Unit root tests allowing for breaks in panels with fixed T

Ioannis Karavias and Elias Tzavalis*

Department of Economics

Athens University of Economics & Business

Athens 104 34, Greece

May 2010

Abstract

In this paper we suggest panel data unit root tests which allow for a potential structural break in the individual effects of each series of the panel, assuming that the time-dimension of the panel (T) is fixed. Both cases of known and unknown time of a single break point are considered and extensions are provided allowing for serial correlation in the error term. In the case of unknown break point the test statistics converge to a theoretically known distribution and Monte Carlo evidence suggest that they have size which is very close to its nominal level and satisfactory power. The latter is clearly higher than that of unit root tests for single time series allowing for structural breaks.

JEL classification: C22, C23

Keywords: Panel data models; unit roots; structural breaks;

*The authors would like to thank the participants at the 3rd International Conference on Computational and Financial Econometrics (CFE09) , Cyprus 2009 and Arellano-Valle R.B and Genton M.G. for their useful comments in a previous version of this paper.

Ioannis Karavias: jkaravia@aueb.gr
Elias Tzavalis: etzavalis@aueb.gr

1 Introduction

The AR(1) model for short panels, whose time dimension T is fixed (or small) relatively to the cross-sectional one, denoted as N, has been extensively used in the literature in studying the dynamic behaviour of many economic series across different economic units [see Baltazi and Kao (2000), and Arelano and Honore (2002), Arelano (2003), *inter alia*]. Of particular interest is to use this model to examine if many economic series contain a unit root in their autoregressive component [see Hlouskova and Wagner (2006), for a recent survey]. More specifically, recent applications of unit roots tests to panel data include the examination of the following hypotheses: the economic growth convergence hypothesis [see de la Fuente (1997), for a survey], the hypothesis that stock prices and dividends follow the random walk model [see Lo and MacKinlay (1995) and Harris and Tzavalis (2004), *inter alia*], the purchasing power parity hypothesis [see Culver and Papell (1999), *inter alia*], and, finally, the effects of trade liberalization policies [see, e.g. Wacziarg and Welch (2004)].

In this paper we extend the panel data unit roots test statistics introduced by Harris and Tzavalis (1999), which assume that the time dimension of the panel is fixed and the cross-section dimension grows large, to allow for a potential structural break in the deterministic components (i.e. individual effects) of the AR(1) panel model of known and unknown date. This extension seems natural given recent evidence suggesting that many economic series may exhibit structural shifts in their deterministic components. As first pointed out by Perron (1989) for a single time series, such a type of shifts are expected to bias the unit root tests towards falsely accepting the null hypothesis of a unit root. For panel data sets, this has been shown by Carrion-i-Silvestre, Barrio-Castro and Lopez-Bazo (2001) based on an AR (1) panel data model which treats the break date as known.

In the literature, there are a few studies which consider panel data unit root tests allowing for structural breaks in the deterministic components of dynamic panel data models (see, e.g. Carrion-i-Silvestre, Barrio-Castro and Lopez-Bazo (2005) and Bai and Carrion-i-Silvestre (2009)). However, these studies assume that the time-dimension of the panel is large and grows larger than N. Thus, the tests that they suggest are more appropriate for panels with large number of time series observations. As shown in Harris and Tzavalis (1999) and Hadri (2000), application of this category of panel unit root tests to small-T panels will lead to serious size distortions and power decreases in testing the null hypothesis against its stationary alternative.

When the date (or time-point, as alternatively said) of the break is known, the test statistic that the paper suggests has a limiting distribution which is normal. The variance of this distribution will depend on the fraction of the sample that the break occurs, denoted as λ, and the time dimension of the panel, T. To implement our test statistic to the case that the time-point of the break is unknown, we suggest a sequential procedure in the line of Zivot and Andrews (1992). This procedure is based on the minimum value of the one-sided test statistic assuming a known date break. This test statistic is sequentially computed for each possible break point of the sample. The distribution of the minimum value of this is calculated by that of the minimum value of a fixed number of correlated normal variables. The paper derives analytically the correlation matrix of these variables and provides critical values of the distribution of their minimum value using the results from Arellano-Valle and Genton (2007) who derive the analytic form of the probability density function of the maximum of absolutely continuous dependent random variables. Afterwards we provide an extension of the test for the case of serial correlation. This is done for the cases that the deterministic components of the panel series contain individual effects.

The paper is organized as follows. The limiting distributions of the suggested test statistics of the paper are derived in Section 2 for known break points and in section 3 for unknown. In Section 3, we extend the test to allow for serial correlation. In section 4 we provide critical values for all situations described before. Also we conduct a Monte Carlo study with the aim of examining the small-N and T sample performance of the tests. Section 5 concludes the paper.

2 The test statistics and their limiting distribution

2.1 The date of the break point is known

Consider the following first-order autoregressive panel data model, AR(1):

$$y_i = \phi y_{i,-1} + X_i^{(\lambda)}\gamma_i^{(\lambda)} + u_i, \quad i = 1, 2, ..., N \tag{1}$$

where $y_i = (y_{i1}, ..., y_{iT})'$ is a $(TX1)$-dimension vector of the time series observations of the dependent variable of each cross-section unit of the panel i, $y_{i,-1} = (y_{i0}, ..., y_{iT-1})'$ is the vector y_i lagged one period back, $u_i = (u_{i1}, ..., u_{iT})$ is a $(TX1)$-dimension vector of disturbance terms and $X_i^{(\lambda)}$ is a $(TX2)$-dimension matrix of deterministic components defined bellow. Index λ denotes the date of the break point T_0 with $\lambda \in I = \{2, 3, ..., T-1\}$. The column-vectors of matrix $X_i^{(\lambda)}$ are appropriately designed to capture a common structural break in the vector $\gamma_i^{(\lambda)}$ of deterministic components for all i. These intercepts capture the individual effects of the panel.[1] More specifically, the column vectors of $X_i^{(\lambda)}$ are defined as follows: $e_t^{(\lambda)} = 1$ if $t \leq T_0$ and 0 otherwise, and $e_t^{(1-\lambda)} = 1$ if $t > T_0$ and 0 otherwise i.e $X_i^{(\lambda)} \equiv \left(e_t^{(\lambda)}, e_t^{(1-\lambda)}\right)$. The vector $\gamma_i^{(\lambda)}$ is then defined as $\gamma_i^{(\lambda)} = (a_i^{(\lambda)}(1-\varphi), a_i^{(1-\lambda)}(1-\varphi))'$.

The above specification of model (1) under the null and alternative hypotheses is appropriate if one takes the view that breakpoints are data dependent, which can be endogenouly determined by the data [see Zivot and Andrews (1992)].

Test statistics of the null hypothesis $\varphi = 1$ based on the above alternative specifications can be derived by noticing that the pooled LS estimator of φ, denoted $\widehat{\varphi}$, under the null hypothesis $\varphi = 1$ satisfies

$$\hat{\phi} - 1 = \left[\sum_{i=1}^{N} y_{i,-1}' Q^{(\lambda)} y_{i,-1}\right]^{-1} \left[\sum_{i=1}^{N} y_{i,-1}' Q^{(\lambda)} u_i\right], \tag{2}$$

where $Q^{(\lambda)} = \left[I - X_i^{(\lambda)} \left(X_i^{(\lambda)\prime} X_i^{(\lambda)}\right)^{-1} X_i^{(\lambda)\prime}\right]$ is the (TXT) "within" transformation matrix of the time series of the panel [Baltagi (1995), *inter alia*]. As shown in the Appendix, under the null hypothesis this estimator has the interesting property to be invariant to the initial conditions of the panel y_{i0}. However, it is an inconsistent estimator of ϕ due to the within transformation of the data (see, e.g. Nickel (1981), Kiviet (1995)). Thus, the test statistic that we will suggest for testing the null hypothesis $\phi = 1$ will rely on a correction of $\hat{\phi}$ for its inconsistency, along the line of Harris and Tzavalis (1999). The limiting distribution of $\hat{\phi} - 1$ corrected for the inconsistency of estimator $\hat{\phi}$ can be derived by making the following assumption about the sequence of the disturbance terms $\{u_{it}\}$.

