Generalized fixed-T Panel Unit Root Tests Allowing for Structural Breaks

Yiannis Karavias and Elias Tzavalis*

Department of Economics

Athens University of Economics & Business

Athens 104 34, Greece

June 2012

Working Paper

Abstract

In this paper we suggest panel data unit root tests which allow for serial correlation in the disturbance terms and structural breaks in the individual effects or linear trends of panel data models. The limiting distributions of the tests are derived under the assumption that the time-dimension of the panel (T) is fixed, while the cross-section (N) grows large. Thus, they are appropriate for short panels, where T is small. The tests consider the cases of a known and unknown date break. For the latter case, the distribution of the tests is nonstandard. The paper gives an analytic form of this distribution, based on the maximum of T-elliptically contoured distributions, which faciliates the application of the tests. The paper proves the consistency of the tests and derives their local power functions. It is shown that the tests have local power on a root-N neighbourhood of the null hypothesis, even for the case that consider individual linear trends. Monte Carlo evidence suggest that our tests have size which is very close to its nominal level and satisfactory power in small-T panels. This is true even for cases where the degree of serial correlation is large and negative, where single time series unit root tests are found to be oversized.

JEL classification: C22, C23

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

*The authors would like to thank the participants at the Conference in Honour of Professor Hashem Pesaran in Cambridge 2011 and Arellano-Valle and Genton for their useful comments on a previous version of this paper.

Yiannis Karavias: jkaravia@aueb.gr

Elias Tzavalis: etzavalis@aueb.gr

1 Introduction

A vast amount of work has been recently focused on drawing inference about unit roots based on dynamic panel data models (see, Hlouskova and Wagner (2006), for a more recent survey). Since many empirical panel data studies rely on short panels, of particular interest is testing for a unit root in dynamic panel data model when the time dimension of the panel, denoted as T, is fixed (finite) and its cross-section, denoted as N, grows large (see, e.g. Harris and Tzavalis (1999, 2004), Hadri (2000), Binder et al (2005) and De Wachter et al (2007)). These tests have better small-T sample performance, compared to large-T panel unit root tests (see, e.g., Levin et al (2002), Im et al (2003)), given that they assume finite T. Implementations of fixed-T panel unit root tests include many interesting applications in economics such as testing for the economic convergence hypothesis [see de la Fuente (1997), for a survey], the purchasing power parity hypothesis under different economic (or exchange rate) regimes [see Culver and Papell (1999), *inter alia*] and the effects of liberalization policies on trade (see, e.g., Warziarg and Welch (2004)).

In this paper, we extend the fixed-T panel data unit roots test statistics of Harris and Tzavalis (1999) to allow for a common structural break in the deterministic components of panel data models, namely their individual effects or linear trends of a known and unknown date. This is done in a dynamic panel data framework allowing for serial correlation of the disturbance terms. This extension is very useful given plethora of evidence supporting the view that the presence of unit roots in economic time series can be falsely attributed to the existence of structural breaks in their deterministic components (see, e.g., Perron (2006), for a survey). On this front, the panel data approach offers an interesting and unique perspective that it is not shared by single univariate tests. The cross-sectional dimension of the panel can provide useful sample information which can help to distinguish the type of shifts (breaks) in the deterministic components of the panel from the effects of stochastic permanent shocks. As pointed out by Bai (2010), this framework can more accurately trace out structural break points of the panel data. Allowing for serial correlation in the unit root tests is critical due to its inherent nature in economic time series data, as is stated in Schwert (1989), Said and Dickey (1984) and Phillips (1987). The paper assumes that the maximum order of serial correlation of the disturbance terms of the panel data model is a function of the time dimension T. Both the variance and the serial correlation effects in the disturbance terms of the panel data models are allowed to be heterogenous across the individual units of the panel. This assumption makes the tests applicable under quite general panel data generating processes, observed in reality.

There are a few studies in the literature which suggest fixed-T panel data unit root tests allowing for a common structural break in the deterministic components of the panel data model (see, e.g., Carrion-i-Silvestre et al (2002) and Tzavalis (2002), or more recently, Karavias and Tzavalis (2012a), and Hadri (2012)). These studies however suggest unit root tests using the simple AR(1) panel data model as an auxiliary regression model, which may not be operational in practice due to the assumption of no serial correlation made for the disturbance terms. As recently is noted by De Blander and Dhaene (2011), this may lead to erroneous inference about unit roots if there is substantial serial correlation in the disturbance terms. The main goal of the above studies is to pass ideas how to test for unit roots in the presence of

breaks of known or unknown date, using the AR(1) panel data model as an example. In addition to the above, there are also studies in the literature which suggest panel unit root tests allowing for a common structural break, but they assume that the time-dimension of the panel, T, is large and grows faster than its cross-section, denoted as N (see, e.g., Carrion-i-Silvestre et al (2005), Bai and Carrion-i-Silvestre (2009), and Kim (2011)). These are appropriate for large-T panel data sets. Application of this category of panel unit root tests to small-T panels will lead to serious size distortions and critical power reductions in testing the null hypothesis of unit roots against its stationary alternative (see Harris and Tzavalis (1999)). As shown recently by Karavias and Tzavalis (2012a), the existence of a break in the data generating process requires panel data sets with a quite large time-dimension, T (e.g. $T > 100$), so as the large-T panel unit root tests to have satisfactory small sample properties.

The sequential panel data test statistics suggested by the paper for the case of an unknown break are in line to those suggested by Andrews (1993), Zivot and Andrews (1992), Perron (1997), Perron and Vogelsang (1998), *inter alia*), for single time series. In this paper the limiting distribution of these test statistics can be obtained as the minimum value of a finite number of correlated variates; $T-2$ for the panel data model with individual effects and $T-3$ for this model allowing also for individual linear trends. This distribution is shown analytically based on recent results of Arellano-Valle and Genton (2008), who have derived the analytic form of the probability density function of the maximum of absolutely continuous dependent random variables. The analytic form of this distribution enables us to derive critical values of the suggested test statistics without having to rely on Monte Carlo analysis, as in Karavias and Tzavalis (2012a). Thus, it substantially facilitates application of the tests in practice.

In addition to the limiting distribution of the test statistics, the paper provides a number of other useful results which have theoretical and practical interest. First, it proves the consistency of the tests and provides their local power. The paper shows that, when the break date is known, there are circumstances where the tests have non trivial asymptotic power in a neighborhood of unity which shrinks at $\sqrt{N}$ rate. This happens even if individual (incidental) linear trends are considered. Second, it shows how to apply the tests for the case that the break point is in the individual effects of the panel under the null hypothesis, as is assumed by some studies (see, e.g., Kim (2011)). Application of our test in this case requires a consistent estimator of the break-point under the null hypothesis. This estimator is provided by Bai (2010). Finally, based on a Monte Carlo study, the paper shows that the tests work satisfactory for small-T or N panel data sets.

The paper is organized as follows. In Section 2, we derive the limiting distributions of the panel unit root test statistics allowing for a known, or an unknown, date break under the assumption that the disturbance terms of panel data models are white noise processes. This analysis helps us to better understand the testing principles of the test statistics. In Section 3, we generalize the results of Section 2 to allow for serial correlation in the disturbance terms, while in Section 4 we extend the tests to allow also for individual linear trends. In Section 4, we show how to carry out the tests when there is a break in the individual effects of panel data models under the null hypothesis. Section 5 conducts a Monte Carlo simulation study to examine the small sample performance of the tests. Section 6 concludes the paper. All the mathematical derivations are provided in the Appendix of the paper.

2 Test statistics and their limiting distribution

In this section, we present panel unit root test statistics for the cases that the break date is known and unknown. This is done under the assumption that the disturbance terms of the AR(1) panel data model considered are independently, identically normally distributed ($NIID$). Extensions of the tests to the more general case of serially correlated and heterogenous disturbance terms are made in the next section.

2.1 Known date break

Consider the following AR(1) nonlinear dynamic panel data model:

$$y_{it} = a_{it}^{(\lambda)}(1-\varphi) + \varphi y_{it-1} + u_{it}, \quad i = 1, 2, .., N, \tag{1}$$

where $\varphi \in (-1, 1]$, $a_{it}^{(\lambda)} = a_i^{(1)}$ if $t \leq T_0$ and $a_i^{(2)}$ if $t > T_0$, where T_0 denotes the time-point of the sample, referred to as break-point, where a common break in the individual effects of panel data model (1) α_i occurs, for all cross-section units of the panel i. $a_i^{(1)}$and $a_i^{(2)}$ denote the individual effects of model (1) before and after the break point T_0, respectively. Throughout the paper, we will denote the fraction of the sample that this break occurs as λ, i.e. $\lambda = \frac{T_0}{T} \in I = \left\{\frac{2}{T}, \frac{3}{T}, ..., \frac{T-1}{T}\right\}$.

Under the null hypothesis of a unit root (i.e. $\varphi = 1$), model (1) reduces to the pure random walk model $y_{it} = y_{it-1} + u_{it}$, for all i, while, under the alternative of stationarity (i.e. $\varphi < 1$), it considers a common structural break in individual effects a_i. The above specification of the null and the alternative hypotheses is very common in single time series inference procedures allowing for structural breaks (see, e.g., Zivot and Andrews (1992), Andrews (1993), Perron and Vogelsang (1998). The main focus of these procedures is to diagnose whether evidence of unit roots can be spuriously attributed to the ignorance of structural breaks in nuisance parameters of the data generating processes like individual effects a_i. The common break assumption across all units of the panel i can be attributed to a monetary regime shift, which is common across all economic units, or to a structural economic shock which is independent of the disturbance terms u_{it}, like a credit crunch or an exchange rate realignment. As aptly noted by Bai (2010), even if each series of the panel data model has its own break point, the common break assumption across i is useful in practice not only for its computational simplicity, but also because it allows for estimating the mean of possibly random break points. This mean will have an economic policy interest.

The AR(1) panel data model (1) can be employed to carry out unit root tests allowing for a structural break in individual effects $a_{it}^{(\lambda)}$ based on the within groups least squares (LS) estimator of autoregressive coefficient of φ, denoted as $\hat{\varphi}^{(\lambda)}$. This estimator is also known as least square dummy variable (LSDV) estimator (see, e.g., Baltagi (1995), *inter alia*). Under the null hypothesis of $\varphi = 1$, it implies:

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

where $y_i = (y_{i1}, ..., y_{iT})'$ is a $(TX1)$-dimension vector collecting the time series observations of dependent variable y_{it} of each cross-section unit of the panel i, $y_{i,-1} = (y_{i0}, ..., y_{iT-1})'$ is vector y_i lagged one period

back, $u_i = (u_{i1}, ..., u_{iT})$ is a $(TX1)$-dimension vector of disturbance terms u_{it} and $Q^{(\lambda)}$ is the (TXT) "within" transformation matrix of the individual series of the panel data model, y_{it}. Let us define $X^{(\lambda)} \equiv \left(e^{(1)}, e^{(2)}\right)$ to be a matrix of deterministic components used by the LSDV estimator to demean the levels of series y_{it}, for all i, where $e^{(1)}$ and $e^{(2)}$ are $(TX1)$-column vectors whose elements are defined as follows: $e_t^{(1)} = 1$ if $t \leq T_0$ and 0 otherwise, and $e_t^{(2)} = 1$ if $t > T_0$ and 0 otherwise. Then, matrix $Q^{(\lambda)}$ will be defined as $Q^{(\lambda)} = I_T - X^{(\lambda)}(X^{(\lambda)\prime}X^{(\lambda)})^{-1}X^{(\lambda)\prime}$, where I_T is an identity matrix of dimension (TXT).

Panel data unit root testing procedures based on above LSDV estimator $\hat{\varphi}^{(\lambda)}$ have the very interesting property that, under the null hypothesis, are invariant (similar) to the initial conditions of the panel y_{i0} and, after appropriate specification of matrix $X^{(\lambda)}$, to the individual effects of the panel data model, as will be seen in Section 4. The latter happens if matrix $X^{(\lambda)}$ also contains broken linear trends. Similarity of the tests with respect the initial conditions of the panel does not require any mean or covariance stationarity conditions on the panel data processes y_{it}, as assumed by generalized method of moments, or conditional and unconditional maximum likelihood based panel data unit root inference procedures (see, e.g., Hsiao et al (2002) and Madsen (2008)). These conditions may be restrictive in practice**.** However, $\hat{\varphi}^{(\lambda)}$ is an inconsistent (asymptotic biased) estimator of φ due to the within transformation of the data, which wipes off individual effects $a_{it}^{(\lambda)}$ (or initial conditions y_{i0} under the null hypothesis $\varphi = 1$) while allowing for a structural break under the alternative hypothesis of stationarity, $\varphi < 1$. Thus, the test statistics that we suggest testing for hypothesis $\varphi = 1$ will rely on a correction of estimator $\hat{\varphi}^{(\lambda)}$ for its inconsistency (asymptotic bias) due to the above transformation of the data (see, e.g., Harris and Tzavalis (1999, 2004)). To derive the limiting distribution of these tests, we make the following assumption about the sequence of the disturbance terms $\{u_{it}\}$.

Assumption 1 (a1) $\{u_i\}$ constitutes a sequence of independent identically distributed (IID) $(TX1)$-dimension vectors with means $E(u_i) = 0$ and variance-autocovariance matrices $\Gamma_i \equiv E(u_i u_i') = \sigma_u^2 I_T < +\infty$ and nonzero, for all i. (a2) $E(u_{it}y_{io}) = E\left(u_{it}a_{it}^{(1)}\right) = E\left(u_{it}a_{it}^{(2)}\right) = 0$ and $\forall\ i \in \{1, 2, ..., N\}$, $t \in \{1, 2, ..., T\}$. (a3) $E\left(u_{it}^4\right) < +\infty$, $E(y_{i0}^4) < +\infty$, $E\left((a_{it}^{(1)})^4\right) < +\infty$, $E\left((a_{it}^{(2)})^4\right) < +\infty$ and $E\left(y_{i0}^2\left(a_{it}^{(1)}\right)^2\right) < +\infty$, $E\left(y_{i0}^2\left(a_{it}^{(2)}\right)^2\right) < +\infty$.

Condition (a1) of Assumption 1 enables us to derive the limiting distribution of a panel data unit root test statistic based on estimator $\hat{\varphi}^{(\lambda)}$ under the null hypothesis $\varphi = 1$ by applying standard asymptotic theory for IID processes, while (a2) and (a3) are simple regularity conditions under which the test statistic can be proved that is consistent under the alternative hypothesis, $\varphi < 1$. The following theorem provides the limiting distribution of such a test statistic, based on estimator $\hat{\varphi}^{(\lambda)}$ corrected for its bias. For analytic convenience, this is done under the assumption that u_{it} is also normally distributed, i.e. $u_{it} \sim NIID(0, \sigma_u^2)$, for all i and t.

