implementing 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 tabulted for the multivariate Dickey-Fuller test without a constant. When the alternative hypothesis is that the series have converged to the same common trend, the power of the ERS test is roughly the same as that of the no constant Dickey-Fuller test.

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. All the homogeneous tests are viable even if the number of series exceeds the number of time periods and they can be be implemented when there are explanatory variables that are constant over time.

The power of the ERS test is relatively insensitive to the choice of $\overline{c}$, the parameter that fixes the autoregressive coefficient for the purposes of detrending. There is a good case for letting $\overline{c} = 0$, 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. There is much to recommend these tests 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 an ADF regression since the simple mean is distorted by the initial conditions.

The error correction mechanism may be embedded in an unobserved components model, as in Harvey and Carvalho (2002), or in an autoregressive model. The autoregressive model provides a straightforward framework for conducting augmented Dickey-Fuller tests, based on the inclusion of lagged first differences, and this extends to the proposed multivariate tests.[13] 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.

Finally a word of caution should perhaps be sounded. The tests typically carried out on economic time series are often rather blunt instruments and the tests here are no exception. Harvey and Carvalho (2002) set out a descriptive approach based on extracting the long-term movements from an unobserved components model. Coupling the results of multivariate tests and pairwise comparisons with a careful study of the observed long-term movements may be the best way forward.

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 Busetti, 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.

A Multivariate LM and Sargan-Bhargava tests

Consider the model

$$\mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\beta} t + \boldsymbol{\mu}_t, \qquad \boldsymbol{\mu}_t = \boldsymbol{\Phi}\boldsymbol{\mu}_{t-1} + \boldsymbol{\eta}_t, \quad t = 1, ..., T. \tag{27}$$

[13] An assessment of the performance of tests in an unobserved component framework is a topic for future research.

with $\mathbf{\Phi} = \phi\mathbf{I}$. Let $\widetilde{\boldsymbol{\mu}}_t$ denote $\mathbf{y}_t$ detrended under the null hypothesis, that is $\widetilde{\boldsymbol{\mu}}_t = \mathbf{y}_t - \widetilde{\boldsymbol{\alpha}}_0 - \widetilde{\boldsymbol{\beta}}t$, where $\widetilde{\beta} = (\mathbf{y}_T - \mathbf{y}_1)/(T-1)$ and $\widetilde{\boldsymbol{\alpha}}_0 = \mathbf{y}_1 - \widetilde{\boldsymbol{\beta}}$, where $\boldsymbol{\alpha}_0 = \boldsymbol{\alpha} + \boldsymbol{\mu}_0$. Note that $\widetilde{\boldsymbol{\mu}}_1 = \mathbf{0}$, as a consequence of fitting the constant, while $\widetilde{\boldsymbol{\mu}}_T = \mathbf{0}$ because of the slope; $\widetilde{\boldsymbol{\mu}}_0$ 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

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

On evaluating the second derivative, we find that

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

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

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

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

If $\boldsymbol{\Sigma}_\eta$ is estimated by $T^{-1}\sum_{t=1}^{T} \Delta \widetilde{\boldsymbol{\mu}}_t \Delta \widetilde{\boldsymbol{\mu}}_t'$, the left hand side of the above expression reduces to NT because $\sum_{t=1}^{T} \Delta \widetilde{\boldsymbol{\mu}}_t' \widehat{\boldsymbol{\Sigma}}_\eta^{-1} \Delta \widetilde{\boldsymbol{\mu}}_t = tr\left[\widehat{\boldsymbol{\Sigma}}_\eta^{-1} \sum_{t=1}^{T} \Delta \widetilde{\boldsymbol{\mu}}_t \Delta \widetilde{\boldsymbol{\mu}}_t'\right]$ and so, provided the slope is estimated, $\widetilde{\boldsymbol{\mu}}_T = 0$ and it follows that $\sum_{t=1}^{T} \Delta \widetilde{\boldsymbol{\mu}}_t' \widehat{\boldsymbol{\Sigma}}_\eta^{-1} \widetilde{\boldsymbol{\mu}}_{t-1} = -NT/2$. 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,

$$\begin{aligned} \tau - GLS_0(N) &= \sqrt{LM} = \frac{-NT + \widetilde{\boldsymbol{\mu}}_T' \widehat{\boldsymbol{\Sigma}}_\eta^{-1} \widetilde{\boldsymbol{\mu}}_T}{2(\sum_{t=1}^{T} \widetilde{\boldsymbol{\mu}}_{t-1}' \widehat{\boldsymbol{\Sigma}}_\eta^{-1} \widetilde{\boldsymbol{\mu}}_{t-1})^{1/2}} \qquad (28) \\ &= \frac{(\mathbf{y}_T - \mathbf{y}_1)' \widehat{\boldsymbol{\Sigma}}_\eta^{-1} (\mathbf{y}_T - \mathbf{y}_1) - NT}{2(\sum_{t=1}^{T} (\mathbf{y}_t - \mathbf{y}_1)' \widehat{\boldsymbol{\Sigma}}_\eta^{-1} (\mathbf{y}_t - \mathbf{y}_1))^{1/2}}. \end{aligned}$$

The fact that this has the distribution in (9) is immediately apparent.