Assumption 1: a1) $\{u_{it}\}$ is a sequence of independently and identically distributed (IID) random variables with $E(u_{it}) = 0$, $Var(u_{it}) = \sigma_u^2 < \infty$, $E(u_{it}^4) = k + 3\sigma_u^4$, $\forall\ i \in \{1, 2, ..., N\}$ and $\forall\ t \in \{1, 2, ..., T\}$, where $k < \infty$. a2) $E(a_i^m) = 0$, $V(a_i^m) = \sigma_{a^m}^2$, $E(a_i^m u_{it}) = 0$ for m$=\lambda, (1-\lambda)$ and $\forall\ i \in \{1, 2, ..., N\}$, $t \in \{1, 2, ..., T\}$. a3) The initial observation y_{i0} satisfies $E(y_{i0}) = 0$, $V(y_{i0}) = \sigma_0^2$, $E(y_{i0}u_{it}) = 0$, $E(a_i^m y_{i0}) = 0$ for m$=\lambda, (1-\lambda)$and $\forall\ i \in \{1, 2, ..., N\}$, $t \in \{1, 2, ..., T\}$.

The next theorem presents the limiting distribution of $\hat{\phi} - 1$ corrected for the inconsistency of $\hat{\phi}$ under the null hypothesis $\phi = 1$.

[1] As is argued below, this non-linear specification of the individual effects (intercepts) of model (1) nests the random walk model $y_i = y_{i,-1} + u_i$, for all i, and the stationary model in the same framework. This framework is more appropriate for evaluating the power of the test statistics suggested by the paper under the stationary alternative hypothesis.

Theorem 1 Let the sequence $\{y_{it}\}$ be generated according to model (1) and the break-point T_0 be known.

Then, under the null hypothesis $\phi = 1$ and Assumption 1, we have:

$$Z(\lambda, T) \equiv C(k, \sigma_u^2, \lambda, T)^{-1/2}\sqrt{N}(\hat{\phi} - 1 - B(\lambda, T)) \xrightarrow{L} N(0,1) \tag{3}$$

as $N \rightarrow \infty$, where

$$B(\lambda, T) = \underset{N\rightarrow\infty}{p\lim}(\hat{\phi} - 1) = tr[\Lambda' Q^{(\lambda)}]\{tr(\Lambda' Q^{(\lambda)}\Lambda)\}^{-1},$$

and

$$C(k, \sigma_u^2, \lambda, T) = \left\{ k\sum_{j=1}^{T} a_{jj}^{(\lambda)^2} + 2\sigma_u^4 tr(A^{(\lambda)^2}) \right\} \left\{ \sigma_u^2 tr(\Lambda' Q^{(\lambda)}\Lambda) \right\}^{-2},$$

where Λ is a (TXT) matrix defined as $\Lambda_{r,c} = 1$, if $r > c$ and 0 otherwise, $A^{(\lambda)} \equiv [a_{ij}]$ is a (TXT) dimension symmetric matrix, defined as $A^{(\lambda)} = \frac{1}{2}(\Lambda' Q^{(\lambda)} + Q^{(\lambda)}\Lambda) - B(\lambda, T)(\Lambda' Q^{(\lambda)}\Lambda)$ and '$\xrightarrow{L}$' signifies convergence in distribution.

The proof of the theorem is given in Tzavalis (2003). Its results indicate that the inconsistency of the pooled LS estimator $\hat{\phi}$, given by $B(\lambda, T)$, is a deterministic function of the fraction of the time series observations of the sample that the break occurs λ and the time dimension of the panel T. Below, we make some interesting remarks on the results of the theorem that provide extensions of them which may be proved very useful in practice.

Remark 1 *It can be easily shown that the results of Theorem 1 can be extended to the case that the disturbance terms u_{it} are heterogenous across i and thus, we have $IID(0, \sigma_{u_i}^2)$ and $E(u_{it}^4) = k_i + 3\sigma_{u_i}^4$, where $k_i < \infty \ \forall \ i \in \{1, 2, ..., N\}$. In this case, the nuisance parameters σ_u^2 and k will be given as $\sigma_u^2 = \frac{1}{N}\sum_{i=1}^{N}\sigma_{u_i}^2$ and $k = \frac{1}{N}\sum_{i=1}^{N} k_i$, respectively (see White (2000)). These parameters can be estimated based on consistent estimates of the second and fourth moments of the first difference of the panel data series y_{it} under the null hypothesis $\phi = 1$ (see Harris and Tzavalis (2004)).*

Remark 2 *When $u_{it} \sim NIID(0, \sigma_u^2)$, then $k = 0$ and thus, the variance of limiting distribution of $\sqrt{N}(\hat{\phi} - 1 - B(\lambda, T))$ becomes $\left\{2tr(A^{(\lambda)^2})\right\}\left\{tr(\Lambda' Q^{(\lambda)}\Lambda)\right\}^{-2}$. This variance function means that the limiting distribution of $\sqrt{N}(\hat{\phi} - 1 - B(\lambda, T))$ and the standardized test statistic $Z(\lambda, T)$ will no longer depend on the nuisances parameters k and σ_u^2.*

2.2 The date of the break point is unknown

The results of Theorem 1 are based on the assumption that the break point is known. In this subsection, we relax this assumption. We propose a test statistic of the null hypothesis $\phi = 1$ which allows for a common structural break in the deterministic components of model (1) of an unknown date. As in Perron and Vogelsang (1992), and Zivot and Andrews (1992), we will view the selection of the break point as the outcome of minimizing the standardized test statistic $Z(\lambda, T)$ given by Theorem 1 over all possible break points of the sample, after trimming the initial and final parts of the time series of the panel model (1).[2] The minimum value of the test statistic $Z(\lambda, T)$, for all $\lambda \in (0, 1)$, will give the least favorable result of the null hypothesis $\phi = 1$.

Let $\hat{\lambda}_{\min}$ denote the break point at which the minimum value of $Z(\lambda, T)$, over all $\lambda \in (0, 1)$, is obtained. Then, the null hypothesis will be rejected if we have:

$$\min_{\lambda \in (0,1)} Z(\lambda, T) < c_{\min},$$

where $c_{\min}$ denotes the size a left-tail critical value of the limiting distribution of $\min\limits_{\lambda \in (0,1)} Z(\lambda, T)$. The following theorem enables us to tabulate the critical values of this distribution at any significance (size) level a.

Theorem 2 Let Assumption 1 hold. Then, as $N \to \infty$, we have

$$\min_{\lambda \in (0,1)} Z(\lambda, T) \xrightarrow{d} \min_{\lambda \in (0,1)} N(0, \Sigma), \tag{4}$$

where $\Sigma \equiv [\sigma_{\lambda s}]$ is the covariance matrix of the test statistics $Z(\lambda, T)$, for all possible pairs of break fractions (λ, s).This matrix has elements given as follows:

$$\sigma_{\lambda s} = \frac{k \sum_{j=1}^{T} a_{jj}^{(\lambda)} a_{jj}^{(s)} + 2\sigma_u^4 tr(A^{(\lambda)} A^{(s)})}{\sqrt{k \sum_{j=1}^{T} a_{jj}^{(\lambda)} a_{jj}^{(\lambda)} + 2\sigma_u^4 tr(A^{(\lambda)} A^{(\lambda)})} \sqrt{k \sum_{j=1}^{T} a_{jj}^{(s)} a_{jj}^{(s)} + 2\sigma_u^4 tr(A^{(s)} A^{(s)})}}$$

for all different pairs (λ, s).

The result of this theorem follows immediately from Theorem 1 using the continuous mapping theorem. The functional form of the covariance elements $\sigma_{\lambda s}$ of the standardized test statistic $Z(\lambda, T)$, for all pairs of

[2] For trimming the initial and final parts of the series of the panel, note that $\lambda \equiv \frac{T_0}{T}$ will range from $\frac{1}{T}$ to $\frac{T-1}{T}$ for the panel data model (1) without the individual trends, while it will range from $\frac{2}{T}$ to $\frac{T-2}{T}$ for model (1) with the trends.