Theorem 1 *Let $u_{it} \sim NIID(0, \sigma_u^2)$, then, under the null hypothesis $\varphi = 1$ and known λ, we have*

$$Z^{(\lambda)} \equiv \widehat{V}^{(\lambda)-1/2}\hat{\delta}^{(\lambda)}\sqrt{N}\left(\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{\delta}^{(\lambda)}} - 1\right) \xrightarrow{d} N(0, 1) \tag{3}$$

as $N \to \infty$, *where*

$$\frac{\hat{b}^{(\lambda)}}{\hat{\delta}^{(\lambda)}} \equiv \frac{\hat{\sigma}_u^2 tr(\Lambda' Q^{(\lambda)})}{\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1} Q^{(\lambda)} y_{i,-1}} \tag{4}$$

is a consistent estimate of the asymptotic bias of $\hat{\varphi}^{(\lambda)}$ *which, under the null hypothesis, is given as* $\frac{b^{(\lambda)}}{\delta^{(\lambda)}} = \frac{\sigma_u^2 tr(\Lambda' Q^{(\lambda)})}{\sigma_u^2 tr(\Lambda' Q^{(\lambda)}\Lambda)}$, $\hat{\sigma}_u^2$ *is a consistent estimator of variance* σ_u^2 *under the null hypothesis, which is given as* $\hat{\sigma}_u^2 = \frac{\sum_{i=1}^{N} \Delta y'_i \Psi^{(\lambda)} \Delta y_i}{N tr(\Psi^{(\lambda)})}$ *where* Δ *is the difference operator and* $\Psi^{(\lambda)}$ *is a* (TXT)*-dimension matrix having in its main diagonal the corresponding elements of matrix* $\Lambda' Q^{(\lambda)}$*, and zeros elsewhere, and* $V^{(\lambda)}$ *is a variance function given as*

$$V^{(\lambda)} = \sigma_u^4 F^{(\lambda)\prime}(K_{T^2} + I_{T^2}) F^{(\lambda)}, \tag{5}$$

where $F^{(\lambda)} = vec(Q^{(\lambda)}\Lambda - \Psi^{(\lambda)\prime})$, K_{T^2} *is a* $(T^2 X T^2)$*-dimension commutation matrix and* I_{T^2} *is a* $(T^2 X T^2)$*-dimension identity matrix. The proof is given in the appendix.*

The test statistic $Z(\lambda)$, given by Theorem 1, can be easily implemented to test unit root hypothesis $\varphi = 1$ based on the tables of the standard normal distribution. Theorem 1 shows that the asymptotic bias of estimator $\hat{\varphi}^{(\lambda)}$ stems from the "within" transformation matrix $Q^{(\lambda)}$, which induces correlation between vectors $y_{i,-1}$ and u_i (see, e.g. Nickel (1981)). Since disturbance terms u_{it} are IID, the correlation between $y_{i,-1}$ and u_i comes only from the main diagonal elements of the variance-autocovariance matrices of u_{it}, defined by Assumption 1 as $\Gamma_i \equiv E(u_i u'_i) = \sigma_u^2 I_T$, for all i. The above bias can be estimated by the nonparametric estimator $\frac{\hat{b}^{(\lambda)}}{\hat{\delta}^{(\lambda)}}$ and, thus, it can be subtracted from $\hat{\varphi}^{(\lambda)} - 1$ to obtain a test statistic which is normally distributed and is asymptotically net of nuisance parameter effects. To test the null hypothesis $\varphi = 1$, this statistic is based on the off-diagonal elements of the sample moments of variance-autocovariance matrices Γ_i which are equal to zero, i.e. $E(u_{it}u_{is}) = 0$ for $s \neq t$. This can be better seen by writing test statistic $Z^{(\lambda)}$ as

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} u'_i (\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)}) u_i = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} tr\left[(\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)}) u_i u'_i \right], \tag{6}$$

(see Appendix) where $(\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)})$ is matrix with zeros in its main diagonal due to the subtraction of matrix $\Psi^{(\lambda)}$ from $\Lambda' Q^{(\lambda)}$, implying $tr\left[(\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)}) E(u_i u'_i)\right] = 0$, for all i.[1]Matrix $\Psi^{(\lambda)}$ allows us to capture the correlation effects between vectors $y_{i,-1}$ and u_i, which are induced by the "within" transformation of the data through matrix $Q^{(\lambda)}$ and, thus, generate the bias of LSDV estimator $\hat{\varphi}^{(\lambda)}$. Subtracting $\Psi^{(\lambda)}$ from $\Lambda' Q^{(\lambda)}$ enables us to adjust $\hat{\varphi}^{(\lambda)}$ for this bias. The adjusted LS estimator relies on sample moments of variance-autocovariance Γ_i with zero elements, i.e. $E(u_{it}u_{is}) = 0$, for $s \neq t$. These moments are weighted by the elements of matrix $\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)}$. They can be consistently estimated under the null hypothesis $\varphi = 1$. Writing analytically matrix $\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)}$ can be easily seen that the elements of this matrix put more weights to sample moments of $E(u_{it}u_{is})$, for $s \neq t$, with s and t defined immediately before break point, T_0.

The next two theorems establish the consistency and the asymptotic local power of test statistic $Z^{(\lambda)}$. The latter is derived under the sequence of local to unity alternatives $\varphi_N = 1 - \frac{c}{\sqrt{N}}$, for $c \geq 0$.

[1]Note that matrix $\Psi^{(\lambda)}$ is used to estimate σ_u^2, based on estimator $\hat{\sigma}_u^2 = \frac{\sum_{i=1}^{N} \Delta y'_i \Psi^{(\lambda)} \Delta y_i}{N tr(\Psi^{(\lambda)})}$ where Δ is the difference operator.

Theorem 2 *Under conditions (a1)-(a3) of Assumption 1, it can be proved that*

$$\lim_{N\to+\infty} P(Z^{(\lambda)} < z_a \mid H_a) = 1, \qquad \lambda \in I, \tag{7}$$

where z_a is the critical value of standard normal distribution at significance level a. The proof is given in the appendix.

Theorem 3 *Under conditions (a1) and (a2) of assumption 1 and the assumption that u_{it} is normally distributed, the limiting distribution of test statistic $Z^{(\lambda)}$ under the local to unity sequence $\varphi_N = 1 - \frac{c}{\sqrt{N}}$ where $c \geq 0$, is given as*

$$Z^{(\lambda)} \equiv \widehat{V}^{(\lambda)-1/2}\hat{\delta}^{(\lambda)}\sqrt{N}\left(\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{\delta}^{(\lambda)}} - 1\right) \xrightarrow{d} N\left(-ck, 1\right), \tag{8}$$

for $N \to \infty$, where

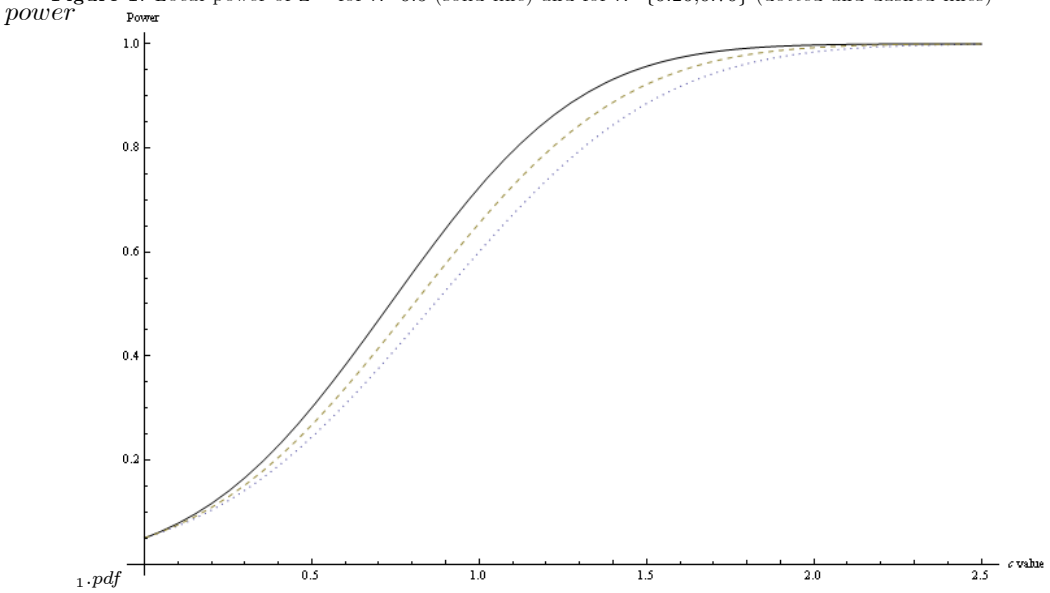
$$k = \frac{\sqrt{3}(T-2)}{\sqrt{T^2(2\lambda^2 - 2\lambda + 1) + 6T + 10 - \frac{4(-\frac{1}{T} + 2(\lambda-1)\lambda T)}{(\lambda-1)\lambda}}}. \tag{9}$$

The limiting distribution given by Theorem 3 implies that local power function of test statistic $Z^{(\lambda)}$ is given as

$$P(Z^{(\lambda)} < z_a | \varphi_N) = \Phi(z_a + ck), \tag{10}$$

where $\Phi(.)$ denotes the cumulative density function (cdf) of standard normal density function $N(0,1)$. This power function always takes values greater than 0.05 since $k > 0$, for $T > 2$.

Figure 1: Local power of $Z^{(\lambda)}$for λ=0.5 (solid line) and for λ={0.20,0.70} (dotted and dashed lines)



Thus, test statistic $Z^{(\lambda)}$ has non trivial asymptotic local power. Due to strict monotonicity of the standard normal distribution $\Phi(.)$, a larger value of k means greater power of $Z^{(\lambda)}$, for all c. Thus, the highest local

power of $Z^{(\lambda)}$ is reached at the point where k achieves its maximum value. From the above formula of k, it can be easily seen that the local power of $Z^{(\lambda)}$ depends on the time dimension T and the fraction of the sample that the break occurs, λ. It does not depend on the initial conditions of the panel data model y_{i0} or individual effects $\alpha_{it}^{(\lambda)}$. Figure 1 presents values of the above local power function across different values of the local parameter c. This is done for the case that $T = 10$ and $\lambda = \{0.20, 0.5, 0.70\}$. Inspection of this figure indicates that test statistic $Z^{(\lambda)}$ has higher power for $\lambda = 0.5$, i.e. when the break is located in the middle of the sample, rather than for $\lambda = \{0.20, 0.70\}$. Analogous graphs of local power function $P(Z^{(\lambda)} < z_a|\varphi_N)$ can be taken for values of T larger than $T = 10$. Note that this local power function is found to increase with T.

2.2 Unknown break point

The results of Theorem 1 are based on the assumption that the break point T_0 is known. Next, we relax this assumption and propose a test statistic of the null hypothesis $\varphi = 1$ which, under the alternative hypothesis of stationarity, allows for a common structural break in the individual effects of model (1) of an unknown date. As in single time series literature (see, e.g., Zivot and Andrews (1992) and Perron and Vogelsang (1998)), we will view the selection of the break point as the outcome of minimizing the standardized test statistic $Z^{(\lambda)}$, given by Theorem 1, over all possible break points of the sample, after trimming out the initial and final parts of the time series observations of the panel data. The minimum value of test statistics $Z^{(\lambda)}$, for all $\lambda \in I$, defined as $z \equiv \min_{\lambda \in I} Z^{(\lambda)}$, will give the least favorable result of the null hypothesis $\varphi = 1$. Let $\hat{\lambda}_{\min}$ denote the break point at which the minimum value of $Z^{(\lambda)}$, over all $\lambda \in I$, is obtained. Then, the null hypothesis will be rejected if we have:

$$Z^{(\hat{\lambda}_{\min})} < c_{\min},$$

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

Theorem 4 *Let condition (a1) of Assumption 1 hold and u_{it} is normally distributed. Then, under the null hypothesis $\varphi = 1$ and unknown λ, we have*

$$z \equiv \min_{\lambda \in I} Z^{(\lambda)} \xrightarrow{d} \zeta \equiv \min_{\lambda \in I} N(0, \Sigma) \tag{11}$$

as $N \to \infty$, where $\Sigma \equiv [\sigma_{\lambda s}]$ is the variance-covariance matrix of the test statistics $Z^{(\lambda)}$, with elements $\sigma_{\lambda s}$ given by the following formula:

$$\sigma_{\lambda s} = \frac{F^{(\lambda)\prime}(K_{T^2} + I_{T^2})F^{(s)}}{\sqrt{F^{(\lambda)\prime}(K_{T^2} + I_{T^2})F^{(\lambda)}}\sqrt{F^{(s)\prime}(K_{T^2} + I_{T^2})F^{(s)}}}, \tag{12}$$

where λ and s denote two different fractions of the sample that the break can occur. See Appendix for a proof.

The result of Theorem 2 implies that critical values of the limiting distribution of the standardized test statistic $\min_{\lambda \in I} Z^{(\lambda)}$, denoted $c_{\min}$, can be obtained from the distribution of the minimum value of a fixed number of $T-2$ correlated normal variables $Z^{(\lambda)}$ with covariance matrix Σ, whose elements are given by Theorem 2. Since $\min\{Z^{(\frac{2}{T})}, Z^{(\frac{3}{T})}..., Z^{(\frac{T-1}{T})}\} = \max\{-Z^{(\frac{2}{T})}, -Z^{(\frac{3}{T})}..., -Z^{(\frac{T-1}{T})}\}$, we can use the distribution of the maximum of normal variables $-Z^{(\lambda)}$ to calculate critical value $c_{\min}$ for a significance level a, i.e.

$$P(\zeta < c_{\min}) = \ P(-\zeta > -c_{\min}) = a. \tag{13}$$

The integral function $P(\zeta > -c_{\min}) = a$ can be calculated numerically based on the probability density function (pdf) of $-\zeta$. This density function has been recently derived by Arellano-Valle and Genton (2008) for the more general case of the maximum of absolutely continuous dependent random variables of elliptically contoured distributions. For the case of normal random variables, it is given as

$$f_{\zeta}(x) = \sum_{\lambda} \phi(x; \mu_{\lambda}, \Sigma_{\lambda,\lambda}) \Phi(x e_{T-3}; \boldsymbol{\mu}_{-\lambda,\lambda}, \boldsymbol{\Sigma}_{-\lambda-\lambda,\lambda}), \quad x \in R, \tag{14}$$

where e_{T-3} is a $(T-3)$-column vector of unities, $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the normal distribution with arguments given as follows:

$$\mu_{-\lambda,\lambda}(x) = \mu_{-\lambda} + (x - \mu_{\lambda})\Sigma_{-\lambda,\lambda}(\Sigma_{\lambda,\lambda})^{-1} \quad \text{and} \quad \Sigma_{-\lambda-\lambda,\lambda} = \Sigma_{-\lambda-\lambda} - \Sigma_{-\lambda,\lambda}\Sigma'_{-\lambda,\lambda}(\Sigma_{\lambda,\lambda})^{-1},$$

where $\mu = (\mu_{-\lambda} \vdots \mu_{\lambda})'$ and $\Sigma = \begin{bmatrix} \Sigma_{-\lambda,-\lambda} & \Sigma_{-\lambda,\lambda} \\ \Sigma_{\lambda,-\lambda} & \Sigma_{\lambda,\lambda} \end{bmatrix}$ are respectively the vector of means and the variance-autocovariance matrix of the $(T-2)$-column vector Z which consists of random variables $Z^{(\boldsymbol{\lambda})}$, for $\lambda \in I$, partitioned as $Z = (Z^{(-\boldsymbol{\lambda})} \vdots Z^{(\boldsymbol{\lambda})})'$, where $Z^{(-\boldsymbol{\lambda})}$ is a $(T$-3)-column vector consisting of the remaining elements of Z excluding $Z^{(\boldsymbol{\lambda})}$.