B Asymptotic distribution of OLS t-test

As $N \to \infty$, the numerator of (9) divided by $\sqrt{N}$ converges to a normal distribution with mean zero and variance 1/2, that is

$$(1/2)(\chi^2_N - N)/\sqrt{N} \to N(0, 1/2)$$

The denominator can be expressed as an infinite weighted sum of independent chi-squared variates, that is

$$CvM_0(N) = \sum_{j=1}^{\infty} \pi^{-2}(j - 1/2)^{-2}\chi^2_N;$$

see, for example, Nyblom (1989). Hence $CvM_0(N)$ has mean $N/2$ and variance $N/3$. Thus

$$p\lim\ CvM_0(N)/N = 1/2$$

and so $\tau_0(N) \to N(0, 1)$.

Now suppose the covariance matrix has equal correlations, ω, as in (14). Without loss of generality denote let the first eigenvalue be $N\omega + 1$. The distribution of the numerator of (13), the t-statistic from the OLS estimator in a pooled regression, is now

$$\frac{N\omega + 1}{2}(\chi^2_1 - 1) + \frac{(1 - \omega)}{2}(\chi^2_{N-1} - N + 1)$$

for large T. Now consider what happens as $N \to \infty$. Dividing numerator and denominator by $\sqrt{N}$, we find that the distribution of the square of the denominator converges to $\omega CvM(1) + (1 - \omega)/2$. The second term in the numerator converges to zero when divided by N so

$$N^{-1/2}\tau^*_0 \to \frac{(\omega/2)(\chi^2_1 - 1)}{\{\omega CvM(1) + (1 - \omega)/2\}^{1/2}}.$$

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] Carvalho, V.M and A.C. Harvey., 2002. Growth, cycles and convergence in US regional time series. DAE Working paper 0221, University of Cambridge.

[8] Chang, Y. 2002. Nonlinear IV unit root tests in panels with cross-sectional dependency. Journal of Econometrics. 110, 261-92.

[9] 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.

[10] Doornik, J. A., 1999. Ox: An Object-Oriented Matrix Language. 3rd edition. Timberlake Consultants Press: London.

[11] 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.

[12] Elliot, G., Rothenberg, T.J. and J.H.Stock., 1996. Efficient tests for an autoregressive unit root. Econometrica, 64, 813-36.

[13] Evans, G.B.A. and N.E. Savin., 1981. Testing for unit roots I. Econometrica, 49, 753-779.

[14] Evans, P. and G. Karras., 1996. Convergence revisited. Journal of Monetary Economics 37, 249-265.

[15] Harvey A.C. and Carvalho, V.M., 2002. Models for converging economies. DAE Working paper 0216, University of Cambridge.

[16] Hendry, D.F., 2001. The impact of initial conditions on autoregressive estimation. Mimeo, Oxford.

[17] Hobijn, B. and P.H. Franses., 2000. Asymptotically perfect and relative convergence of productivity. Journal of Applied Econometrics 15, 59-81.

[18] Im, K.S., Pesaran, M H and Y Shin., 2002. Testing for unit roots in heterogeneous panels. Journal of Econometrics (forthcoming).

[19] Johansen, S., 1988. Statistical analysis of cointegration vectors. Journal of Economic Dynamics and Control 12, 13-54.

[20] Johansen, S., 1995. Likelihood-Based Inference in Co-Integrated Vector Autoregressive Models. Oxford. Oxford University Press.

[21] 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.

[22] 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.

[23] Lee, H.Y. and J.L. Wu., 2001. Mean reversion of inflation rates: Evidence from 13 OECD countries. Journal of Macroeconomics 23, 477-487.

[24] Levin, A., Lin, C-F. and C.-S. Lu., 2002. Unit root tests in panel data: asymptotic and finte sample properties. Journal of Econometrics 108, 1-24.

[25] Linden, M., 2000. Testing growth convergence with time series data - a nonparametric approach. International Review of Applied Econometrics 14, 361-70.

[26] 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.

[27] 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.

[28] Nyblom, J.(1989). Testing for the constancy of parameters over time. Journal of the American Statistical Association 84*:* 223-30.

[29] Nyblom, J. and A.C. Harvey., 2000. Tests of Common Stochastic Trends. Econometric Theory 16, 176-199.

[30] O'Connell, P., 1998. The overvaluation of purchasing power parity. Journal of International Economics 44, 1-19.

[31] 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.

[32] Phillips, P.C.B and D.Sul., 2002. Dynamic panel estimation and homogeneity testing under cross section dependence. Cowles Foundation discussion paper no. 1362.

[33] 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.

[34] 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.

[35] Schmidt, P. and J. Lee., 1991. A modification of the Schmidt-Phillips unit root test. Economics Letters 36, 285-89.

[36] 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.

[37] 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.

[38] Tanaka, K.,1996. Time Series Analysis. New York: John Wiley and Sons.

[39] 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.