(λ, s), can be derived based on the following result: $E(\xi_i^{(\lambda)}\xi_i^{(s)}) = k\sum_{j=1}^{T} a_{jj}^{(\lambda)}a_{jj}^{(s)} + 2\sigma_u^4 tr(A^{(\lambda)}A^{(s)})$, where $\xi_i^{(j)} = u_i' A^{(j)} u_i$ is defined in the Appendix (see proof of Theorem 1). Below, we present some special cases of Theorem 2 as remarks. These may be proved very useful in practice.

Remark 3 *When $u_{it} \sim NIID(0, \sigma_u^2)$, then $k = 0$ and $\sigma_{\lambda s}$ are given as*

$$\sigma_{\lambda s} = \frac{tr(A^{(\lambda)}A^{(s)})}{\sqrt{tr(A^{(\lambda)}A^{(\lambda)})tr(A^{(s)}A^{(s)})}}$$

The result of Theorem 2 imply that the critical values of the limiting distribution of the standardized test statistic $\min\limits_{\lambda\in(0,1)} Z(\lambda, T)$, $c_{\min}$, can be obtained from the distribution of the minimum value of a fixed number of correlated normal variables with covariance matrix Σ, whose elements are given by the theorem.

The probability density function of the distribution of the maximum of absolutely continuous dependent random variables is given by Arellano-Valle and Genton (2007). Generally, in the case of elliptically contoured distributions denoted by D, if X is a random vector of n elliptically contoured distributions let $F_n(x; \mu, \Sigma, h^{(n)})$ be its cdf where h$^{(n)}$ is the density generator. For a fixed i consider the partition given by

$$X = \begin{pmatrix} X_{-i} \\ X_i \end{pmatrix}, \qquad \mu = \begin{pmatrix} \mu_{-i} \\ \mu_i \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \Sigma_{-i-i} & \Sigma_{-ii} \\ \Sigma_{i-i} & \Sigma_{ii} \end{pmatrix}$$

then $X_i \sim D_1(\mu_i, \Sigma_{ii}, h^{(1)})$ with density $f_1(x; \mu_i, \Sigma_{ii}, h^{(1)}) = h^{(1)}(z_i^2)/\sqrt{\Sigma_{ii}}$, where $z_i = (x - \mu_i)/\sqrt{\Sigma_{ii}}$. After this marginilization the rest of the vector conditional on X$_i$ follows

$(X_{-i}/X_i = x) \sim D_{n-1}(\mu_{-i,i}(x), \Sigma_{-i-i,i}, h_{z_i^2}^{(n-1)})$ where $\mu_{-i,i}(x) = \mu_{-i} + (x - \mu_i)\Sigma_{-ii}/\Sigma_{ii}$, $\Sigma_{-i-i,i} = \Sigma_{-i-i} - \Sigma_{-ii}\Sigma_{-ii}'/\Sigma_{ii}$ and $h_{z_i^2}^{(n-1)} = h^{(n)}(u + z_i^2)/h^{(1)}(z_i^2)$, $u \geq 0$. Then the pdf of the maximum is given by

$$f_{X_{(n)}}(x) = \sum_{i=1}^{n} f_1(x; \mu_i, \Sigma_{ii}, h^{(1)}) F_{n-1}(x1_{n-1}; \mu_{-i,i}(x), \Sigma_{-i-i,i}, h_{z_i^2}^{(n-1)}), \quad x \in R \tag{5}$$

The needed critical values can be easily obtained by the fact that $\min\{X_1, ..., X_k\} = -\max\{-X_1, ..., -X_k\}$. Assume a significance level α, then

$$P(\min_{\lambda\in(0,1)} Z(\lambda, T) < c_{\min}) = a \quad \rightarrow \quad P(\max_{\lambda\in(0,1)} - Z(\lambda, T) > -c_{\min}) = a$$

The last expression is an integral equation with respect to $c_{\min}$ and is solved numerically.

Remark If X is a random vector from a multivariate normal distribution, X$\sim N(\mu, \Sigma)$ then the pdf is given by

$$f_{X_{(n)}}(x) = \sum_{i=1}^{n} \varphi_1(x; \mu_i, \Sigma_{ii}) \Phi_{n-1}(x1_{n-1}; \mu_{-i,i}(x), \Sigma_{-i-i,i}), \quad x \in R$$

where φ and Φ are the pdf and cdf of a normal distribution.

Theorem 3 Under assumption 1 the test is consistent:

$$\lim_{n \to +\infty} P(\min_{\lambda \in (0,1)} Z(\lambda, T) < c_{\min} \mid H_a) = 1$$

The proof is given in the appendix.

3 Extension of the tests for serial correlation

We generalize the model by allowing for serial correlation in the error term. Theoritically, we treat the cases of known and unknown break point as before but now the limiting distribution of the test statistic is different. To derive this distribution, we make the following assumption for the sequence of disturbance terms, $\{u_{it}\}$.

Assumption 2:(b1) $\{u_i\}$ constitutes a sequence of independent random $T \times 1$ vectors with means $E(u_i) = 0$ and $(T \times T)$ autocovariance matrices $E(u_i u_i') = \Gamma_i \equiv [\gamma_{i,rs}]$ of unknown form apart from $\gamma_{i,1T} = \gamma_{i,T1} = 0$. (b2) The smallest eigenvalue of the average population covariance matrix $\bar{\Gamma}_N \equiv \frac{1}{N} \sum_{i=1}^{N} \Gamma_i$ is bounded away from zero for sufficiently large N. (b3)$E(u_{it} y_{io}) = E(u_{it} a_i^m) = 0$ for m=$\lambda, (1-\lambda)$and $\forall$ $i \in \{1, 2, ..., N\}$, $t \in \{1, 2, ..., T\}$. (b4) $E(u_{it}^4) < +\infty$, $E(y_{i0}^4) < +\infty$, $E((a_i^m)^4) < +\infty$, $E(y_{i0}^2 \gamma_i^m \gamma_i^{m\prime}) < +\infty$.

Assumption 2 allows for quite general conditions for the data generating process and enables us to derive the limiting distribution of $\hat{\phi} - 1$ under the null hypothesis by employing standard results from asymptotic theory under the assumption that N tends to infinity, while T remains fixed. Below, we discuss some of the implications of the assumption in more detail.

Condition (b1) allows for heterogeneity in the cross section dimension of the panel as well as variation in the values of the autocovariance parameters over time. Ignoring intertemporal differences among the elements of the Γ_i may lead to erroneous inference about ϕ. The condition $\gamma_{i,rs} = \gamma_{i,sr} = 0$ implies that the maximum order of serial correlation in the u_{it}'s should be less than $T-1$. It can encounter for a parametric, moving average (MA) model of serial correlation in u_{it} of lag order up to $q = T-2$.[3] This assumption is required because the time dimension of the panel is fixed and our test statistics correct the LS estimator for the inconsistency due to the serial correlation by using a non-parametric estimator of the average population covariance matrix, $\bar{\Gamma}_N$, based on $\frac{1}{N}\sum_{i=1}^{N}(\Delta y_i \Delta y_i')$. It enables us to correct for the inconstency of the LS estimator coming from the non-zero elements of $\bar{\Gamma}_N$, while preserving sample variation in the corrected estimator arising from the zero elements of $\bar{\Gamma}_N$ which will allow us to test for the null hypothesis $\phi = 1$.

Under the null we do not need to make any assumption about the initial observations of the panel, y_{i0} or the individual effects a_i^m, for both models considered because the test statistics that we propose are asymptotically invariant under the null hypothesis, $\phi = 1$. Under the alternative y_{i0} and a_i^m are not eliminated but we only need the weak assumptions b3) and b4) to prove the consistency of the test.