The above pdf of random variable $-\zeta$, defined as $f_{\zeta}(x)$, is a mixture of the normal marginal densities $\phi(x; \mu_{\lambda}, \Sigma_{\lambda,\lambda})$ corresponding to all possible break fractions of the sample λ. These densities are weighed with the cdf values of the $(T$-3)-column vector xe_{T-3}, given $\Phi(xe_{T-3}; \mu_{-\lambda,\lambda}(x), \Sigma_{-\lambda-\lambda,\lambda})$, giving the probability that $-Z^{(\boldsymbol{\lambda})}$ takes the largest value across all λ (implying that $Z^{(\lambda)}$ takes its minimum value) and the remaining variables of vector Z, collected in vector $Z^{(-\boldsymbol{\lambda})}$, having smaller values.

The consistency of the test given by Theorem 2 follows immediately from Theorem 2, which proves the consistency of $Z^{(\lambda)}$ for a known date break. This can be seen by noting that if, under the alternative hypothesis $\varphi < 1$, test statistic $Z^{(\boldsymbol{\lambda})}$ converges to minus infinity, for $\lambda \in I$, then so does their minimum.

3 Generalizing the test statistics for serially correlated and heterogenous disturbance terms

In this section, we generalize the test statistics presented in the previous section to allow for serially correlated and heterogenous disturbance terms u_{it}, for all i. Due to the fixed-T dimension of panel data model (1)

and the allowance for a common structural break in the individual effects $\alpha_{it}^{(\lambda)}$, the maximum order of serial correlation, denoted as $p_{\max}$, considered by our tests will be a function of T. This will be assumed to be the same for both sample intervals before and after break point T_0. Later on, we will give a table of values of $p_{\max}$ which do not depend on the location of the break, T_0, and thus can be proved very useful for the application of our tests, in practice.

To derive the limiting distribution of test statistics based on estimator $\hat{\varphi}^{(\lambda)}$ under the above more general assumptions of panel data sets, we will make the following assumption about the sequence of the disturbance terms $\{u_i\}$.

Assumption 2 (b1): $\{u_i\}$ constitutes a sequence of independent random vectors of dimension $(TX1)$ with means $E(u_i) = 0$ and variance-autocovariance matrices $E(u_i u_i') = \Gamma_i \equiv [\gamma_{i,ts}]$, where $\gamma_{i,ts} = E(u_{it}u_{is}) = 0$ for $s = t + p_{\max} + 1, ..., T$ and $t < s$. (b2): The average population covariance matrix $\Gamma_N \equiv \frac{1}{N}\sum_{i=1}^{N}\Gamma_i$ is bounded away from zero in large samples: $\bar{\gamma}_{N,tt} > \eta'$ for some $\eta' > 0$ and for all $N > N_0$, for some N_0, and for at least one $t \in \{1, ..., T\}$. (b3): The $4 + \eta$-th population moments of Δy_i, $i = 1, ..., N$, are uniformly bounded i.e. for every real $(TX1)$ vector l such that $l'l = 1$, $E(|l'\Delta y_i|^{4+\eta}) < B < \infty$ for some B. (b4): $\frac{1}{N}\sum_{i=1}^{N} l'Var(vec(\Delta y_i \Delta y_i'))l > \eta'$ for some $\eta' > 0$, and for all $N > N_1$, for some N_1 and for every real $(\frac{1}{2}T(T+1)X1)$ vector l with $l'l = 1$. (b5): $E(u_{it}y_{io}) = E\left(u_{it}a_i^{(1)}\right) = E\left(u_{it}a_i^{(2)}\right) = 0$ and $\forall\ i \in \{1, 2, ..., N\}$, $t \in \{1, 2, ..., T\}$.

Assumption 2 enables us to derive the limiting distribution of $\hat{\varphi}^{(\lambda)} - 1$ by employing standard asymptotic theory under more general conditions than those of Assumption 1, which considers the simple case that $u_{it} \sim NIID(0, \sigma_u^2)$, for all i. More specifically, condition (b1) allows the variance-autocovariance matrices of disturbance terms u_{it}, $\Gamma_i = E(u_i u_i')$, to be heterogenous across the cross-sectional units of the panel i with a degree of serial correlation p_{max} less than T. The pattern of serial correlation considered by matrices Γ_i can capture that implied by moving average (MA) processes of u_{it}, often assumed for many economic series (see, e.g. Schwert (1989)). It can be also though of as approximating that implied by AR models of u_{it} whose autocorrelation dies out after $p_{\max}$.[2] Condition (b2) qualifies application of a central limit theorem (CLT) to derive the limiting distribution of test statistic $\hat{\varphi}^{(\lambda)} - 1$ adjusted for the asymptotic bias (inconsistency) of estimator $\hat{\varphi}^{(\lambda)}$ as $N \to \infty$, under the more general assumptions than condition (b1). More specifically, Condition (b2) along with condition (b4) guarantees that, the variance and the test will be different than zero. Finally, conditions (b5) and (b3) constitute weak conditions under which the consistency of the tests can be proved. These two conditions correspond to conditions (a2) and (a3) of Assumption 1.

The limiting distribution of a normalized panel unit root test statistic based on estimator $\hat{\varphi}^{(\lambda)}$ corrected for its inconsistency under the above assumption is given in the next section. This statistic assumes that the fraction of the sample λ that the break occurs is known.

Theorem 5 *Let conditions (b1) - (b5) of Assumption 2 hold. Then, under the null hypothesis* $\varphi = 1$ *and* λ

[2] In single time series literature, $p_{\max}$ is assumed to increase with T with an order of $o(T^{1/2})$ see Chang and Park (2002).

known, we have

$$Z_1^{(\lambda)} \equiv \hat{V}_1^{(\lambda)-1/2}\hat{\delta}_1^{(\lambda)}\sqrt{N}\left(\hat{\varphi}^{(\lambda)} - \frac{\hat{b}_1^{(\lambda)}}{\hat{\delta}_1^{(\lambda)}} - 1\right) \xrightarrow{d} N(0,1) \tag{15}$$

as $N \to \infty$*, where*

$$\frac{\hat{b}_1^{(\lambda)}}{\hat{\delta}_1^{(\lambda)}} = \frac{tr(\Psi_1^{(\lambda)}\hat{\Gamma}_N)}{\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1}Q^{(\lambda)}y_{i,-1}} \tag{16}$$

is a consistent estimate of the asymptotic bias of $\hat{\varphi}^{(\lambda)}$ *which, under the null hypothesis, is given as*

$$\frac{b_1^{(\lambda)}}{\delta_1^{(\lambda)}} = \frac{tr(\Lambda' Q^{(\lambda)}\Gamma_N)}{tr(\Lambda' Q^{(\lambda)}\Lambda\Gamma_N)}, \tag{17}$$

where matrix $\Psi_1^{(\lambda)}$ *is a* (TXT)*-dimension matrix having in its main diagonal, and its p-lower and* $p-$*upper diagonals of the main diagonal the corresponding elements of matrix* $\Lambda' Q^{(\lambda)}$*, and zero otherwise,* $\hat{\Gamma}_N = \frac{1}{N}\sum_{i=1}^{N}(\Delta y_i \Delta y'_i)$ *is a consistent estimator of population variance-autocovariance matrix* Γ_N *and* $V_1^{(\lambda)}$ *is a variance function given as*

$$V_1^{(\lambda)} = F_1^{(\lambda)\prime}\Theta F_1^{(\lambda)\prime}, \tag{18}$$

where $F_1^{(\lambda)} = vec(Q^{(\lambda)}\Lambda - \Psi_1^{(\lambda)\prime})$ *and* $\Theta = \frac{1}{N}\sum_{i=1}^{N} Var(vec(u_i u'_i))$ *is the variance-covariance matrix of* $vec(u_i u'_i)$*. The proof is given in the appendix.*

To implement the test statistic given by Theorem 5, $Z_1^{(\lambda)}$, to test hypothesis $\varphi = 1$, we need consistent estimates of the variance-covariance matrix of vector $vec(u_i u'_i)$, defined as Θ.This can be done under the null hypothesis based on the following estimator:

$$\hat{\Theta} = \frac{1}{N}\sum_{i=1}^{N}\left(vec(\Delta y_i \Delta y'_i)vec(\Delta y_i \Delta y'_i)'\right). \tag{19}$$

As $\Psi^{(\lambda)}$ for $Z^{(\lambda)}$, matrix $\Psi_1^{(\lambda)}$ plays a crucial role in constructing test statistic $Z_1^{(\lambda)}$. It adjusts LS estimator $\hat{\varphi}^{(\lambda)}$ for its asymptotic bias. This bias now comes from two sources: the "within" transformation of the data through matrix $Q^{(\lambda)}$ and the serial correlation of disturbance terms u_{it}.[3] Subtracting $\Psi_1^{(\lambda)}$ from $\Lambda' Q^{(\lambda)}$ enables to adjust $\hat{\varphi}^{(\lambda)}$ for the above two sources of bias. The adjusted LS estimator $\hat{\varphi}^{(\lambda)}$ enables us to test the null hypothesis $\varphi = 1$, as it relies on sample moments of the elements of variance-autocovariance matrices Γ_i, for all i, which are serially uncorrelated, i.e. $E(u_{it}u_{is}) = 0$, for $s = t + p_{\max} + 1, ..., T$ and $t < s$. These moments are weighted by elements of matrix $\Lambda' Q^{(\lambda)} - \Psi_1^{(\lambda)}$, which assign higher weights to the moments which are immediately before the break point T_0 than those which are away than it. They can be consistently estimated under the null hypothesis through the variance-covariance estimator $\hat{\Theta}$. The weights that matrix $\Lambda' Q^{(\lambda)} - \Psi_1^{(\lambda)}$ assigns to the above elements of variance-autocovariance matrices Γ_i obviously depend on the break point and the maximum order of serial correlation $p_{\max}$ is assumed by test statistic $Z_1^{(\lambda)}$. Based on the specification of this matrix, Table 1 and (20) give values of $p_{\max}$ which enable us to apply our tests independently on the location of the break T_0, or the fraction of the sample λ. These values

[3]Note that, under conditions of Assumption 1, test statistic $Z_1^{(\lambda)}$ becomes identical to $Z^{(\lambda)}$.

are found based on the criterion that the elements of matrix $\Lambda' Q^{(\lambda)} - \Psi_1^{(\lambda)}$ do not assign weights to zero elements of Γ_i which result in a value of variance function $V_1^{(\lambda)}$ which is zero, i.e. $V_1^{(\lambda)} = 0$. They are useful in choosing the maximum order of serial correlation $p_{\max}$, considered by test statistic $Z_1^{(\lambda)}$, especially when the break is of an unknown date.

Table 1: Maximum order of serial correlation

T	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$p_{\max}$	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8

$$p_{\max} = [\frac{T}{2} - 2]^* \tag{20}$$

where $[.]^*$ denotes the greatest integer function. Note that, in the case that disturbance tests u_{it} are normally distributed, the variance function $V_1^{(\lambda)}$ can be written more analytically as follows:

$$V_1^{(\lambda)} = F_1^{(\lambda)\prime}(K_{T^2} + I_{T^2})(\Gamma_N \otimes \Gamma_N)F_1^{(\lambda)}, \tag{21}$$

where $\otimes$ denotes the Kronecker product.[4] This form of $V_1^{(\lambda)}$ can be easily calculated by replacing Γ_N with its consistent estimate $\hat{\Gamma}_N = \frac{1}{N}\sum_{i=1}^{N}(\Delta y_i \Delta y_i')$.

The test statistic given by Theorem 5, $Z_1^{(\lambda)}$, can be proved that is consistent and has asymptotic local power. Next, we derive the limiting distribution of $Z_1^{(\lambda)}$ under the local to unity sequence of alternatives $\varphi_N = 1 - \frac{c}{\sqrt{N}}$, for $c \geq 0$. Our analysis will enable us to examine how the presence of serial correlation in the disturbance terms u_{it} affects the local power of $Z_1^{(\lambda)}$.

Theorem 6 *Under conditions (b1)-(b4) of Assumption 2, the assumption that u_{it} are normally distributed and the local to unity sequence $\varphi_N = 1 - \frac{c}{\sqrt{N}}$, for $c \geq 0$, the limiting distribution of $Z_1^{(\lambda)}$ is given as*

$$Z_1^{(\lambda)} \equiv V_1^{(\lambda)-1/2}\hat{\delta}_1^{(\lambda)}\sqrt{N}\left(\hat{\varphi}^{(\lambda)} - \frac{\hat{b}_1^{(\lambda)}}{\hat{\delta}_1^{(\lambda)}} - 1\right) \xrightarrow{d} N(-ck_1, 1), \tag{22}$$

for $N \to \infty$, where

$$k_1 = \frac{tr(\Xi' Q^{(\lambda)}\Gamma_N) + tr(\Lambda' Q^{(\lambda)}\Lambda\Gamma_N) - tr(\Psi_1^{(\lambda)}\Lambda\Gamma_N) - tr(\Lambda'\Psi_1^{(\lambda)}\Gamma_N)}{\sqrt{F_1^{(\lambda)\prime}(K_{T^2} + I_{T^2})(\Gamma_N \otimes \Gamma_N)F_1^{(\lambda)}}}, \tag{23}$$

with $\Xi = \frac{d\Omega}{d\varphi}|_{\varphi=1}$, where Ω is a (TXT)-dimension matrix defined in the appendix. See Appendix for the proof.

The result of Theorem 6 implies that the local power of test statistic $Z_1^{(\lambda)}$ is given by the following

[4] This can be easily seen using standard results of the variance of a quadratic form for normally distributed variates (see e.g. Schott(1996)), which imply

$$Var[vec(u_i u_i')] = Var(u_i \otimes u_i) = (I_{T^2} + K_{T^2})(\Gamma_N \otimes \Gamma_N).$$

function:

$$P(Z_1^{(\lambda)} < z_a|\varphi_N) = \Phi(z_a + ck_1). \tag{24}$$

The value of parameter k_1, which determines the power of test statistic $Z_1^{(\lambda)}$, depends on the values of variance-autocovariance matrices Γ_i. To examine the effects of serial correlation on the local power of $Z_1^{(\lambda)}$, Table 2 reports values of k_1 for T=10, $\lambda = \{0.25, 0.5, 0.75\}$ and a set of different values of the moving average parameter θ of the following MA(1) process of disturbance terms u_{it}: $u_{it} = \varepsilon_{it} + \theta\varepsilon_{it-1}$, for all i, i.e., $\theta \in \{-0.5, 0, 0.5\}$. Note the table also reports values of k_1 for the case that the order of serial correlation p assumed by test statistic $Z_1^{(\lambda)}$ is higher than that of the true data generating process, which is $p = 1$. It also provides results for the case that $\theta = 0$, where test statistic $Z_1^{(\lambda)}$ reduces to $Z^{(\lambda)}$ which assumes no serial correlation (i.e. $p_{\max} = 0$). The above cases of p can show someone who assumes higher order of serial correlation than the correct order, which may happen in practice, the impact on the local power of the test, given that the adjusted for its bias LS estimator $\hat{\varphi}^{(\lambda)}$ will rely on fewer moment conditions. Finally Table 2 also contains the Monte Carlo estimates of local power next to the theoretical values. These results were taken from 5000 replications, N=1000, $\varphi = 0.968$, $y_{i0} = 0$, $a_i^{(\lambda)} = 0$. Nuisance parameter values were set to zero as they dot not enter the local power function**.** The results of Table 2 lead to a number of interesting conclusions.

First, they show that the test has local power for all values of λ and θ considered, except for some cases where θ takes large negative values, i.e. $\theta = -0.5$. Second, the power function of the test depends on the fraction of the sample λ that the break occurs. When $\theta \geq 0$, the power function takes its largest values for $\lambda = 0.5$ (i.e. the break is in the middle of the sample). On the other hand, if $\theta < 0$, then the power function takes its largest values for $\lambda = 0.70$. That is, the break occurs towards the end of the sample. In contrast to what was expected, there is not always power reduction if a higher order of serial correlation is assumed than the correct (true) order. For instance, for the following cases of λ and θ: ($\lambda = 0.70$, $\theta < 0$) and ($\lambda = 0.20$, $\theta > 0$), the power slightly increases with the order of serial correlation p, despite the fact that the correct order is less than p. This result can be attributed to the weights that matrix $\Psi_1^{(\lambda)}$ assigns to the moments of matrix Γ_i employed to adjust estimator $\hat{\varphi}^{(\lambda)}$ for its inconsistency. It is an effect that is caused by serial correlation because as is found in Karavias and Tzavalis (2012b) the effect of p on the test is weaker than that of serial correlation. For $\theta = 0$ power drops as p increases. From (23), it can be seen that $\Psi_1^{(\lambda)}$ affects k_1 and, hence, local power $P(Z_1^{(\lambda)} < z_a|\varphi_N)$ through the values of traces $tr(\Psi_1^{(\lambda)}\Lambda\Gamma_N)$ and $tr(\Lambda'\Psi_1^{(\lambda)}\Gamma_N)$, and variance function $V_1^{(\lambda)} = F_1^{(\lambda)\prime}(K_{T^2} + I_{T^2})(\Gamma_N \otimes \Gamma_N)F_1^{(\lambda)}$ but mainly depends on the values of traces. By writing analytically matrix $\Psi_1^{(\lambda)}$, it can be seen that, as p increases, this matrix assigns higher weights to the sample moments of variance-autocovariance matrices $\Gamma_i = E(u_{it}u_{is}) = 0$, for $s = t + p_{\max} + 1, ..., T$ and $t < s$, employed by the adjusted for its bias estimator $\hat{\varphi}^{(\lambda)}$, which are closer to the break point T_0 if they

are before and closer to T if they are after T_0.[5] Under different values of λ and θ, these moments contain very useful sample information which helps to identify the existence of a break point in individual effects $\alpha_{it}^{(\lambda)}$ from the data. Thus, increasing their weights results in an increase of the local power of $Z_1^{(\lambda)}$ when p increases, despite the fact that the adjusted for its bias estimator $\hat{\varphi}^{(\lambda)}$ relies on a smaller number of sample moments of $E(u_{it}u_{is})$ than its maximum possible number.

Table 2: Local power with serial correlation

			$\theta=-0.5$		$\theta=0$		$\theta=0.5$	
T	p	T_0	$\Phi(z_a+k_1)$	$\hat{\Phi}(z_a+k_1)$	$\Phi(z_a+k_1)$	$\hat{\Phi}(z_a+k_1)$	$\Phi(z_a+k_1)$	$\hat{\Phi}(z_a+k_1)$
8	1	2	0.01	0.093	0.47	0.378	0.83	0.58
		4	0.15	0.126	0.59	0.537	0.92	0.84
		7	0.47	0.08	0.49	0.401	0.51	0.69
	2	2	0	0.109	0.44	0.359	0.93	0.601
		4	0.14	0.106	0.40	0.354	0.73	0.662
		7	0.70	0.097	0.47	0.401	0.34	0.660
10	1	2	0	0.106	0.50	0.369	0.83	0.501
		5	0.14	0.113	0.68	0.575	0.92	0.859
		9	0.43	0.070	0.51	0.384	0.51	0.645
	2	2	0	0.106	0.49	0.369	0.88	0.52
		5	0.15	0.117	0.59	0.509	0.88	0.804
		9	0.61	0.069	0.50	0.355	0.43	0.637

The Monte Carlo experiment shows that the effects of T, serial correlation and loss of moments are adequately captured by the estimates. The effect of the break fraction is not so clear. It seems that fixed T tests do not respond with precision to the effects of λ as is also found by Karavias and Tzavalis (2012a).

Following analogous steps to those required by Theorem 4, test statistic $Z_1^{(\lambda)}$ can be easily extended to the case of an unknown break point date, which requires a sequential application of the test. Define this test statistic as $z_1 \equiv \min_{\lambda \in I} Z_1^{(\lambda)}$. The limiting distribution of z_1 is given as

$$z_1 \equiv \min_{\lambda \in I} Z_1^{(\lambda)} \xrightarrow{d} \zeta_1 \equiv \min_{\lambda \in I} N(0, \Sigma_1), \tag{25}$$

$N \to \infty$, where $\Sigma_1 \equiv [\sigma_{1,\lambda s}]$ is the variance-covariance matrix of the test statistics $Z_1^{(\lambda)}$ whose elements,

[5] For instance, for $T=6$, $\lambda=0.5$ and $p=1$ matrix $\Psi_1^{(\lambda)}$ becomes $\Psi_1^{(\lambda)} = \begin{pmatrix} -2/3 & -1/3 & 0 & 0 & 0 & 0 \\ 1/3 & -1/3 & 0 & 0 & 0 & 0 \\ 0 & 2/3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2/3 & -1/3 & 0 \\ 0 & 0 & 0 & 1/3 & -1/3 & 0 \\ 0 & 0 & 0 & 0 & 2/3 & 0 \end{pmatrix}$.

defined as $\sigma_{1,\lambda s}$, are given by the following formula:

$$\sigma_{1,\lambda s} = \frac{F_1^{(\lambda)\prime}\Theta F_1^{(s)}}{\sqrt{F_1^{(\lambda)\prime}\Theta F_1^{(\lambda)}}\sqrt{F_1^{(s)\prime}\Theta F_1^{(s)}}}. \tag{26}$$

Critical values of the distribution of random variable ζ_1, denoted as $f_{\zeta_1}(x_1)$ where $x_1 \in R$, can be calculated by replacing the values of $\sigma_{\lambda s}$ in pdf formula (14) with those of $\sigma_{1,\lambda s}$. This also requires to obtain consistent estimates of variance-covariance matrix Θ, in the first step.

4 Extension of the tests to the case of deterministic trends

In this Section, we will extend the tests presented in the previous section to allow for individual (incidental) linear trends in the panel data generating processes. We will consider two cases of AR(1) panel data models with linear trends. In the first case, we will assume that these trends are present only under the alternative hypothesis of stationarity (see, e.g. Karavias and Tzavalis (2012a), or Zivot and Andews (1992) for single time series), while in the second that they are present under the null hypothesis of $\varphi = 1$ either (see, e.g. Carrion-i-Silvestre et al (2005) and Kim (2011). The first of the above cases is more appropriate in distinguishing between nonstationary panel data series which exhibit persistent random deviations from linear trends, implied by the presence of individual effects under the null hypothesis of unit roots, and stationary panel data series allowing for broken individual linear trends. The second case is more suitable when considering more explosive panel data series under the null hypothesis of unit roots, which can exhibit both deterministic and random persistent shifts from their linear trends.

4.1 Broken trends under the alternative hypothesis of stationarity

Consider the following extension of the nonlinear AR(1) model (1):

$$y_{it} = \alpha_{it}^{(\lambda)}(1-\varphi) + \varphi\beta_i + \beta_{it}^{(\lambda)}(1-\varphi)t + \varphi y_{it-1} + u_{it}, \quad i = 1, ..., N \tag{27}$$

where $\alpha_{it}^{(\lambda)}$ are defined by equation (1) and $\beta_{it}^{(\lambda)} = \beta_i^{(1)}$ if $t \leq T_0$ and $\beta_i^{(2)}$ if $t > T_0$. Under the null hypothesis $\varphi = 1$, β_i's constitute individual effects of the panel data model, which capture linear trends in the level of series y_{it}, for all i. Under the alternative hypothesis $\varphi < 1$, β_i will be given as $\beta_i = \beta_i^{(1)}$ if $t \leq T_0$ and $\beta_i^{(2)}$ if $t > T_0$. That is, they constitute the slope coefficients of individual linear trends t, for all i.

Let us define matrix $X_*^{(\lambda)} = (e^{(1)}, e^{(2)}, \tau^{(1)}, \tau^{(2)})$, where $\tau^{(1)}$ and $\tau^{(2)}$ are $(TX1)$-column vectors whose elements are given as $\tau_t^{(1)} = t$ if $t \leq T_0$, and zero otherwise, and $\tau_t^{(2)} = t$ if $t > T_0$, and zero otherwise. Then, the "within" transformation matrix now will be written as $Q_*^{(\lambda)} = I_T - X_*^{(\lambda)}(X_*^{(\lambda)\prime}X_*^{(\lambda)})^{-1}X_*^{(\lambda)\prime}$ and the LSDV estimator, denoted as $\hat{\varphi}_*^{(\lambda)}$, can be written under null hypothesis $\varphi = 1$ as follows:

$$\hat{\varphi}_*^{(\lambda)} - 1 = \left[\sum_{i=1}^{N} y_{i,-1}' Q_*^{(\lambda)} y_{i,-1}\right]^{-1} \left[\sum_{i=1}^{N} y_{i,-1}' Q_*^{(\lambda)} u_i\right]. \tag{28}$$

Following analogous steps to those for the derivation of test statistics $Z^{(\lambda)}$ or $Z_1^{(\lambda)}$, inference about unit roots can be conducted based on estimator $\hat{\varphi}_*^{(\lambda)}$ adjusted for its asymptotic bias. Under conditions of Assumption 2, this bias now is given as

$$\frac{b_2^{(\lambda)}}{\delta_2^{(\lambda)}} = \frac{tr(\Lambda' Q_*^{(\lambda)} \Gamma_N)}{tr(\Lambda' Q_*^{(\lambda)} \Lambda \Gamma_N)} \tag{29}$$

(see Appendix, proof of Theorem 7). However, in contrast to the case of model (1), the average population variance-autocovariance matrix Γ_N cannot be consistently estimated based on estimator $\hat{\Gamma}_N = \frac{1}{N}\sum_{i=1}^{N}(\Delta y_i \Delta y_i')$, due to the presence of individual effects β_i under the null hypothesis $\varphi = 1$. It can be easily seen that, under $\varphi = 1$, $\Delta y_i = u_i + \beta_i e$, where e is a $(TX1)$-vector of unities, and thus

$$\frac{1}{N}\sum_{i=1}^{N} E(\Delta y_i \Delta y_i') = \Gamma_N + \beta_N^2 J_T, \tag{30}$$

where J_T is a $T \times T$ matrix of ones and $\beta_N^2 = \frac{1}{N}\sum_{i=1}^{N} E((\beta_i)^2)$. The last relationship clearly shows that in order to provide consistent estimates of Γ_N based on $\hat{\Gamma}_N = \frac{1}{N}\sum_{i=1}^{N}(\Delta y_i \Delta y_i')$, we need to substitute out the average of squared individual effects β_N^2 in $\hat{\Gamma}_N$. This can be done with the help of a (TXT)-dimension selection matrix M, defined as follows: M has elements $m_{ts} = 0$ if $\gamma_{ts} \neq 0$ and $m_{ts} = 1$ if $\gamma_{ts} = 0$. That is, the elements of this matrix correspond to the elements of matrix $\Gamma_N + \beta_N^2 J_T$ (or $\frac{1}{N}\sum_{i=1}^{N} E(\Delta y_i \Delta y_i')$ which contain only β_N^2. Based on matrix M, we can derive a consistent estimator of β_N^2 under the null hypothesis, which is given as

$$\frac{1}{tr(MJ_T)N}\sum_{i=1}^{N} \Delta y_i' M \Delta y_i \xrightarrow{p} \beta_N^2, \tag{31}$$

since $tr(M\Gamma_N) = 0$, where "$\xrightarrow{p}$" signifies convergence in probability. As mentioned above, this estimator enables us to substitute out individual effects β_N^2 in the estimator of Γ_N, given by $\hat{\Gamma}_N$. Then, a consistent estimator of the bias of the LSDV estimator $\hat{\varphi}_*^{(\lambda)}$ for model (27), defined as $\frac{b_2^{(\lambda)}}{\delta_2^{(\lambda)}}$, can be obtained. This is given as

$$\frac{\hat{b}_2^{(\lambda)}}{\hat{\delta}_2^{(\lambda)}} = \frac{tr(\Psi_2^{(\lambda)} \hat{\Gamma}_N)}{\frac{1}{N}\sum_{i=1}^{N} y_{i,-1}' Q_*^{(\lambda)} y_{i,-1}}, \tag{32}$$

where $\Psi_2^{(\lambda)} = \Psi_{1*}^{(\lambda)} + \frac{tr(\Lambda' Q_*^{(\lambda)} M)}{trace(MJ_T)} M$ (see Appendix, proof of Theorem 7) where $\Psi_{1*}^{(\lambda)}$ is a (TXT)-dimension matrix having in its main diagonal, and its p-lower and $p-$upper diagonals of the main diagonal the corresponding elements of matrix $\Lambda' Q_*^{(\lambda)}$, and zero otherwise. It can be easily seen that $tr(\Psi_2^{(\lambda)} \hat{\Gamma}_N)$ is a consistent estimator of $\hat{b}_2^{(\lambda)}$, since $tr(\Psi_2^{(\lambda)}(\Gamma_N + \beta_N^{(1)} J_T)) = tr(\Psi_2^{(\lambda)} \Gamma_N)$.

Having derived a consistent estimator of the asymptotic bias of LS estimator $\hat{\varphi}_*^{(\lambda)}$, next we derive the limiting distribution of a test statistic of hypothesis $\varphi = 1$ based on this estimator adjusted for its bias. This is done after trimming out two time series observations from the end of the sample, i.e. $\lambda = \frac{T_0}{T} \in I^* = \left\{\frac{2}{T}, \frac{2}{T}, ..., \frac{T-2}{T}\right\}$, due to the presence of individual effects and linear trends under the alternative hypothesis of stationarity. To derive this limiting distribution and to prove the consistency of the test, we rely the following assumption.

Assumption 3 Let all conditions of Assumption 2 hold and we also have: $E(u_{it}\beta_i) = 0$, $\forall\ i \in \{1, 2, ..., N\}$, $t \in \{1, 2, ..., T\}$, $E(a_{it}^{(\lambda)}\beta_{it}^{(\lambda)}) = 0$, $\forall\ i \in \{1, 2, ..., N\}$.

Then, the following theorem applies.

Theorem 7 *Let the sequence $\{y_{i,t}\}$ be generated according to model (27) and conditions (b1)-(b4) of Assumption 2 hold. Then, under the null hypothesis $\varphi = 1$ and λ known, we have*

$$Z_2^{(\lambda)} \equiv \hat{V}_2^{(\lambda)-0.5}\hat{\delta}_2^{(\lambda)}\sqrt{N}\left(\hat{\varphi}_*^{(\lambda)} - 1 - \frac{\hat{b}_2^{(\lambda)}}{\hat{\delta}_2^{(\lambda)}}\right) \xrightarrow{d} N(0,1), \tag{33}$$

as $N \to \infty$, where $V_2^{(\lambda)} = F_2^{(\lambda)\prime}\Theta F_2^{(\lambda)}$, Θ is defined in Theorem 2, and $F_2^{(\lambda)} = vec(Q_*^{(\lambda)}\Lambda - \Psi_2^{(\lambda)\prime})$. The proof of the theorem is given in the appendix.