To derive a unit root test statistic for model (1) based on $\hat{\phi}_1$, define the following quantities:

$$\hat{b}_1 = vec(Q^{(\lambda)}\Lambda)' S \left(\frac{1}{N}\sum_{i=1}^{N} vec(\Delta y_i \Delta y_i') \right),$$

where Λ is a $(T \times T)$ matrix, defined in the previous section, S is a $(T^2 \times T^2)$ diagonal selection matrix defined as $S_{(c-1)T+r,(c-1)T+r} = 1 - d((\overline{\Gamma}_N)_{r,c} = 0)$, $\{r = 1, 2, ..., T,\ c = 1, 2, ..., T\}$, where $d(\cdot)$ denotes the Dirac function defined as 0 everywhere and 1 for x=0. The symbol $vec(.)$ denotes the vector operator, and

$$\hat{\delta}_1 = \frac{1}{N}\sum_{i=1}^{N} y_{i,-1}' Q^{(\lambda)} y_{i,-1}.$$

Since S selects the non-zero elements of $\Gamma_i = E(u_i u_i')$ (and, hence, $\bar{\Gamma}_N$), it can be easily seen that substracting the ratio $\frac{\hat{b}_1}{\hat{\delta}_1}$ from $(\hat{\phi}_1 - 1)$ adjusts the estimator $\hat{\phi}_1$ for its inconsistency under the null hypothesis, and thus can lead to a test statistic whose mean is net of the nuisance parameters of the data while its variance depends on that of the sample moments of the zero elements of $\bar{\Gamma}_N$. This is given in in the next theorem.

[3]Note that MA models of serial correlation are often observed in many economic series [see Swchert (1989)]. The order of serial correlation considered in u_{it} may also be adequate to capture autoregressive (AR) models of serial correlation or ARMA models whose order of serial correlation dies out after the lag $T-2$.

Theorem 4 Let the sequence $\{y_{i,t}\}$ be generated according to model (1) and assumption 2 holds. Then under the null hypothesis $\phi = 1$, as $N \to \infty$

$$Z_1 \equiv V_1^{-0.5} \hat{\delta}_1 \sqrt{N} \left(\hat{\phi}_1 - 1 - \frac{\hat{b}_1}{\hat{\delta}_1} \right) \xrightarrow{d} N(0,1), \tag{6}$$

where

$$V_1 = vec(Q^{(\lambda)}\Lambda)'(I_{T^2} - S) \left(\frac{1}{N} \sum_{i=1}^{N} \mathcal{V}(vec(\Delta y_i \Delta y_i')) \right) (I_{T^2} - S) vec(Q^{(\lambda)}\Lambda),$$

where I_{T^2} is a $(T^2 \times T^2)$ identity matrix, $\mathcal{V}(\cdot)$ denotes the variance operator and " $\xrightarrow{d}$ " signifies convergence in distribution. The proof of the theorem is given in the appendix.

Given a consistent estimator for V_1, the statistic proposed in Theorem 1 can be readily used in practice by employing the tables of the standard normal distribution to test for the null hypothesis of $\phi = 1$. A consistent estimator for V_1 can be obtained by replacing $\frac{1}{N}\sum_{i=1}^{N} \mathcal{V}(vec(\Delta y_i \Delta y_i'))$ by its consistent estimator

$$\begin{aligned} \hat{\Omega} &= \frac{1}{N} \sum_{i=1}^{N} (vec(\Delta y_i \Delta y_i') vec(\Delta y_i \Delta y_i')') \\ &\quad - \left(\frac{1}{N} \sum_{i=1}^{N} vec(\Delta y_i \Delta y_i') \right) \left(\frac{1}{N} \sum_{i=1}^{N} vec(\Delta y_i \Delta y_i') \right)', \end{aligned}$$

where $\hat{\Omega}$ is a positive semidefinite matrix.

To test for unit roots, the statistic given by Theorem 4 exploits the following moment condition of the data $E(\Delta y_{it} \Delta y_{it}') = E(u_i u_i') = 0$ which holds under the null hypothesis. This can be easily seen by writing the terms in the cross-section summation of the numerator of $\left(\hat{\phi}_1 - 1 - \frac{\hat{b}_1}{\hat{\delta}_1} \right)$ as

$$\begin{aligned} W_{i,T}^{(\lambda)} &= vec(Q^{(\lambda)}\Lambda)' (I_{T^2} - S) \, vec(\Delta y_i \Delta y_i') \\ &= vec(Q^{(\lambda)}\Lambda)' (I_{T^2} - S) \, vec(u_i u_i'), \end{aligned}$$

since $\Delta y_i = u_i$ under the null hypothesis, and noticing that the matrix $(I_{T^2} - S)$ selects the zero elements of the average covariance matrix $\bar{\Gamma}_N$, i.e $E(u_{i1}u_{iT})$.[4] If it is assumed that the order of serial correlation in the disturbance terms u_i is smaller than $q = T-2$, these moments conditions will increase. In this case our test statistic will exploit a linear combination of these moments by appropriately designing the selection matrix S, so that $I_T - S$ to also pick up the additional elements of the sample moment of $\Delta y_i \Delta y_i'$ which correspond to the new zero elements of matrix $\bar{\Gamma}_N$, for $q < T-2$.

For the case of the unknown break point we provide a theorem analogous to Theorem 2:

Theorem 5 Let Assumption 2 hold. Then, as $N \to \infty$, we have

$$\min_{\lambda \in (0,1)} Z_1 \xrightarrow{d} \min_{\lambda \in (0,1)} N(0, \Sigma), \tag{7}$$

where $\Sigma \equiv [\sigma_{\lambda s}]$ is the covariance matrix of the test statistics Z_1, for all possible pairs of break fractions (λ, s).This matrix has elements given as follows:

$$\begin{aligned} \sigma_{\lambda s} \quad = \quad & vec(Q^{(\lambda)}\Lambda)'(I_{T^2} - S)\left(\frac{1}{N}\sum_{i=1}^{N}\mathcal{V}(vec(\Delta y_i \Delta y_i'))\right)(I_{T^2} - S)vec(Q^{(s)}\Lambda) \\ & \frac{1}{\sqrt{vec(Q^{(\lambda)}\Lambda)'(I_{T^2} - S)\left(\frac{1}{N}\sum_{i=1}^{N}\mathcal{V}(vec(\Delta y_i \Delta y_i'))\right)(I_{T^2} - S)vec(Q^{(\lambda)}\Lambda)}} \\ & \frac{1}{\sqrt{vec(Q^{(s)}\Lambda)'(I_{T^2} - S)\left(\frac{1}{N}\sum_{i=1}^{N}\mathcal{V}(vec(\Delta y_i \Delta y_i'))\right)(I_{T^2} - S)vec(Q^{(s)}\Lambda)}} \end{aligned}$$

for all different pairs (λ, s).

The main difference between theorems 2 and 5 is that now we do not know the exact quantity $\mathcal{V}(vec(\Delta y_i \Delta y_i')$, thus we must estimate it by $\hat{\Omega}$.

Theorem 6 Under assumption 2 the test is consistent:

$$\lim_{n \to +\infty} P(\min_{\lambda \in (0,1)} Z_1 < c_{\min} \mid H_a) = 1$$

The proof is given in the appendix.

[4] The authors would like to thank an anonymous referee for making this point to us.

4 Critical Values and Simulation Results

4.1 Critical Values

Below we have tabulated the critical values for the theoritical distributions for T=10,15,25 assuming the errors $u_{it} \sim NIID(0, \sigma_u^2)$ for the case of no serial correlation and MA(1) errors for $\theta = \{-0.5, 0, 0.5\}$ for the case of correlation. The integral equation that gives the critical values is numerically solved with presicion up to the second decimal. The problem in the second case is that we do not know theoritically $\frac{1}{N}\sum_{i=1}^{N} \mathcal{V}(vec(\Delta y_i \Delta y_i'))$ so in a first step we estimate it by $\hat{\Omega}$ for N=10000 and then plug it in to get the theoritical distribution.