Apart from the initial conditions of the panel y_{i0}, the test statistic given by Theorem 7, $Z_2^{(\lambda)}$, is similar under the null hypothesis to the individual effects of the panel β_i, due to the allowance of broken trends in the "within" transformation matrix, $Q_*^{(\lambda)}$. To test the null hypothesis of unit roots, test statistic $Z_2^{(\lambda)}$ relies on the same moments to those assumed by statistic $Z_1^{(\lambda)}$, namely $E(u_{it}u_{is}) = 0$, for $s = t + p_{\max} + 1, ..., T$ and $t < s$. These moments now are weighted by elements of matrix $\Lambda' Q_*^{(\lambda)} - \Psi_2^{(\lambda)}$, where matrix $\Psi_2^{(\lambda)}$ is appropriately adjusted to wipe off the effects of nuisance parameters β_i on the limiting distribution of the test statistic. The maximum order of serial correlation of variance-autocovariance matrices Γ_i assumed by test statistic $Z_2^{(\lambda)}$ is the same to that assumed by test statistic $Z_1^{(\lambda)}$. It can be easily shown that $Z_2^{(\lambda)}$ is a consistent test, based on the conditions of Assumption 3. The limiting distribution of $Z_2^{(\lambda)}$ under the local to unity sequence of alternatives $\varphi_N = 1 - \frac{c}{\sqrt{N}}$, for $c \geq 0$, is given in the next theorem.

Theorem 8 *Under Assumption 2 and Assumption 3, normally distributed disturbance terms u_{it} and the local to unity sequence $\varphi_N = 1 - \frac{c}{\sqrt{N}}$, for $c \geq 0$, the limiting distribution of test statistic $Z_2^{(\lambda)}$ is given*

$$Z_2^{(\lambda)} = \hat{V}_2^{(\lambda)-1/2}\hat{\delta}_2^{(\lambda)}\sqrt{N}\left(\hat{\varphi}_*^{(\lambda)} - \frac{\hat{b}_2^{(\lambda)}}{\hat{\delta}_2^{(\lambda)}} - 1\right) \to N(k_2, 1), \tag{34}$$

where

$$k_2 = \frac{-\Pi_1 - c[\Pi_2 + \Pi_3]}{\sqrt{F_2^{(\lambda)\prime}(K_{T^2} + I_{T^2})(\Gamma_N \otimes \Gamma_N)F_2^{(\lambda)}}}, \tag{35}$$

$\Pi_1 = tr(e_*' X_*^{(\lambda)\prime}\Psi_2^{(\lambda)} X_*^{(\lambda)} e_*\Sigma_b)$, $\Pi_2 = tr(\Lambda' Q_*^{(\lambda)}\Lambda\Gamma_N) + tr(\Xi Q_*^{(\lambda)}\Gamma_N) - tr(\Psi_2^{(\lambda)}\Lambda\Gamma_N) - tr(\Lambda'\Psi_2^{(\lambda)}\Gamma_N)$, $\Sigma_b = \frac{1}{N}\sum_{i=1}^N E(\nu_i\nu_i')$, $\nu_i = (\beta_i^{(1)}, \beta_i^{(2)})'$, $\Pi_3 = -tr(e^{*\prime} X_*^{(\lambda)\prime}\Lambda'\Psi_2^{(\lambda)} X_*^{(\lambda)} e_*\Sigma_b) - tr(e^{*\prime} X_*^{(\lambda)\prime}\Psi_2^{(\lambda)}\Lambda X_*^{(\lambda)} e_*\Sigma_b) + tr(X_*^{(\lambda)\prime}\Psi_2^{(\lambda)} X_*^{(\lambda)} e_*\Sigma_{b\mu}) + tr(e_* X_*^{(\lambda)\prime}\Psi_2^{(\lambda)} X_*^{(\lambda)}\Sigma_{b\mu}')$, $\mu_i = (\alpha_i^{(1)} - \beta_i^{(1)}, \alpha_i^{(2)} - \beta_i^{(2)}, \beta_i^{(1)}, \beta_i^{(2)})'$, $\Sigma_{b\mu} = \frac{1}{N}\sum_{i=1}^N E(\nu_i\mu_i')$ *and* $e_* = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$ *if $c > 0$ and $(1, 1, 0, 0)$ if $c = 0$. The proof is given in the appendix.*

The results of Theorem 8 imply that the local power of test statistic $Z_2^{(\lambda)}$ is given as

$$P(Z_2^{(\lambda)} < z_a|\varphi_N) = \Phi(z_a - k_2). \tag{36}$$

This function shows that $Z_2^{(\lambda)}$ has non trivial power when $-k_2 > 0$. Note that, under the null hypothesis ($c = 0$), the limiting distribution (34) corresponds to that given by (33), since $\Pi_1 = 0$ since e^* becomes $(1, 1, 0, 0)$.[6] If disturbance terms u_{it} are homoscedastic, for all i and t, and not serially correlated, then k_2 simplifies to following:

$$k_2 = \frac{-\Pi_1 - c\Pi_3}{\sqrt{F_2^{(\lambda)\prime}(K_{T^2} + I_{T^2})(\Gamma_N \otimes \Gamma_N)F_2^{(\lambda)}}}, \tag{37}$$

since $\Pi_2 = 0$.

The results of Theorem 8 indicate that the local power of test statistic $Z_2^{(\lambda)}$ depends on the following: the fraction of break λ, the second order moments of nuisance parameters $\beta_i^{(1)}$ and $\beta_i^{(2)}$, defined as $\beta_N^{(j)} = \frac{1}{N}\sum_{i=1}^{N} E((\beta_i^{(j)})^2)$, for $j = 1, 2$, and the order of serial correlation p considered. The latter affects the elements of matrix $\Psi_2^{(\lambda)}$. To examine how λ, p and $\beta_N^{(j)}$ affect the asymptotic local power of test statistic $Z_2^{(\lambda)}$, defined as $P(Z_2^{(\lambda)} < z_a|\varphi_N)$, in Table 3 we present values k_2 for the case that disturbance terms u_{it} are given as $u_{it} = \varepsilon_{it} + \theta\varepsilon_{it-1}$, for all i. This is done for $c = 1$, $T \in \{8, 10\}$, $T_0 \in \{2, 4, 6\}$ for $T = 8$ and $T_0 \in \{2, 5, 8\}$ for $T = 10$ and $\theta \in \{-0.5, 0, 0.5\}$. For $\beta_N^{(1)}$ and $\beta_N^{(2)}$ we assume $\beta_N^{(1)} = \beta_N^{(2)} \in \{0.2, 0.8\}$. The most important result of Table 3 is that test statistic $Z_2^{(\lambda)}$ has non trivial local power in a neighborhood of unity which shrinks at $\sqrt{N}$ rate. In particular, the test has always significant local power independently on the value of θ, when $\lambda = 0.5$ and $p > 1$. That is, the break point is in the middle of the sample and there is allowance for serial correlation of order two although the true order is 1. In fact, for most cases local power increases with the order of serial correlation p, even if this order is higher than the correct one. This result can be explained based on analogous arguments to those followed for the interpretation of the local power increase of test statistic $Z_1^{(\lambda)}$ (see Theorem 6) with lag order p. It can be attributed to the weights that matrix $\Psi_2^{(\lambda)}$ imposes on the sample moments of auto-covariances Γ_i employed by the adjusted for its bias LS estimator $\hat{\varphi}_*^{(\lambda)}$ to carry out the test.

The increase of $P(Z_2^{(\lambda)} < z_a|\varphi_N)$ with the increase of the values of $\beta_N^{(1)}$ and $\beta_N^{(2)}$ means that high variability in the nuisance parameters $\beta_i^{(1)}$ and $\beta_i^{(2)}$ increases the local power . But as T increases from 8 to 10 it wipes out the effect of $\beta_N^{(1)}$ and $\beta_N^{(2)}$. Another interesting conclusion implied by the results of Table 3 is that the power of test statistic $Z_2^{(\lambda)}$ also decreases as we move from negative to positive values of MA parameter θ. This is the opposite behavior from that seen in Table 2 investigating the local power of $Z_1^{(\lambda)}$. Finally although these results are not reported, for $p = 0$, the test is always biased. In the large T panel unit root test literature for $p = 0$ the tests have trivial power. But in this case the test is biased due to the assumption that under the null there is no break. This will further clarified in Section XXX where a break under the null hypothesis is allowed.

[6] The assumption that $a_{it}^{(\lambda)}$ and $\beta_{it}^{(\lambda)}$ are uncorrelated (see Assumption 3) simplifies the analysis because it makes the local power function to depend only on $\beta_N^{(1)}$ and $\beta_N^{(2)}$. However, this assumption can be relaxed. If $E(a_{it}^{(\lambda)}\beta_{it}^{(\lambda)}) \neq 0$, then the local power function will also depend on the covariance between $a_{it}^{(\lambda)}$ and $\beta_{it}^{(\lambda)}$.

Table 3: Local power with trend

				$\theta=-0.5$		$\theta=0$		$\theta=0.5$	
T	p	β_N	λ	$\Phi(z_a+k_1)$	$\hat{\Phi}(z_a+k_1)$	$\Phi(z_a+k_1)$	$\hat{\Phi}(z_a+k_1)$	$\Phi(z_a+k_1)$	$\hat{\Phi}(z_a+k_1)$
8	1	0.2	2	0	0	0	0	0.02	0
			4	0.05	0.056	0.05	0.047	0.04	0.043
			6	0.01	0	0	0	0	0
		0.8	2	0	0	0	0	0	0
			4	0.05	0.047	0.05	0.046	0.04	0.037
			6	0	0	0	0	0	0
	2	0.2	2	0.06	0.10	0.06	0.102	0.05	0.103
			4	0.57	1	0.76	1	0.62	1
			6	0.10	0.129	0.06	0.116	0.03	0.063
		0.8	2	0.10	0.231	0.104	0.283	0.08	0.230
			4	1	1	1	1	1	1
			6	0.16	0.306	0.104	0.355	0.05	0.220
10	1	0.2	2	0	0	0	0	0.01	0
			5	0	0	0	0	0	0
			8	0	0	0	0	0	0
		0.8	2	0	0	0	0	0	0
			5	0	0	0	0	0	0
			8	0	0	0	0	0	0
	2	0.2	2	0.01	0	0	0	0.03	0.003
			5	0.35	0.985	0.41	1	0.30	0.984
			8	0.04	0	0	0	0.01	0.001
		0.8	2	0	0	0	0	0	0
			5	1	1	1	1	1	0.998
			8	0	1	0	1	0	1

Finally, note that test statistic $Z_2^{(\lambda)}$ can be extended to the case of an unknown date break following an analogous procedure to that assumed for sequential tests statistics z and z_1, defined by equations (11) and (25), respectively. This version of the test statistic will be denoted as $z_2 \equiv \min_{\lambda \in I^*} Z_2^{(\lambda)}$. Its limiting distribution is given as

$$z_2 \equiv \min_{\lambda \in I^*} Z_2^{(\lambda)} \xrightarrow{d} \zeta_2 \equiv \min_{\lambda \in I^*} N(0, \Sigma_2), \tag{38}$$

as $N \to \infty$, where $\Sigma_2 \equiv [\sigma_{2,\lambda s}]$ is the variance-covariance matrix of test statistics $Z_2^{(\lambda)}$ whose elements $\sigma_{2,\lambda s}$ are given by the formula: $\sigma_{2,\lambda s} = \frac{F_2^{(\lambda)\prime} \Theta F_2^{(s)}}{\sqrt{F_2^{(\lambda)\prime} \Theta F_2^{(\lambda)}} \sqrt{F_2^{(s)\prime} \Theta F_2^{(s)}}}$. Critical values of the distribution of ζ_2 can be

derived based on the pdf $f_\zeta(x)$, given by (14), following an analogous procedure to that assumed for test statistic z_1.

4.2 Broken trends under the null hypothesis of unit roots

To allow for a common break in the individual effects of the panel data model under the null hypothesis $\varphi = 1$, consider the following extension of $AR(1)$ model (1):

$$y_{it} = \alpha_{it}^{(\lambda)}(1-\varphi) + \varphi\beta_{it}^{(\lambda)} + \beta_{it}^{(\lambda)}(1-\varphi)t + \varphi y_{it-1} + u_{it}, \quad i = 1, ..., N \tag{39}$$

Using vector notation, this model implies that, under hypothesis $\varphi = 1$, the first-difference of vector y_i is given as $\Delta y_i = \beta_i^{(1)} e^{(1)} + \beta_i^{(2)} e^{(2)} + u_i$. As for model (27), this means that estimator $\hat{\Gamma}_N = \frac{1}{N}\sum_{i=1}^{N}(\Delta y_i \Delta y_i')$ will not lead to consistent estimates of the average population variance-autocovariance matrix Γ_N, due to the presence of individual effects $\beta_i^{(1)}$ and $\beta_i^{(2)}$. These imply

$$\frac{1}{N}\sum_{i=1}^{N} E(\Delta y_i \Delta y_i') = \beta_N^{(1)} e^{(1)} e^{(1)\prime} + \beta_N^{(2)} e^{(2)} e^{(2)\prime} + \Gamma_N, \tag{40}$$

where $J_1 = e^{(1)}e^{(1)\prime}$ and $J_2 = e^{(2)}e^{(2)\prime}$. The allowance of a break in incidental parameters β_i under the null hypothesis requires estimation of squared individual effects $\beta_N^{(1)}$ and $\beta_N^{(2)}$ so as to obtain consistent estimates of matrix Γ_N. To this end, we will follow an analogous procedure to that introduced in the previous subsection, based on selection matrix M. We will define two (TXT)-dimension block diagonal selection matrices $M^{(1)}$ and $M^{(2)}$, which select square individual effects $\beta_N^{(1)}$ and $\beta_N^{(2)}$, respectively. Matrix $M^{(1)}$ which has elements $m_{ts}^{(1)} = 0$ if $\gamma_{ts} \neq 0$, and $m_{ts}^{(1)} = 1$ if $\gamma_{ts} = 0$ and, thus, it selects the elements of matrix $\beta_N^{(1)} e^{(1)} e^{(1)\prime} + \beta_N^{(2)} e^{(2)} e^{(2)\prime} + \Gamma_N$ consisting only of effects $\beta_N^{(1)}$, for $t, s \leq T_0$. For t or $s > T_0$, all elements of $M^{(1)}$ are set to $m_{ts}^{(1)} = 0$. On the other hand, Matrix $M^{(2)}$ has elements $m_{ts}^{(2)} = 0$ if $\gamma_{ts} \neq 0$, and $m_{2ts} = 1$ if $\gamma_{ts} = 0$ and, thus, it selects the elements of matrix $\beta_N^{(1)} e^{(1)} e^{(1)\prime} + \beta_N^{(2)} e^{(2)} e^{(2)\prime} + \Gamma_N$ consisting only of effects $\beta_N^{(2)}$, for $t, s > T_0$. For t or $s \leq T_0$, all the elements of $M^{(2)}$ are set to $m_{ts}^{(2)} = 0$.