Table 1 contains the critical values of the theoritical distribution presented in section 2. Table 2 contains the critical values for the case of a moving average model of order 1. θ does affect these values because after the variance is estimated covariances are taken with respect to the mean zero elements of the matrix. Numerical accuracy is 0.02.

Table 1: Critical values of $\min\limits_{\lambda \in (0,1)} N(0, R)$

a/T	10	15	25
1%	-2.87	-2.93	-2.99
5%	-2.25	-2.31	-2.37
10%	-1.93	-1.99	-2.03

Table 2: Critical values of $\min\limits_{\lambda \in (0,1)} N(0, \Sigma)$, $MA(1)$

a/T	10	15	25
1%	-2.95	-3.03	-3.13
5%	-2.37	-2.47	-2.55
10%	-2.07	-2.17	-2.27

4.2 Monte Carlo Simulations

In this section we conduct Monte Carlo experiments for the test statistics of theorems 2 and 5, that is for the case of unknown break point with and without serial correlation and compare them.The data generating process is $y_{it} = \phi y_{it-1} + (1-\varphi)a_i^{(\lambda)} + \varepsilon_{it} + \theta\varepsilon_{it-1}$ before the break and $y_{it} = \phi y_{it-1} + (1-\varphi)a_i^{(1-\lambda)} + \varepsilon_{it} + \theta\varepsilon_{it-1}$ after. All random variables $a_i^{(1-\lambda)}, a_i^{(\lambda)}, \varepsilon_{it}, \varepsilon_{it-1}, y_{i0}$ have a standard normal distribution and each result is taken after 10000 repetitions. Below we provide the size and power for all four cases, no correlation and correlation for $\theta = \{-0.5, 0, 0.5\}$ and for various combinations of N and T all for nominal size 5%.

Table 5: Size and power of nominal level 5% for the test of theorem 2

N	25	25	50	50	100	100	100
T	10	15	15	25	10	15	25
$\phi = 1.00$	0.07	0.08	0.08	0.07	0.05	0.07	0.07
				$\lambda = 0.25$			
$\phi = 0.99$	0.10	0.14	0.15	0.22	0.12	0.18	0.33
$\phi = 0.95$	0.26	0.42	0.61	0.87	0.56	0.83	0.99
$\phi = 0.90$	0.48	0.74	0.93	0.99	0.92	0.99	1
				$\lambda = 0.50$			
$\phi = 0.99$	0.10	0.14	0.15	0.22	0.12	0.18	0.33
$\phi = 0.95$	0.27	0.43	0.60	0.87	0.56	0.84	0.98
$\phi = 0.90$	0.48	0.74	0.93	0.99	0.92	0.99	1
				$\lambda = 0.75$			
$\phi = 0.99$	0.10	0.14	0.15	0.22	0.12	0.19	0.29
$\phi = 0.95$	0.27	0.42	0.61	0.87	0.57	0.83	0.98
$\phi = 0.90$	0.49	0.73	0.93	0.99	0.92	0.99	1

Table 6: Size and power of nominal level 5% for the test of theorem 5, $\theta = -0.5$

N	25	25	50	50	100	100	100
T	10	15	15	25	10	15	25
$\phi = 1.00$	0.07	0.07	0.06	0.07	0.05	0.05	0.05
				$\lambda = 0.25$			
$\phi = 0.99$	0.08	0.07	0.06	0.08	0.06	0.05	0.05
$\phi = 0.95$	0.09	0.08	0.07	0.08	0.08	0.07	0.07
$\phi = 0.90$	0.09	0.09	0.07	0.08	0.08	0.09	0.09
				$\lambda = 0.50$			
$\phi = 0.99$	0.08	0.07	0.06	0.08	0.06	0.06	0.05
$\phi = 0.95$	0.09	0.08	0.07	0.08	0.08	0.07	0.06
$\phi = 0.90$	0.09	0.08	0.07	0.08	0.09	0.08	0.09
				$\lambda = 0.75$			
$\phi = 0.99$	0.08	0.08	0.06	0.08	0.06	0.05	0.05
$\phi = 0.95$	0.09	0.08	0.07	0.08	0.08	0.08	0.07
$\phi = 0.90$	0.10	0.08	0.08	0.09	0.08	0.08	0.09

Table 7: Size and power of nominal level 5% for the test of theorem 5, $\theta = 0$

N	25	25	50	50	100	100	100
T	10	15	15	25	10	15	25
$\phi = 1.00$	0.07	0.09	0.07	0.07	0.06	0.07	0.05
				$\lambda = 0.25$			
$\phi = 0.99$	0.08	0.10	0.09	0.09	0.08	0.11	0.09
$\phi = 0.95$	0.13	0.14	0.17	0.13	0.20	0.20	0.16
$\phi = 0.90$	0.17	0.17	0.21	0.17	0.32	0.29	0.22
				$\lambda = 0.50$			
$\phi = 0.99$	0.09	0.11	0.08	0.09	0.08	0.10	0.08
$\phi = 0.95$	0.14	0.17	0.17	0.13	0.20	0.17	0.16
$\phi = 0.90$	0.17	0.20	0.20	0.17	0.34	0.30	0.23
				$\lambda = 0.75$			
$\phi = 0.99$	0.09	0.10	0.09	0.09	0.08	0.10	0.09
$\phi = 0.95$	0.14	0.15	0.15	0.12	0.19	0.21	0.17
$\phi = 0.90$	0.19	0.18	0.20	0.16	0.35	0.29	0.22

Table 8: Size and power of nominal level 5% for the test of theorem 5, $\theta = 0.5$

N	25	25	50	50	100	100	100
T	10	15	15	25	10	15	25
$\phi = 1.00$	0.06	0.08	0.06	0.07	0.05	0.06	0.06
				$\lambda = 0.25$			
$\phi = 0.99$	0.09	0.11	0.09	0.10	0.09	0.10	0.09
$\phi = 0.95$	0.17	0.16	0.18	0.13	0.30	0.24	0.15
$\phi = 0.90$	0.24	0.23	0.28	0.19	0.55	0.40	0.24
				$\lambda = 0.50$			
$\phi = 0.99$	0.09	0.10	0.09	0.10	0.09	0.10	0.09
$\phi = 0.95$	0.17	0.17	0.18	0.14	0.31	0.24	0.15
$\phi = 0.90$	0.25	0.22	0.28	0.19	0.57	0.41	0.24
				$\lambda = 0.75$			
$\phi = 0.99$	0.09	0.11	0.10	0.10	0.08	0.09	0.10
$\phi = 0.95$	0.17	0.18	0.18	0.13	0.31	0.25	0.15
$\phi = 0.90$	0.24	0.24	0.27	0.19	0.57	0.40	0.24

The results of the tables clearly indicate that both test statistics have size close to the nominal level 5% for all combinations of N and T. The power in the test statistic of theorem 2 increases repidly with both N and T. When $\theta = 0$ the use of the test statistic of theorem 5 depicts the implications of the variance matrix estimation. Estimating $\frac{1}{N}\sum_{i=1}^{N}\mathcal{V}(vec(\Delta y_i \Delta y_i'))$ in the second statistic brings noise that heavily affects the power of the test. For $\theta = 0.5$, the power function increases although in a slower rate than the previous case and when $\theta = -0.5$ the rate is even slower. Still the size is not distorted and the test is unbiased unlike Agiakoglou Newbold (1996) and Schwert (1989) . In the case of correlation the effect of an increase in T on the power of the tests per se is ambiguous since it does not only increase the number of observations, but also the number of nuisance parameters of the variance matrix that is estimated. The time of the break doesn't significally affect the results.

5 Empirical Application

We now apply the statistic of theorem 5 to examine whether there is high persistence in the consumption of the Eurogroup countries or the introduction of euro led to a structural break in the series. Final consumption expenditure of households is broken down by consumption purposes in twelve categories and annual data are collected for a time span of eleven years, from 1996 to 2006 and for fifteen countries of the eurogroup (Greece is not included due to missing data). All variables were divided by the respective country's gdp to eliminate the trend. In order to mitigate for possible effects of cross section correlation on the tests, the individual series of our panel data set were taken in deviations from their cross-section mean at each point in time, following O'Connel (1998). The critical value used for this test is taken from table 1 for T=10 and we assume a moving average model of order 1.

The results of the table clearly indicate that the null hypothesis of a unit root in the level of final consumpion variable is rejected in favour of its stationary alternative. The break point seems to occur in year 2003. This can be due to uncertainty and consumer reservation in the first years of the new currency.

Year	1998	1999	2000	2001	2002	2003	2004	2005
Statistic	3.2152	0.8988	2.1654	0.9685	0.1090	-0.2375	0.7072	1.0030

6 Conclusions

In this paper we proposed panel data unit root test statistics that allow for a structural break in the individual effects of panel data sets, assuming that the time dimension of the panel is fixed. These tests allow for a break point at either a known or unknown date. When the break point is considered as known, we show that the test statistic suggested has normal limiting distribution whose variance depend on the fraction of the sample that the break occurs. On the other hand, when the break is considered as unknown the proposed test relies on a sequential testing procedure of the null hypothesis. This entails in computing the values of the test statistic for a known structural break over all possible break points of the sample, in first step, and then selecting that with the minimum value to test the null hypothesis. The minimum value of this sequential test statistic has a limiting distribution whose critical values can be tabulated as those of the minimum value of a fixed number of correlated normal variables, after trimming for the initial and final time-points of the sample. The distribution of the minimum of correlated random variables is not known, but the distribution of the maximum is, and through this we derive the exact critical values. We extend the model to allow for serial correlation. In this case, the covariance between the test statistics for different break dates is not theoritically known but is consistenly estimated. Finally we prove consistency of the tests for both statistics.

To evaluate the small sample performance of the proposed test statistics, we conduct a small Monte Carlo study assuming the break point unknown. This shows that both types of the test statistics suggested, i.e. with and without correlation in the error term, have empirical size which is very close to its nominal level of 5% and power which increases with both dimensions of the panel. As an empirical illustration, we employed the statistic of theorem 2 to examine whether there is high persistence in the consumption of the Eurogroup countries or the introduction of euro led to a structural break in the series. We find that there is evidence of a break in the series in 2003 which is different from 2002, the date euro came into circulation.

A Appendix

Proof of Theorem 1: To derive the limiting distribution of the test statistic of the theorem, we will proceed into stages. We first show that the pooled LS estimator, $\hat{\phi}$, is inconsistent, as $N \to \infty$. We then construct a normalized statistic based on estimator $\hat{\phi}$ corrected for its inconsistency and we derive its limiting distribution under the null hypothesis of $\hat{\phi}$, as $N \to \infty$.

Decompose the vector $y_{i,-1}$ under the null hypothesis $\phi = 1$ as

$$y_{i,-1} = ey_{i0} + \Lambda' u_i, \tag{8}$$

where e is a $(TX1)$-dimension vector of ones and matrix Λ is defined in the theorem. Premultiplying equation (8) with the matrix $Q^{(\lambda)}$ yields

$$Q^{(\lambda)} y_{i,-1} = Q^{(\lambda)} \Lambda' u_i, \tag{9}$$

since $Q^{(\lambda)} e = (0, 0, ..., 0)$.

Substituting (9) into (2) and noticing that $Q^{(\lambda)}$ is an idempotent and symmetric matrix yields

$$\hat{\phi} - 1 = \left[\sum_{i=1}^{N} u_i' \Lambda' Q^{(\lambda)} u_i \right] \left[\sum_{i=1}^{N} u_i' \Lambda' Q^{(\lambda)} \Lambda u_i \right]^{-1}. \tag{10}$$

Taking probability limits of equation (10) gives the inconsistency of $\hat{\phi}$ as follows:

$$
\begin{aligned}
B(\lambda, T) &= p \lim_{N\to\infty} (\hat{\phi} - 1) = E\left[u_i'\Lambda' Q^{(\lambda)} u_i\right] E\left[u_i'\Lambda' Q^{(\lambda)} \Lambda u_i\right]^{-1} \\
&= tr\left[\Lambda' Q^{(\lambda)}\right] \left\{tr\left[\Lambda' Q^{(\lambda)}\Lambda\right]\right\}^{-1}, \qquad (11)
\end{aligned}
$$

by the LLN.

Subtracting the term $B(\lambda, T)$ from (10) gives the corrected for its inconsistency estimator of ϕ:

$$
\begin{aligned}
&\hat{\phi} - 1 - B(\lambda, T) \\
= &\left\{\sum_{i=1}^{N}\left[u_i'\Lambda' Q^{(\lambda)} u_i - B(\lambda, T)(u_i'\Lambda' Q^{(\lambda)}\Lambda u_i)\right]\right\}\left\{\sum_{i=1}^{N} u_i'\Lambda' Q^{(\lambda)}\Lambda u\right\}^{-1} \\
= &\left\{\sum_{i=1}^{N}\xi_i^{(\lambda)}\right\}\left\{\sum_{i=1}^{N} u_i'\Lambda' Q^{(\lambda)}\Lambda u\right\}^{-1}, \qquad (12)
\end{aligned}
$$

where $\xi_i^{(\lambda)} = u_i'\Lambda' Q^{(\lambda)} u_i - B(\lambda, T)(u_i'\Lambda' Q^{(\lambda)}\Lambda u_i)$ is a random variable which has zero mean by construction and constant variance for all i, denoted $Var(\xi_i^{(\lambda)})$. Using standard results on quadratic forms, $\xi_i^{(\lambda)}$ can be written as follows:

$$
\begin{aligned}
\xi_i^{(\lambda)} &= u_i'\frac{1}{2}\left(\Lambda' Q^{(\lambda)} + Q^{(\lambda)}\Lambda\right) u_i - B(\lambda, T)(u_i'\Lambda' Q^{(\lambda)}\Lambda u_i) \\
&= u_i'\left[\frac{1}{2}\left(\Lambda' Q^{(\lambda)} + Q^{(\lambda)}\Lambda\right) - B(\lambda, T)(\Lambda' Q^{(\lambda)}\Lambda)\right] u_i \\
&= u_i' A^{(\lambda)} u_i, \qquad (13)
\end{aligned}
$$

where $A^{(\lambda)} = \frac{1}{2}\left(\Lambda' Q^{(\lambda)} + Q^{(\lambda)}\Lambda\right) - B(\lambda, T)(\Lambda' Q^{(\lambda)}\Lambda)$ is a symmetric matrix, given that $\frac{1}{2}\left(\Lambda' Q^{(\lambda)} + Q^{(\lambda)}\Lambda\right)$ and $(\Lambda' Q^{(\lambda)}\Lambda)$ are symmetric matrices. Using results on quadratic forms for symmetric matrices, it can be seen that $Var(\xi_i^{(\lambda)})$ is given by

$$
\begin{aligned}
Var(\xi_i^{(\lambda)}) &= Var[u_i' A^{(\lambda)} u_i] \\
&= k \sum_{j=1}^{T} a_{jj}^{(\lambda)^2} + 2\sigma_u^4 tr\left(A^{(\lambda)^2}\right) \qquad (14)
\end{aligned}
$$

[see Anderson (1971)].

The result of the theorem can be proved by scaling (16) appropriately and using the following two asymptotic results, as $N \to \infty$:

$$
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi_i^{(\lambda)} \xrightarrow{d} N(0, Var(\xi_i)) \qquad (15)
$$

by the CLT, and

$$
p\lim \frac{1}{N} \sum_{i=1}^{N} u_i' \Lambda' Q^{(\lambda)} \Lambda u_i = \sigma_u^2 tr\left[\Lambda' Q^{(\lambda)} \Lambda\right] \qquad (16)
$$

by the LLN. These results hold under the conditions of Assumption 1. Note that the condition $k < \infty$ of the assumption guarantees that $Var(\xi_i^{(\lambda)})$ exists.