Based on the above definitions of selection matrices $M^{(1)}$ and $M^{(2)}$, we can obtain the following two consistent estimators of $\beta_N^{(1)}$ and $\beta_N^{(2)}$:

$$\frac{1}{tr(M^{(1)}J_1)N}\sum_{i=1}^{N}\Delta y_i' M^{(1)}\Delta y_i \xrightarrow{p} \beta_N^{(1)} \quad \text{and} \quad \frac{1}{tr(M^{(2)}J_2)N}\sum_{i=1}^{N}\Delta y_i' M^{(2)}\Delta y_i \xrightarrow{p} \beta_N^{(2)}, \tag{41}$$

respectively, since $tr(M^{(j)}\Gamma_N) = 0$ for $j = 1, 2$ and $tr(M^{(j)}J_r) = 0$ for $j, r = 1, 2$ and $j \neq r$. These estimators can be employed to obtain consistent estimates of matrix Γ_N, which are net of square individual effects $\beta_N^{(1)}$ and $\beta_N^{(2)}$. Then, a consistent estimator of the bias of the LSDV estimator $\hat{\varphi}_*^{(\lambda)}$ for model (39), defined as $\frac{b_3^{(\lambda)}}{\delta_3^{(\lambda)}}$, can be derived as

$$\frac{\hat{b}_3^{(\lambda)}}{\hat{\delta}_3^{(\lambda)}} = \frac{tr(\Psi_3^{(\lambda)}\hat{\Gamma}_N)}{\frac{1}{N}\sum_{i=1}^{N} y_{i,-1}' Q_*^{(\lambda)} y_{i,-1}}, \tag{42}$$

where $\Psi_3^{(\lambda)} = \Psi_{1*}^{(\lambda)} + \frac{tr(\Lambda' Q_*^{(\lambda)} M^{(1)})}{trace(M^{(1)} J_1)} M^{(1)} + \frac{tr(\Lambda' Q_*^{(\lambda)} M^{(2)})}{trace(M^{(2)} J_2)} M^{(2)}$. Adjusting $\hat{\varphi}_*^{(\lambda)}$ by the above estimator of its bias will lead to a panel unit root test statistic whose limiting distribution will be net of squared individual effects $\beta_N^{(1)}$ and $\beta_N^{(2)}$ under the null hypothesis. In the next theorem, we derive this distribution. This corresponds to the case of known date break. If the date is unknown, then it can be estimated, in a first step, based on the first differences of the individual series y_{it} of the panel data model under the null hypothesis $\varphi = 1$, i.e. $\Delta y_i = \beta_i^{(1)} e^{(1)} + \beta_i^{(2)} e^{(2)} + u_i$. As shown by Bai (2010), this estimator provides consistent estimates of the break point T_0, which converges at on $o(\sqrt{N})$ rate. Based on a consistent estimate of T_0, then we can apply the test statistic given by the next theorem to conduct inference about unit roots.

Theorem 9 *Let the sequence $\{y_{i,t}\}$ be generated according to model (39) and conditions (b1)-(b4) of Assumption 2 hold. Then, under the null hypothesis $\varphi = 1$ and λ known, we have*

$$Z_3^{(\lambda)} \equiv \hat{V}_3^{(\lambda)-0.5} \hat{\delta}_3^{(\lambda)} \sqrt{N} \left(\hat{\varphi}_*^{(\lambda)} - 1 - \frac{\hat{b}_3^{(\lambda)}}{\hat{\delta}_3^{(\lambda)}} \right) \xrightarrow{d} N(0,1), \tag{43}$$

as $N \to \infty$, where

$$V_3^{(\lambda)} = F_3^{(\lambda)\prime} \Theta F_3^{(\lambda)} \tag{44}$$

and $F_3^{(\lambda)} = vec(Q_*^{(\lambda)} \Lambda - \Psi_3^{(\lambda)\prime})$. The proof of the theorem is given in the appendix.

Corollary 1 *k_3 of the test in (43) is the same with k_2 (see (35)) with $\Psi_3^{(\lambda)}$ instead of $\Psi_2^{(\lambda)}$, but $\Pi_1 = \Pi_2 = 0$ which means that k_3 has the same form with k_1 with $\Psi_3^{(\lambda)}$ instead of $\Psi_1^{(\lambda)}$:*

$$k_3 = \frac{tr(\Xi' Q_*^{(\lambda)} \Gamma_N) + tr(\Lambda' Q_*^{(\lambda)} \Lambda \Gamma_N) - tr(\Psi_3^{(\lambda)} \Lambda \Gamma_N) - tr(\Lambda' \Psi_3^{(\lambda)} \Gamma_N)}{\sqrt{F_3^{(\lambda)\prime} (K_{T^2} + I_{T^2})(\Gamma_N \otimes \Gamma_N) F_3^{(\lambda)}}} \tag{45}$$

It is very interesting that the local power functions are no longer affected by the nuisance parameters of the individual effects. As test statistic $Z_2^{(\lambda)}$, the test statistic given by Theorem 9, $Z_3^{(\lambda)}$, is similar under the null hypothesis to the individual effects $\beta_i^{(1)}$ and $\beta_i^{(2)}$, due to the inclusion of broken trends in the "within" transformation matrix $Q_*^{(\lambda)}$. Due to the presence of a break under the null hypothesis $\varphi = 1$, the maximum order of serial correlation of the disturbance terms u_{it}, $p_{\max}$, allowed by the test is not given by Table 1. This is given by

$$a) \frac{T}{2} - 3, \text{when } T \text{ is even and } T_0 = \frac{T}{2} \tag{46}$$

$$b) \text{equal or less than } \min\{T_0 - 2, T - T_0 - 2\} \text{ in all other cases of } T \text{ or } T_0$$ [7]

Based on conditions of Assumption 3, it can be proved that test statistic $Z_3^{(\lambda)}$ is consistent, following analogous steps to those for the proof of the consistency of test statistic $Z_2^{(\lambda)}$. The test is also consistent, if

[7] Again the criterion of choosing $p_{\max}$ is the value of variance function $V_3^{(\lambda)}$ not to be zero. If T is even then $p_{\max}$=min$\{T_0 - 2, T - T_0 - 2\}$ always except from the case that $T_0 = \frac{T}{2}$ where it becomes $p_{\max}$=$\frac{T}{2} - 3$. Example: $T = 10$, $T_0 = 3 : p\max = \min\{T_0 - 2, T - T_0 - 2\} == \min\{1, 5\} = 1$

If $T_0 = \frac{T}{2} = 5$: then $p_{\max} = \frac{T}{2} - 3 = 2$. If we used the results from (20) we would have $p\max = \min\{T_0 - 2, T - T_0 - 2\} = \min\{3, 3\} = 3$ which does not apply.

the break point is unknown and is estimated, in the first step, based on the procedure mentioned above. In the next table, we present values of k_3 for different values of the MA parameter θ, corresponding to those of Table 2. This is done for $\lambda = 0.5$ and $T = 15$ so as the conditions of $p_{\max}$ are satisfied[8].

Table 4: Local power with break under the null

			$\theta = -0.5$		$\theta = 0$		$\theta = 0.5$	
T	p	T_0	$\Phi(z_a + k_1)$	$\hat{\Phi}(z_a + k_1)$	$\Phi(z_a + k_1)$	$\hat{\Phi}(z_a + k_1)$	$\Phi(z_a + k_1)$	$\hat{\Phi}(z_a + k_1)$
8	1	4	0.057	0.053	0.05	0.053	0.041	0.045
10	1	4	0.067	0.055	0.05	0.059	0.038	0.039
	2	4	0.062	0.059	0.05	0.057	0.037	0.036

Table 4 has very interesting results. For $p = 0$, the test has trivial power, as the other tests in the literature and it is not biased like the test that was constrained with no break under the null. For negative θ the test gains power while for positive θ the test loses power. This behavior is exhibited also in Karavias and Tzavalis (2012b). When a break under the null is not allowed θ has the opposite effect as shown in Table 3. Finally, as with $Z_1^{(\lambda)}$, the test loses power as p increases.

5 Simulation Results

In this section, we conduct a Monte Carlo study to investigate the small sample performance of the tests suggested in the previous sections. For reasons of space, in our study, we consider only the case that the break date is unknown. We consider experiments of different sample sizes of N and T, i.e. $N = \{50, 100, 200\}$ and $T = \{6, 10, 15\}$, while the fractions of sample that the break occurs are assumed to be $\lambda = \{0.25, 0.5, 0.75\}$, which facilitate the choice of the break point T_0. For all experiments, we conduct 10000 iterations. In each iteration, we assume that the data generating processes are given by models (1) and (27) respectively, where disturbance terms u_{it} follow a MA(1) process, i.e. $u_{it} = \varepsilon_{it} + \theta\varepsilon_{it-1}$, with $\varepsilon_{it} \tilde{} NIID(0, 1)$, for all i and t, and $\theta = \{-0.5, 0.0, 0.5\}$. The values of the nuisance parameters of the simulated models, namely the individual effects or the slope coefficients of individual linear trends are assumed that they are driven from the following distributions: $\alpha_i^{(1)} \sim U(-0.5, 0)$, $\alpha_i^{(2)} \sim U(0, 0.5)$, $\beta_i \sim U(0, 0.05)$, $\beta_i^{(1)} \sim U(0, 0.025)$ and $\beta_i^{(2)} \sim U(0.025, 0.05)$, where $U(\cdot)$ stands for the uniform distribution.

The small magnitude of individual effects $\alpha_i^{(j)}$ or slope coefficients $\beta_i^{(j)}$ assumed above correspond to evidence found in the empirical literature, see e.g. Hall and Mairesse (2005). This small magnitude also makes very hard the tests to distinguish the null hypothesis of unit root from its alternative of stationarity. For all simulation experiments, we assume that the order of serial correlation p is set to $p = 1$.

[8] If $T = 15$ then $p_{\max}$=min$\{T_0 - 2, T - T_0 - 2\}$ always (e.g. if $T_0 = 7$, $p_{\max} = \min\{5, 6\} = 5$).

Table 4(a): Size and power of $z_1 \equiv \min Z_1^{(\lambda)}$, for $\theta = 0.5$									
N	50	50	50	100	100	100	200	200	200
T	6	10	15	6	10	15	6	10	15
				$\lambda = 0.25$					
$\varphi = 1.00$	0.059	0.062	0.065	0.053	0.056	0.063	0.056	0.053	0.053
$\varphi = 0.95$	0.211	0.236	0.222	0.332	0.360	0.295	0.514	0.572	0.461
$\varphi = 0.90$	0.445	0.449	0.328	0.714	0.699	0.504	0.945	0.934	0.759
				$\lambda = 0.5$					
$\varphi = 1.00$	0.060	0.064	0.065	0.053	0.055	0.063	0.052	0.050	0.058
$\varphi = 0.95$	0.215	0.241	0.223	0.321	0.359	0.297	0.512	0.587	0.462
$\varphi = 0.90$	0.452	0.440	0.330	0.712	0.698	0.505	0.947	0.935	0.766
				$\lambda = 0.75$					
$\varphi = 1.00$	0.060	0.060	0.065	0.052	0.054	0.065	0.051	0.050	0.054
$\varphi = 0.95$	0.214	0.245	0.213	0.324	0.365	0.293	0.528	0.585	0.465
$\varphi = 0.90$	0.463	0.452	0.342	0.711	0.703	0.500	0.942	0.934	0.760

Table 4(b): Size and power of $z_1 \equiv \min Z_1^{(\lambda)}$, for $\theta = -0.5$									
N	50	50	50	100	100	100	200	200	200
T	6	10	15	6	10	15	6	10	15
				$\lambda = 0.25$					
$\varphi = 1.00$	0.059	0.064	0.076	0.054	0.060	0.066	0.057	0.053	0.065
$\varphi = 0.95$	0.076	0.075	0.078	0.079	0.074	0.068	0.090	0.079	0.071
$\varphi = 0.90$	0.083	0.076	0.078	0.092	0.075	0.075	0.109	0.083	0.078
				$\lambda = 0.50$					
$\varphi = 1.00$	0.057	0.064	0.072	0.056	0.061	0.066	0.051	0.053	0.069
$\varphi = 0.95$	0.082	0.070	0.073	0.074	0.072	0.068	0.087	0.079	0.071
$\varphi = 0.90$	0.083	0.073	0.079	0.093	0.079	0.072	0.116	0.082	0.073
				$\lambda = 0.75$					
$\varphi = 1.00$	0.059	0.064	0.073	0.056	0.061	0.065	0.051	0.057	0.069
$\varphi = 0.95$	0.076	0.069	0.077	0.074	0.070	0.073	0.088	0.078	0.071
$\varphi = 0.90$	0.083	0.074	0.076	0.093	0.078	0.074	0.116	0.086	0.078

This means that, for $\theta = 0$, we assume an order of serial correlation which is higher than the appropriate order. This experiment will show if our tests underperform when a higher order of serial correlation is assumed, which may happen in practice. According to the asymptotic local power functions investigated in our previous sections, this should not lead to a power underperformance of the tests, especially for model (27). The results of our Monte Carlo analysis for sequential test statistics z_1 and z_2, corresponding to models

(1) and (27), are summarized in Tables 4(a)-(c) and 5(a)-(c), respectively, for values of $\theta \in \{0.5, -0.5, 0.0\}$. These tables present values of the size and power of the tests. The size of the tests is calculated for $\varphi = 1.00$, while the power for values of $\varphi \in \{0.95, 0.9\}$. Note that, in all experiments, the power is calculated against the nominal 5% significance level of the distribution of the tests.

Table 4(c): Size and power of $z_1 \equiv \min Z_1^{(\lambda)}$, for $\theta = 0$

N	50	50	50	100	100	100	200	200	200
T	6	10	15	6	10	15	6	10	15
				$\lambda = 0.25$					
$\varphi = 1.00$	0.057	0.061	0.076	0.051	0.058	0.063	0.060	0.052	0.064
$\varphi = 0.95$	0.140	0.169	0.167	0.184	0.225	0.217	0.282	0.355	0.307
$\varphi = 0.90$	0.234	0.258	0.228	0.365	0.393	0.322	0.575	0.629	0.488
				$\lambda = 0.50$					
$\varphi = 1.00$	0.056	0.060	0.076	0.054	0.058	0.063	0.053	0.052	0.064
$\varphi = 0.95$	0.139	0.159	0.165	0.182	0.224	0.216	0.291	0.357	0.311
$\varphi = 0.90$	0.235	0.246	0.232	0.365	0.395	0.326	0.592	0.632	0.490
				$\lambda = 0.75$					
$\varphi = 1.00$	0.057	0.060	0.070	0.053	0.059	0.069	0.049	0.052	0.064
$\varphi = 0.95$	0.144	0.161	0.165	0.192	0.229	0.211	0.280	0.357	0.312
$\varphi = 0.90$	0.234	0.248	0.230	0.362	0.401	0.326	0.589	0.635	0.495

The results of the tables indicate that all three tests examined have size which is close to the nominal level 5% considered. This is true for all combinations of N and T, examined. The size performance of the tests is close to its nominal level even if the MA parameter θ takes a large negative value, i.e. $\theta = -0.5$. Note that, for this case of θ, single time series unit root tests are critically oversized (see, e.g., Schwert (1989)). The size of the tests does not also deteriorate, if a higher order of serial correlation $p = 1$ is assumed than the true order, i.e. $\theta = 0$. The size performance of the tests improves as N increases relative to T. This can be attributed to the fact that, as N increases relative to T, variance-covariance matrix Θ is more precisely estimated by estimator $\hat{\Theta}$. The above results hold independently on the fraction of the sample that the break occurs, λ.

Table 5(a): Size and power of $z_2 \equiv \min Z_2^{(\lambda)}$, for $\theta = 0.5$

N	50	50	50	100	100	100	200	200	200
T	6	10	15	6	10	15	6	10	15
				$\lambda = 0.25$					
$\varphi = 1.00$	0.058	0.084	0.093	0.051	0.071	0.072	0.048	0.060	0.068
$\varphi = 0.95$	0.056	0.085	0.117	0.055	0.083	0.106	0.056	0.081	0.118
$\varphi = 0.90$	0.057	0.111	0.207	0.056	0.129	0.259	0.058	0.153	0.358
				$\lambda = 0.5$					
$\varphi = 1.00$	0.060	0.078	0.093	0.054	0.072	0.072	0.051	0.060	0.068
$\varphi = 0.95$	0.055	0.090	0.117	0.054	0.084	0.106	0.054	0.081	0.120
$\varphi = 0.90$	0.062	0.120	0.205	0.058	0.125	0.254	0.057	0.154	0.350
				$\lambda = 0.75$					
$\varphi = 1.00$	0.033	0.077	0.093	0.052	0.069	0.071	0.052	0.060	0.067
$\varphi = 0.95$	0.058	0.091	0.117	0.054	0.083	0.105	0.053	0.081	0.116
$\varphi = 0.90$	0.059	0.118	0.205	0.057	0.127	0.256	0.052	0.152	0.349

Table 5(b): Size and power of $z_2 \equiv \min Z_2^{(\lambda)}$, for $\theta = -0.5$

N	50	50	50	100	100	100	200	200	200
T	6	10	15	6	10	15	6	10	15
				$\lambda = 0.25$					
$\varphi = 1.00$	0.055	0.061	0.073	0.051	0.057	0.066	0.050	0.055	0.064
$\varphi = 0.95$	0.057	0.075	0.102	0.052	0.080	0.109	0.052	0.090	0.135
$\varphi = 0.90$	0.053	0.092	0.131	0.050	0.113	0.162	0.051	0.148	0.220
				$\lambda = 0.50$					
$\varphi = 1.00$	0.053	0.061	0.076	0.051	0.058	0.066	0.050	0.055	0.064
$\varphi = 0.95$	0.050	0.073	0.104	0.052	0.080	0.107	0.052	0.086	0.133
$\varphi = 0.90$	0.055	0.093	0.131	0.051	0.105	0.153	0.051	0.131	0.195
				$\lambda = 0.75$					
$\varphi = 1.00$	0.053	0.065	0.076	0.051	0.063	0.066	0.050	0.058	0.064
$\varphi = 0.95$	0.052	0.074	0.103	0.053	0.085	0.108	0.051	0.086	0.132
$\varphi = 0.90$	0.055	0.088	0.132	0.050	0.101	0.154	0.052	0.136	0.202

Table 5(c): Size and power of $z_2 \equiv \min Z_2^{(\lambda)}$, for $\theta = 0$									
N	50	50	50	100	100	100	200	200	200
T	6	10	15	6	10	15	6	10	15
$\lambda = 0.25$									
$\varphi = 1.00$	0.056	0.076	0.089	0.051	0.066	0.075	0.051	0.064	0.070
$\varphi = 0.95$	0.050	0.079	0.117	0.051	0.079	0.123	0.049	0.083	0.128
$\varphi = 0.90$	0.054	0.103	0.184	0.053	0.119	0.231	0.055	0.141	0.322
$\lambda = 0.50$									
$\varphi = 1.00$	0.055	0.068	0.089	0.052	0.064	0.075	0.051	0.063	0.068
$\varphi = 0.95$	0.051	0.082	0.118	0.049	0.078	0.122	0.050	0.086	0.125
$\varphi = 0.90$	0.055	0.107	0.181	0.058	0.111	0.226	0.050	0.145	0.309
$\lambda = 0.75$									
$\varphi = 1.00$	0.053	0.076	0.091	0.055	0.061	0.075	0.051	0.064	0.069
$\varphi = 0.95$	0.051	0.078	0.114	0.052	0.078	0.107	0.052	0.082	0.123
$\varphi = 0.90$	0.055	0.107	0.189	0.050	0.110	0.229	0.059	0.137	0.306

Regarding the power of the tests, the results of the tables indicate that, as was expected, the tests have better power performance for the case of model (1), where there are only individual effects under the alternative. For model (27) where linear trends are considered, the power of the tests reduces. However, the test is not biased and, consistently with the theory, has power which increases as the value of φ moves away from unity. The power performance of all test statistics is not affected by the fraction of the sample that the break occurs, λ, see also Karavias and Tzavalis (2012a). However, it depends on the value of θ. When θ is negative (i.e., $\theta = -0.5$), the power performance of all three tests deteriorates. Finally, note that the power of the tests tends to increase as N increases relative to T.

6 Conclusion

This paper suggests panel unit root tests which allow for a common structural break in the individual effects or linear trends of dynamic panel data models. Common breaks in panel data models can arise in cases of a credit crunch, an oil price shock or a change in tax policy (see e.g. Bai (2010)). The suggested tests assume that the time-dimension of the panel T is fixed (or finite), while the cross-section N grows large. Thus, they are appropriate for short panels applications, where T is smaller than N. They are based on the corrected for its inconsistency (asymptotic bias) least squares dummy variable (LSDV) estimator of the autoregressive coefficient of dynamic panel data models (see Harris and Tzavalis (1999)), and they allow for serial correlation in the disturbance terms. This estimator provides unit root tests which have the very useful property of being invariant (similar) under the null hypothesis of a unit root in the autoregressive component of panel data models to the initial conditions of the panel or the individual effects. This property does not restrict application of the tests to panel data where conditions of mean or covariance stationarity

of the initial conditions or individual effects are required.

The paper derives the limiting distributions of the tests and study their asymptotic (over N) local power. When the break is unknown, it shows that the limiting distribution of the tests is the minimum of a fixed number of correlated normals. This distribution is given as a mixture of normals. It can be analytically obtained based on recent results of Arellano-Valle and Genton (2008), for absolutely continuous dependent variables. Knowledge of the analytic form of the limiting distribution of the tests considerably facilitates calculation of critical values for the implementation of the tests in practice. To investigate the power of the tests under a sequence of local alternatives, the paper considers a MA(1) process of the disturbance terms, which is often assumed for many economic series. The paper shows that the tests can have non trivial asymptotic local power in a neighborhood of unity which shrinks at $\sqrt{N}$ rate, even if individual (incidental) linear trends are considered. Positive values of serial correlation increase power, while negative values decrease power, in the cases of individual effects only and incidental trends with no break under the null. When a break is allowed under the null hypothesis power decreases with positive values of serial correlation and increases with negative. This is in accordance to the finding of Karavias and Tzavalis (2012b). The date of the break interacts with serial correlation in terms of power. In the cases of individual effects only and incidental trends with break under the null for negative values of serial correlation the tests have greater power when the break is towards the end of the sample, while for positive values of serial correlation the tests have more power when the break is in the middle of the sample. When no breaks are allowed under the null the test has more power when the break is in the middle of the sample irrespective to the sign of serial correlation. The variability of the slope coefficients only affects the test when no break under the null is allowed. Greater the variability means more power. Assuming higher order of serial correlation creates loss of power for tests in cases of individual effects only and incidental trends with break under the null but a great increase in power for case with incidental trends and no break under the null when the break is in the middle of the sample.

Finally, the paper examines the small sample size and power performance of the tests by conducting a Monte Carlo study. This is done for the case that the break is of an unknown date. The results of this exercise indicate that our tests have the correct nominal size and power which is bigger than their size. As was expected, the power of the tests is higher for the dynamic panel data model which consider individual effects rather than for the model which also allows for individual linear trends. For all cases, the power is found to increase if N increases faster than T, due to the fixed-T assumption of the tests.

7 Appendix

In this appendix, we provide proofs of the theorems or any other theoretical results presented in the main text of the paper.

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 LSDV estimator $\hat{\varphi}^{(\lambda)}$ is inconsistent, as $N \to \infty$. Then, will

construct a normalized statistic based on $\hat{\varphi}^{(\lambda)}$ corrected for its inconsistency (bias) and derive its limiting distribution under the null hypothesis of $\varphi = 1$, as $N \to \infty$.

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

$$y_{i,-1} = ey_{i0} + \Lambda u_i, \tag{47}$$

where the matrix Λ is is a (TXT) matrix defined as $\Lambda_{r,c} = 1$, if $r > c$ and 0 otherwise.

Premultiplying (47) with matrix $Q^{(\lambda)}$ yields

$$Q^{(\lambda)} y_{i,-1} = Q^{(\lambda)} \Lambda u_i, \tag{48}$$

since $Q^{(\lambda)} e = (0, 0, ..., 0)'$. Substituting (48) into (2) yields

$$\hat{\varphi}^{(\lambda)} - 1 = \frac{\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1} Q^{(\lambda)} u_i}{\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1} Q^{(\lambda)} y_{i,-1}} = \frac{\frac{1}{N}\sum_{i=1}^{N} u'_i \Lambda' Q^{(\lambda)} u_i}{\frac{1}{N}\sum_{i=1}^{N} u'_i \Lambda' Q^{(\lambda)} \Lambda u_i}. \tag{49}$$

By Kitchin's Weak Law of Large Numbers (KWLLN), we have

$$\frac{1}{N}\sum_{i=1}^{N} u'_i \Lambda' Q^{(\lambda)} u_i \xrightarrow{p} b^{(\lambda)} = \sigma_u^2 tr(\Lambda' Q^{(\lambda)}) \text{ and } \frac{1}{N}\sum_{i=1}^{N} u'_i \Lambda' Q^{(\lambda)} \Lambda u_i \xrightarrow{p} \delta^{(\lambda)} = \sigma_u^2 tr(\Lambda' Q^{(\lambda)} \Lambda), \tag{50}$$

where "$\xrightarrow{p}$" signifies convergence in probability. Using the last results, the yet non standardized statistic $Z^{(\lambda)}$ can be written by (49) as

$$\begin{aligned} \sqrt{N}\hat{\delta}^{(\lambda)} \left(\hat{\varphi}^{(\lambda)} - 1 - \frac{\hat{b}^{(\lambda)}}{\hat{\delta}^{(\lambda)}} \right) &= \sqrt{N}\hat{\delta}^{(\lambda)} \left(\frac{\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1} Q^{(\lambda)} u_i}{\hat{\delta}^{(\lambda)}} - \frac{\hat{\sigma}_u^2 tr(\Lambda' Q^{(\lambda)})}{\hat{\delta}^{(\lambda)}} \right) \\ &= \sqrt{N} \left(\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1} Q^{(\lambda)} u_i - tr(\Lambda' Q^{(\lambda)}) \frac{\sum_{i=1}^{N} \Delta y'_i \Psi^{(\lambda)} \Delta y_i}{N tr(\Psi^{(\lambda)})} \right). \end{aligned} \tag{51}$$

Since, under the null hypothesis $\varphi = 1$, we have $u_i = \Delta y_i$, the last relationship can be written as follows:

$$\begin{aligned} & \sqrt{N}\hat{\delta}^{(\lambda)} \left(\hat{\varphi}^{(\lambda)} - 1 - \frac{\hat{b}^{(\lambda)}}{\hat{\delta}^{(\lambda)}} \right) \\ = \; & \sqrt{N} \left(\frac{1}{N}\sum_{i=1}^{N} u'_i \Lambda' Q^{(\lambda)} u_i - \frac{tr(\Lambda' Q^{(\lambda)})}{tr(\Psi^{(\lambda)})} \frac{1}{N}\sum_{i=1}^{N} u'_i \Psi^{(\lambda)} u_i \right) \\ = \; & \frac{1}{\sqrt{N}}\sum_{i=1}^{N} u'_i (\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)}) u_i = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} tr\left[(\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)}) u_i u'_i \right] \\ = \; & \frac{1}{\sqrt{N}}\sum_{i=1}^{N} W_i^{(\lambda)}, \end{aligned} \tag{52}$$

where $W_i^{(\lambda)}$ constitute random variables with mean

$$\begin{aligned} E(W_i^{(\lambda)}) &= E[u_i'(\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)})u_i] = tr[(\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)})E(u_i u_i')] \\ &= \sigma_u^2 tr(\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)}) = 0, \text{ for all } i, \end{aligned}$$

since $tr(\Lambda' Q^{(\lambda)}) = tr(\Psi^{(\lambda)})$ (or $tr(\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)}) = 0$) and variance

$$\begin{aligned} Var(W_i^{(\lambda)}) &= Var(u_i'(\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)})u_i) = Var[F^{(\lambda)\prime} vec(u_i u_i')] = \\ &= F^{(\lambda)} Var[vec(u_i u_i')] F^{(\lambda)\prime}, \text{ for all } i. \end{aligned}$$

The results of Theorem 1 follows by applying Lindeberg-Levy central limit theorem (CLT) to the sequence of IID random variables $W_i^{(\lambda)}$. Following standard linear algebra results (see e.g. Schott(1997), variance $Var[vec(u_i u_i')]$ can be analytically written as $Var[vec(u_i u_i')] = Var(u_i \otimes u_i) = \sigma_u^4(I_{T^2} + K_{T^2})$, where $\otimes$ denotes the Kroenecker product.

Proof of Theorem 2: Assume that the break point T_0 is known. Define vector $w = (1, \varphi, \varphi^2, ..., \varphi^{T-1})'$ and matrix

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

Under null hypothesis $\varphi = 1$, we have $\Omega = \Lambda$. Based on the above definitions of w and Ω, vector $y_{i,-1}$ can be writen as

$$y_{i,-1} = w y_{i0} + \Omega X^{(\lambda)} \gamma_i^{(\lambda)} + \Omega u_i, \tag{53}$$

where $d_i^{(\lambda)} = (a_i^{(1)}(1-\varphi), a_i^{(2)}(1-\varphi))'$. Using last expression of $y_{i,-1}$, test statistic $Z^{(\lambda)}$ can be written

under the alternative hypothesis $\varphi < 1$ as follows:

$$\begin{aligned}
Z^{(\lambda)} &= \sqrt{N}\widehat{V}^{(\lambda)-1/2}\hat{\delta}^{(\lambda)}\left(\hat{\varphi}^{(\lambda)} - 1 - \frac{\hat{b}^{(\lambda)}}{\hat{\delta}^{(\lambda)}}\right) \\
&= \sqrt{N}\hat{V}^{(\lambda)-1/2}\hat{\delta}^{(\lambda)}\left(\varphi + \frac{\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1}Q^{(\lambda)}u_i}{\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1}Q^{(\lambda)}y_{i,-1}} - 1 - \frac{\hat{\sigma}_u^2 tr(\Lambda' Q^{(\lambda)})}{\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1}Q^{(\lambda)}y_{i,-1}}\right) \\
&= \sqrt{N}\hat{V}^{(\lambda)-1/2}\hat{\delta}^{(\lambda)}(\varphi - 1) + \sqrt{N}\hat{V}^{(\lambda)-1/2}\left(\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1}Q^{(\lambda)}u_i - \hat{\sigma}_u^2 tr(\Lambda' Q^{(\lambda)})\right) \qquad (54) \\
&= \underbrace{\left\{\sqrt{N}\hat{V}^{(\lambda)-1/2}\hat{\delta}^{(\lambda)}(\varphi - 1)\right\}}_{(I)} + \underbrace{\left\{\hat{V}^{(\lambda)-1/2}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(y'_{i,-1}Q^{(\lambda)}u_i - \Delta y'_i \Psi^{(\lambda)}\Delta y_i)\right\}}_{(II)}.
\end{aligned}$$

Next, we will show that summand (I) diverges to $-\infty$ and summand (II) is bounded in probability. These two results imply that, as $N \to \infty$, test statistic $Z^{(\lambda)}$ converges to $-\infty$, which proves its consistency. To prove the above results, we will use the following identities:

$$u_i = y_i - \varphi y_{i-1} - X^{(\lambda)}d_i^{(\lambda)} \qquad (55)$$

and

$$\Delta y_i = u_i + (\varphi - 1)y_{i-1} + X^{(\lambda)}d_i^{(\lambda)}, \qquad (56)$$

which hold under the alternative hypothesis $\varphi < 1$.