Proof of Theorem 2: $E(\xi_i^{(\lambda)} \xi_i^{(s)}) = E(u_i' A^{(\lambda)} u_i u_i' A^{(s)} u_i) = k \sum_{j=1}^{T} a_{jj}^{(\lambda)^2} + 2\sigma_u^4 tr(A^{(\lambda)^2})$.From combining this with (16) and standardizing it with the respectives $C(k, \sigma_u^2, \lambda, T)$ we take the result.

Proof of Theorem 3: The estimator can be written as

$$
\hat{\phi} = \varphi + \left[\sum_{i=1}^{N} y_{i,-1}' Q^{(\lambda)} y_{i,-1}\right]^{-1} \left[\sum_{i=1}^{N} y_{i,-1}' Q^{(\lambda)} u_i\right] \qquad (17)
$$

and

$$
y_{i-1} = w y_{i0} + \Omega X_i^{(\lambda)} \gamma_i^{(\lambda)} + \Omega u_i \qquad (18)
$$

where $w = (1, \varphi, \varphi^2, ..., \varphi^{T-1})'$ and Ω is given by

$$\Omega = \begin{pmatrix} 0 & . & . & . & . & . & 0 \\ 1 & 0 & & & & & . \\ \varphi & 1 & . & & & & . \\ \varphi^2 & \varphi & . & . & & & . \\ . & & . & . & . & & . \\ . & & & . & 1 & 0 & . \\ \varphi^{T-2} & \varphi^{T-3} & . & . & \varphi & 1 & 0 \end{pmatrix}$$

Notice that for $\varphi = 1$, $\Omega = \Lambda$. Under H_a and the assumption 1 combining (17) and (18), by application of the Markov LLN:

$$\frac{\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1} Q^{(\lambda)} u_i}{\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1} Q^{(\lambda)} y_{i,-1}} \longrightarrow \frac{\sigma_u^2 tr(\Omega' Q^{(\lambda)})}{\sigma_u^2 tr(\Omega' Q^{(\lambda)} \Omega) + tr(X_i^{(\lambda)\prime} \Omega' Q^{(\lambda)} \Omega X_i^{(\lambda)} \Sigma_\gamma) + \sigma_0^2 w' Q^{(\lambda)} w}$$

where Σ_γ is the variance-covariance matrix of $\gamma_i^{(\lambda)}$. Therefore

$$\begin{aligned} C(k, \sigma_u^2, \lambda, T)^{-1/2}\sqrt{N}(\hat{\phi} - 1 - B(\lambda, T) &= \sqrt{N}(\varphi - 1) C(k, \sigma_u^2, \lambda, T)^{-1/2} + \\ &+ C(k, \sigma_u^2, \lambda, T)^{-1/2}\sqrt{N}\left(\frac{\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1} Q^{(\lambda)} u_i}{\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1} Q^{(\lambda)} y_{i,-1}} - B(\lambda, T)\right) \end{aligned}$$

since $\varphi < 1$, $\sqrt{N}(\varphi - 1) C(k, \sigma_u^2, \lambda, T)^{-1/2} \longrightarrow -\infty$ and the second term converges to a normal distribution with finete mean and variance.

Proof of Theorem 4: The ideas are the same with the proof of theorem 1. To derive the limiting distribution of the test statistic, we first show that the LSDV estimator, $\hat{\phi}_{LSDV}$, is inconsistent. We then construct a normalised statistic based on $\hat{\phi}_{LSDV}$ corrected for the inconsistency.

Decompose the vector $y_{i,-1}$ for model (1) under the null hypothesis $\phi = 1$ as

$$y_{i,-1} = e y_{i0} + \Lambda u_i, \tag{19}$$

where the matrix Λ is defined in Theorem 1.

Premultiplying (19) with the matrix M yields

$$Q^{(\lambda)}y_{i,-1} = Q^{(\lambda)}\Lambda u_i, \tag{20}$$

since $Me = (0, 0, ..., 0)'$. Under the null hypothesis $\phi = 1$, we also have $u_i = \Delta y_i$. It follows that

$$y'_{i,-1}Q^{(\lambda)}u_i = u'_i\Lambda' Q^{(\lambda)}u_i = \Delta y'_i\Lambda' Q^{(\lambda)}\Delta y_i. \tag{21}$$

It can now easily be seen that the numerator of $\hat{\phi}-1$ converges in probability by the Markov LLN as $N \to \infty$, i.e.

$$\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1}Q^{(\lambda)}u_i - tr(\Lambda' Q^{(\lambda)}\overline{\Gamma}_N) \xrightarrow{p} 0, \tag{22}$$

where $\overline{\Gamma}_N = \frac{1}{N}\sum_{i=1}^{N}\Gamma_i$. Since S picks up all elements of $vec(u_iu'_i)$ that have a mean which is different from zero, $\frac{1}{N}\sum_{i=1}^{N} Svec(\Delta y_i \Delta y'_i)$ is a consistent estimator of $vec(\overline{\Gamma}_N)$ under the null hypothesis.

Subtract the term $\widehat{b}_1/\hat{\delta}_1$ from $\hat{\phi} - 1$. This yields

$$\hat{\phi} - 1 - \frac{\hat{b}_1}{\hat{\delta}_1} = \left[\sum_{i=1}^{N} D_{i,T}^{(\lambda)}\right]^{-1}\left[\sum_{i=1}^{N} W_{i,T}^{(\lambda)}\right] \tag{23}$$

where

$$\begin{aligned} W_{i,T}^{(\lambda)} &= y'_{i,-1}Q^{(\lambda)}u_i - vec(Q^{(\lambda)}\Lambda)'S\,(vec(\Delta y_i\Delta y'_i)) \\ &= \Delta y'_i\Lambda' Q^{(\lambda)}\Delta y_i - vec(Q^{(\lambda)}\Lambda)'S\,(vec(\Delta y_i\Delta y'_i)) \\ &= vec(Q^{(\lambda)}\Lambda)'[vec(\Delta y_i\Delta y'_i) - Svec(\Delta y_i\Delta y'_i)] \\ &= vec(Q^{(\lambda)}\Lambda)'\,(I_{T^2} - S)\,vec(\Delta y_i\Delta y'_i) \end{aligned} \tag{24}$$

is a typical term in the numerator of the inconsistency corrected estimator $\hat{\phi}$, and

$$\begin{aligned} D_{i,T}^{(\lambda)} &= y'_{i,-1}Q^{(\lambda)}y_{i,-1} \\ &= u'_i\Lambda' Q^{(\lambda)}\Lambda u_i \\ &= \Delta y'_i\Lambda' Q^{(\lambda)}\Lambda\Delta y_i \end{aligned} \tag{25}$$

is a typical term in the denominator of $\hat{\phi}$. The equalities in (25) also follow from (9), and the equality $u_i = \Delta y_i$, which is only valid under the null hypothesis. The terms $W_{i,T}^{(\lambda)}$ have expectation zero by construction, since the matrix $(I_{T^2} - S)$ selects those elements of $vec(u_i u_i')$ that have mean zero. The variance of $W_{i,T}^{(\lambda)}$, $V_{i,T}^{(\lambda)}$, is given by $vec(Q^{(\lambda)}\Lambda)'\,(I_{T^2} - S)\,\mathcal{V}(vec(\Delta y_i \Delta y_i'))\,(I_{T^2} - S)\,vec(Q^{(\lambda)}\Lambda)$. These variances are different across i. Notice that the variance of $W_{i,T}^{(\lambda)}$ is well defined, when $\mathcal{V}(vec(\Delta y_i \Delta y_i'))$ exists. The latter follows by condition (b1) of Assumption 1, which requires the existence of the $4+\delta$-th population moments of Δy_i.

Application of the Markov LLN yields that $\hat{\delta}_1 - tr(\Lambda' Q^{(\lambda)} \Lambda \overline{\Gamma}_N) = N^{-1}\sum_{i=1}^N D_{i,T}^{(\lambda)} - tr(\Lambda' Q^{(\lambda)} \Lambda \overline{\Gamma}_N) \xrightarrow{p} 0$ (under the null hypothesis) by condition (b1) of Assumption 1. Condition (a2) ensures that $\text{plim}_{N\to\infty}\hat{\delta}_1$ is different from zero. If conditions (b1) and (b2) of Assumption 1 are satisfied, then $N^{-0.5}\sum_{i=1}^N V_1^{-0.5} W_{i,T}^{(\lambda)} \xrightarrow{d} N(0,1)$ by the Lindeberg-Feller CLT, where $V_1^{(\lambda)} = N^{-1}\sum_{i=1}^N V_{i,T}^{(\lambda)}$ is the cross-sectional average of the variances of the $W_{i,T}^{(\lambda)}$, $i = 1, 2, ..., N$. Finally, by applying the Cramér theorem, we obtain

$$\sqrt{N}(\frac{\hat{\delta}_1}{V_1^{0.5}})(\hat{\phi} - 1 - \frac{\hat{B}_1}{\hat{\delta}_1}) \xrightarrow{d} N\,(0, 1)\,. \tag{26}$$

Proof of Theorem 5: The same with the proof of theorem 2 by taking $W_{i,T}^{(\lambda)}$ this time.

Proof of Theorem 6: Under assumption 2 combining equations (17) and (18)

$$\frac{\frac{1}{N}\sum_{i=1}^N y_{i,-1}' Q^{(\lambda)} u_i}{\frac{1}{N}\sum_{i=1}^N y_{i,-1}' Q^{(\lambda)} y_{i,-1}} \longrightarrow \frac{tr(\Omega' Q^{(\lambda)} \overline{\Gamma})}{tr(\Omega' Q^{(\lambda)} \Omega \overline{\Gamma}) + tr(X_i^{(\lambda)\prime} \Omega' Q^{(\lambda)} \Omega X_i^{(\lambda)} \Sigma_\gamma) + \sigma_0^2 w' Q^{(\lambda)} w}$$

$\overline{\Gamma}$ is consistenly estimated by $\frac{1}{N}\sum(\Delta y_i \Delta y_i')$ because under the null $\Delta y_i = u_i$. This is not the case under the alternative so we must examine the limit behavior of the estimator. Under the alternative

$$u_i = y_i - \phi y_{i-1} - X_i^{(\lambda)} \gamma_i^{(\lambda)}$$

and by subtracting from both sides y_{i-1} and by rearranging

$$\Delta y_i = u_i + (\varphi - 1) y_{i-1} + X_i^{(\lambda)} \gamma_i^{(\lambda)}$$

After some algebra it can be shown that

$$\begin{aligned}\frac{1}{N}\sum(\Delta y_i \Delta y_i') \quad &\longrightarrow \quad \overline{\Gamma} + (\varphi-1)\overline{\Gamma}\Omega' + (\varphi-1)\Omega\overline{\Gamma}' + (\varphi-1)^2(\sigma_0^2 ww' + \Omega\overline{\Gamma}\Omega' + \Omega X_i^{(\lambda)}\Sigma_\gamma X_i^{(\lambda)\prime}\Omega') + \\ &+(\varphi-1)\Omega X_i^{(\lambda)}\Sigma_\gamma X_i^{(\lambda)\prime} + (\varphi-1)X_i^{(\lambda)}\Sigma_\gamma X_i^{(\lambda)\prime}\Omega' + X_i^{(\lambda)}\Sigma_\gamma X_i^{(\lambda)\prime}\end{aligned}$$

which is finite. Notice that for $\varphi = 1$ all terms after the first one disappear. Using the same argument as in the proof of theorem 3 we get the result.

References

[1] Arelano E. (2003), Panel Data Econometrics, Oxford University Press.

[2] Arelano E. and B. Honoré (2002), Panel data models: Some recent developments, Handbook of Econometrics, edited by Heckman J. and E. Leamer editors, Vol 5, North Holland.

[3] Bai J. and J.L. Carrion-i-Silvestre (2009), "Structural changes, common stochastic trends and unit roots in panel data", *Review of Economic Studies*, 76, 471-501.

[4] Baltagi B.H. (1995), Econometric analysis of panel data, Wiley, Chichester.

[5] Baltagi B.H. (2000), Nonstationary panels, cointegration in panels and dynamic panels: A survey in advances in econometrics: nonstationary panels, panel cointegration and dynamic panels, edited by Baltagi B.H, T.B. Fomby and R. C. Hill, 15, 7-51.

[6] Barro R.J. and X. Sala-i-Martin (1995), Economic growth, McCraw-Hill, New York.

[7] Carrion-i-Silvestre J.L., Del Barrio-Castro, and E. Lopez-Bazo (2001), "Level shifts in a panel data based unit root test. An application to the rate of unemployment", mimeo, Department of Econometria, University of Barcelona, Spain.

[8] Carrion-i-Silvestre J.L., Del Barrio-Castro, and E. Lopez-Bazo (2005), "Breaking the panels. An application to real per capita GDP", *Econometrics Journal*, 8, 159-175.

[9] Culver S.E. and D.H. Papell (1999), Long-run power parity with short-run data: evidence with a null hypothesis of stationarity, *Journal of International Money and Finance*, 18, 751-768.

[10] Granger C.W.J. and P. Newbold (1986), Forecasting economic time series, Academic Press, London.

[11] Faini R. (2004), "Trade liberalization in a Globalizing word", IZA, Discussion Paper No. 1406.

[12] Hadri K. (2000), "Testing for stationarity in heterogeneous panel data", Econometrics Journal, 3, 148-161.

[13] Hahn J. and G. Kuersteiner (2002), "Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large", *Econometrica*, 70, 1639-1657.

[14] Harris R.D.F. and E. Tzavalis (1999), "Inference for unit roots in dynamic panels where the time dimension is fixed", *Journal of Econometrics*, 91, 201-226.

[15] Harris R.D.F and E.Tzavalis (2004), "Inference for unit roots for dynamic panels in the presence of deterministic trends: Do stock prices and dividends follow a random walk ?",*Econometric Reviews* (2004), 23, 149-166.

[16] Hlouskova J. and M. Wagner (2006), "The performance of panel unit root and stationary tests: Results from a large scale simulation study",*Econometric Reviews*, 25, 85-117.

[17] Kiviet Y. (1995), "On bias , inconsistency, and efficiency of various estimators in dynamic panel data models", *Journal of Econometrics*, 74, 119-147.

[18] Levin A., and C.F., Lin, and C-S. J, Chu (2002), "Unit root tests in panel data: asymptotic and finite-sample properties", *Journal of Econometrics*, 108, 1-24.

[19] Lo A.W. and A.C. MacKinlay (1995), A non-random walk down Wall Street, Princeton University Press.

[20] O'Connell P.G.J (1998), The overvaluation of purchasing power parity, *Journal of International Economics*, 44, 1-19.

[21] Perron P. (1989), "The great crash, the oil price shock, and the unit root hypothesis", *Econometrica*, 57, 1361-1401.

[22] Perron P. (1990), "Testing for a unit root in a time series with a changing mean", *Journal of Business & Economic Statistics*, 8, 153-162.

[23] Perron P. (2006), "Dealing with structural breaks", in Palgrave Handbook of Econometrics, Vol. 1: Econometric Theory, Mills. T and K. Patterson (eds), Palgrave MacMillan, 278-352.

[24] Perron P. and T., Vogelsang (1992), "Nonstationarity and level shifts with an application to purchasing power parity", *Journal of Business & Economic Statistics, 10, 301-320.*

[25] Stock J.H. (1994), "Unit roots, structural breaks and trends", in Handbook of Econometrics, Vol. IV, Engle R.F. and D.L. McFadden (eds), Elsevier, Amsterdam.

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

[27] Wacziarg R. and K.H. Welch (2004), "Trade liberalization and growth: new evidence", NBER Working Paper 10152.

[28] White H. (2000), Asymptotic theory for econometricians, Academic Press.