To prove that summand (I), defined by (54), diverges to $-\infty$, it is sufficient to show that $p\lim \hat{\delta}^{(\lambda)}$ is $O_p(1)$ and positive, and $p\lim \hat{\sigma}_u^2 = O_p(1)$ and nonzero. The last result implies that variance function $\hat{V}^{(\lambda)} = \hat{\sigma}_u^4 F^{(\lambda)\prime}(K_{T^2} + I_{T^2})F^{(\lambda)}$ is bounded in probability. Using equations (53), (55) and (56), it can be seen that $\hat{\delta}^{(\lambda)}$ is $O_p(1)$ as follows:

$$\begin{aligned}
\hat{\delta}^{(\lambda)} &= \frac{1}{N}\sum_{i=1}^{N} y'_{i,-1}Q^{(\lambda)}y_{i,-1} = \frac{1}{N}\sum_{i=1}^{N}(wy_{i0} + \Omega X^{(\lambda)}d_i^{(\lambda)} + \Omega u_i)'Q^{(\lambda)}(wy_{i0} + \Omega X^{(\lambda)}d_i^{(\lambda)} + \Omega u_i) \\
&= \frac{1}{N}\sum_{i=1}^{N}(y_{i0}^2 w'Q^{(\lambda)}w + y_{i0}w'Q^{(\lambda)}\Omega X^{(\lambda)}d_i^{(\lambda)} + y_{i0}wQ^{(\lambda)}\Omega u_i + ... + u_i'\Omega' Q^{(\lambda)}\Omega u_i \\
&\xrightarrow{p} E(y_{i0}^2)w'Q^{(\lambda)}w + tr(X^{(\lambda)\prime}\Omega' Q^{(\lambda)}\Omega X^{(\lambda)}\Sigma_d) + \sigma_u^2 tr(\Omega' Q^{(\lambda)}\Omega) = O_p(1),
\end{aligned}$$

where $\Sigma_d = E(d_i^{(\lambda)}d_i^{(\lambda)\prime})$. The last result holds by condition a3 of Assumption 1. All quantities involved in the above limit are positive because they are either variances or quadratic forms. Based on condition a3 of

Assumption 1, we can also show that the following result also holds:

$$\begin{aligned}\hat{\sigma}_u^2 &= \frac{1}{tr(\Psi^{(\lambda)})}\frac{1}{N}\sum_{i=1}^{N}\Delta y_i'\Psi^{(\lambda)}\Delta y_i \\ &= \frac{1}{tr(\Psi^{(\lambda)})}\frac{1}{N}\sum_{i=1}^{N}(u_i+(\varphi-1)y_{i-1}+X^{(\lambda)}d_i^{(\lambda)})'\Psi^{(\lambda)}(u_i+(\varphi-1)y_{i-1}+X^{(\lambda)}d_i^{(\lambda)}) \\ &= O_p(1).\end{aligned}$$

This limit is a nonzero quantity because at least $\sigma_u^2 > 0$. The remaining terms entered into this limit are zero or positive quantities.

To prove that summand (II) is bounded in probability note that, by Assumption 1, we have

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(y_{i,-1}'Q^{(\lambda)}u_i - \Delta y_i'\Psi^{(\lambda)}\Delta y_i) = O_p(1).$$

See also proof of Theorem 1.

Proof of Theorem 3: See proof of the more general case given by Theorem 6.

Proof of Theorem 4: The proof of this theorem follows as an extension of Theorem 1, by applying the continuous mapping theorem to the joint limiting distribution of standardized test statistic $Z^{(\lambda)}$, for all $\lambda \in I$. The elements of the covariance matrix between random variables $Z^{(\lambda)}$ and $Z^{(\mu)}$, for all $\lambda \neq \mu$, can be derived by writing

$$\begin{aligned}Z^{(\lambda)}Z^{(\mu)} &= \sqrt{N}\left(\frac{\hat{\delta}^{(\lambda)}}{\sqrt{V^{(\lambda)}}}\right)\left(\hat{\varphi}-1-\frac{\hat{b}^{(\lambda)}}{\hat{\delta}^{(\lambda)}}\right)\sqrt{N}\left(\frac{\hat{\delta}^{(\mu)}}{\sqrt{V^{(\mu)}}}\right)\left(\hat{\varphi}-1-\frac{\hat{b}^{(\mu)}}{\hat{\delta}^{(\mu)}}\right) \\ &= \frac{\hat{\delta}^{(\lambda)}\hat{\delta}^{(\mu)}}{\sqrt{V^{(\lambda)}}\sqrt{V^{(\mu)}}}N\frac{(\frac{1}{N}\sum_{i=1}^{N}W_i^{(\lambda)})(\frac{1}{N}\sum_{i=1}^{N}W_i^{(\mu)})}{\hat{\delta}^{(\lambda)}\hat{\delta}^{(\mu)}} \\ &= \frac{1}{\sqrt{V^{(\lambda)}}\sqrt{V^{(\mu)}}}\frac{1}{N}\sum_{i=1}^{N}W_i^{(\lambda)}\sum_{i=1}^{N}W_i^{(\mu)} \qquad (57)\end{aligned}$$

By the definition of $W_i^{(\lambda)}$ (see (52)) and the assumption of cross-section independence between $W_i^{(\lambda)}$ and $W_j^{(m)}$, for $i \neq j$, we have $E(W_i^{(\lambda)}W_j^{(\mu)}) = 0$, for $i \neq j$. Based on this result, we can show that

$$\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}W_i^{(\lambda)}\sum_{i=1}^{N}W_i^{(\mu)} = \operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}W_i^{(\lambda)}W_i^{(\mu)} = E(W_i^{(\lambda)}W_i^{(\mu)}).$$

$E(W_i^{(\lambda)}W_i^{(\mu)})$ can be analytically derived as

$$\begin{aligned} E(W_i^{(\lambda)}W_i^{(\mu)}) &= E[u_i'(\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)})u_i u_i'(\Lambda' Q^{(\mu)} - \Psi^{(\mu)})u_i] \\ &= E[F^{(\lambda)\prime} vec(u_i u_i') vec(u_i u_i')' F^{(\mu)}] \\ &= F^{(\lambda)\prime} E[vec(u_i u_i') vec(u_i u_i')'] F^{(\mu)} \\ &= F^{(\lambda)\prime} E[vec(u_i u_i') vec(u_i u_i')'] F^{(\mu)}, \quad \text{or} \end{aligned} \tag{58}$$

$$E(W_i^{(\lambda)}W_i^{(\mu)}) = \sigma_u^4 F^{(\lambda)\prime}[(I_{T^2} + K_{T^2}) + vec(I_T)vec(I_T)']F^{(\mu)}, \tag{59}$$

using the following result:

$$\begin{aligned} E[vec(u_i u_i')vec(u_i u_i')'] &= Var(u_i \otimes u_i) + E(vec(u_i u_i'))E(vec(u_i u_i'))' \\ &= \sigma_u^4[(I_{T^2} + K_{T^2}) + vec(I_T)vec(I_T)']. \end{aligned}$$

Based on (59), it can be found that the probability limit of (57) is given as

$$\begin{aligned} & E(Z^{(\lambda)}Z^{(\mu)}) \\ = & \frac{F^{(\lambda)\prime}\sigma_u^4[(I_{T^2} + K_{T^2}) + vec(I_T)vec(I_T)']F^{(\mu)}}{\sqrt{F^{(\lambda)\prime}\sigma_u^4[(I_{T^2} + K_{T^2} + vec(I_T)vec(I_T)']F^{(\lambda)}}\sqrt{F^{(\mu)\prime}\sigma_u^4[(I_{T^2} + K_{T^2}) + vec(I_T)vec(I_T)']F^{(\mu)}}} \\ = & \frac{F^{(\lambda)\prime}(I_{T^2} + K_{T^2})F^{(\mu)}}{\sqrt{F^{(\lambda)\prime}(I_{T^2} + K_{T^2})F^{(\lambda)}}\sqrt{F^{(\mu)\prime}(I_{T^2} + K_{T^2})F^{(\mu)}}}, \end{aligned}$$

where the result of the last row follows from $F^{(\lambda)\prime}vec(I_T)vec(I_T)' = 0$.

Proof of Theorem 5: The theorem can be proved following analogous steps to those for the proof of Theorem 1 and using the following results:

$$\frac{1}{N}\sum_{i=1}^{N} u_i'\Lambda' Q^{(\lambda)} u_i \xrightarrow{p} tr(\Lambda' Q^{(\lambda)}\Gamma_N) \text{ and } \frac{1}{N}\sum_{i=1}^{N} u_i'\Lambda' Q^{(\lambda)}\Lambda u_i \to^p tr(\Lambda' Q^{(\lambda)}\Lambda\Gamma_N).$$

Based on the definition of matrix $\Psi_1^{(\lambda)}$ and conditions $(b1)$ and $(b2)$ of Assumption 2, it can be easily seen that

$$E(W_i^{(\lambda)}) = tr((\Lambda' Q^{(\lambda)} - \Psi_1^{(\lambda)})\Gamma_N) = 0, \text{ for all } i.$$

Proof of Theorem 6: Write the non standardized version of test statistic $Z_1^{(\lambda)}$ under the sequence of

local alternative hypotheses $\varphi_N = 1 - \frac{c}{\sqrt{N}}$ as follows:

$$\begin{aligned}
& \hat{\delta}_1^{(\lambda)} \sqrt{N} \left(\hat{\varphi}^{(\lambda)} - \frac{\hat{b}_1^{(\lambda)}}{\hat{\delta}_1^{(\lambda)}} - \varphi_N \right) \\
= & \sqrt{N} \hat{\delta}_1^{(\lambda)} \left(\frac{\frac{1}{N} \sum_{i=1}^N y'_{i,-1} Q^{(\lambda)} u_i - tr(\Psi_1^{(\lambda)} \hat{\Gamma}_N)}{\hat{\delta}_1^{(\lambda)}} \right) \\
= & \sqrt{N} \left(\frac{1}{N} \sum_{i=1}^N y'_{i,-1} Q^{(\lambda)} u_i - \frac{1}{N} \sum_{i=1}^N tr(\Psi_1^{(\lambda)} \Delta y_i \Delta y'_i) \right) \\
= & \underbrace{\frac{1}{\sqrt{N}} \sum_{i=1}^N y'_{i,-1} Q^{(\lambda)} u_i}_{(A)} - \underbrace{\frac{1}{\sqrt{N}} \sum_{i=1}^N \Delta y'_i \Psi_1^{(\lambda)} \Delta y_i}_{(B)}. \qquad (60)
\end{aligned}$$

The proof of the theorem follows immediately based on the following two results:

$$(A): \quad \frac{1}{\sqrt{N}} \sum_{i=1}^N y'_{i,-1} Q^{(\lambda)} u_i \xrightarrow{d} N \left(-ctr(\Xi' Q^{(\lambda)} \Gamma_N) + tr(\Lambda' Q^{(\lambda)} \Gamma_N),\ 2tr((\Upsilon^{(\lambda)} \Gamma_N)^2) \right), \qquad (61)$$

where $\Xi = \frac{d\Omega}{d\varphi}|_{\varphi=1}$ and $\Upsilon^{(\lambda)} = \frac{1}{2}(\Lambda' Q^{(\lambda)} + Q^{(\lambda)} \Lambda)$, and

$$(B): \quad -\frac{1}{\sqrt{N}} \sum_{i=1}^N \Delta y'_i \Psi_1^{(\lambda)} \Delta y_i \xrightarrow{d} N \left(c[tr(\Psi_1^{(\lambda)} \Lambda \Gamma_N) + tr(\Lambda' \Psi_1^{(\lambda)} \Gamma_N)] - tr(\Psi_1^{(\lambda)} \Gamma_N),\ 2tr((\Upsilon_\psi^{(\lambda)} \Gamma_N)^2) \right), \qquad (62)$$

where $Y_\psi^{(\lambda)} = \frac{1}{2}(\Psi_1^{(\lambda)\prime} + \Psi_1^{(\lambda)})$. These results can be proved based on the following Taylor's expansions of w and Ω around $\varphi = 1$:

$$\Omega = \Lambda + \Xi(\varphi_N - 1) + o_p(1) \quad \text{and} \quad w = e + \varpi(\varphi_N - 1) + o_P(1), \qquad (63)$$

where $\Xi = \frac{d\Omega}{d\varphi}|_{\varphi=1}$ (defined above) and $\varpi = \frac{dw}{d\varphi}|_{\varphi=1}$.

In particular, to prove result (A), given by (61), write $\frac{1}{\sqrt{N}} \sum_{i=1}^N y'_{i,-1} Q^{(\lambda)} u_i$ as follows:

$$\begin{aligned}
\frac{1}{\sqrt{N}} \sum_{i=1}^N y'_{i,-1} Q^{(\lambda)} u_i &= \frac{1}{\sqrt{N}} \sum_{i=1}^N (w' y_{i0} + d_i^{(\lambda)\prime} X^{(\lambda)\prime} \Omega' + u'_i \Omega') Q^{(\lambda)} u_i \qquad (64) \\
&= \frac{1}{\sqrt{N}} \sum_{i=1}^N (y_{i0} w' Q^{(\lambda)} u_i + d_i^{(\lambda)\prime} X^{(\lambda)\prime} \Omega' + u'_i \Omega' Q^{(\lambda)} u_i).
\end{aligned}$$

Based on (63) and conditions (b1)-(b3) of Assumption 2, it can be proved that the three summands involved in the last relationship (i.e. (64)) converge to the following:

$$\begin{aligned}
\frac{1}{\sqrt{N}} \sum_{i=1}^N y_{i0} w' Q^{(\lambda)} u_i &= \frac{1}{\sqrt{N}} \sum_{i=1}^N (y_{i0} e' Q^{(\lambda)} u_i + y_{i0} w (\varphi_N - 1) Q^{(\lambda)} u_i) + o_p(1) \\
&= -\frac{1}{N} c \sum_{i=1}^N w Q^{(\lambda)} u_i y_{i0} + o_p(1) \xrightarrow{p} 0,
\end{aligned}$$

$$\begin{aligned}\frac{1}{\sqrt{N}}\sum_{i=1}^{N} d_i^{(\lambda)\prime} X^{(\lambda)\prime}\Omega' Q^{(\lambda)} u_i &= (1-\varphi)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(a_i^{(1)}, a_i^{(2)})X^{(\lambda)\prime}(\Lambda + \Xi(\varphi_N - 1))' Q^{(\lambda)} u_i + o_p(1) \\ &= \frac{c}{N}\sum_{i=1}^{N}(a_i^{(1)}, a_i^{(2)})X^{(\lambda)\prime}\Lambda' Q^{(\lambda)} u_i + \frac{c^2}{N^{3/2}}\sum_{i=1}^{N}(a_i^{(1)}, a_i^{(2)})X^{(\lambda)\prime}\Xi' Q^{(\lambda)} u_i + o_p(1) \\ &\xrightarrow{p} 0,\end{aligned}$$

and

$$\begin{aligned}\frac{1}{\sqrt{N}}\sum_{i=1}^{N} u_i'\Omega' Q^{(\lambda)} u_i &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N} u_i'(\Lambda' + \Xi'(\varphi_N - 1))Q^{(\lambda)} u_i + o_p(1) \\ &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N} u_i'\Lambda' Q^{(\lambda)} u_i - \frac{c}{N}\sum_{i=1}^{N} u_i'\Xi' Q^{(\lambda)} u_i + o_p(1) \\ &\xrightarrow{d} N\left(-ctr(\Xi' Q^{(\lambda)}\Gamma_N) + tr(\Lambda' Q^{(\lambda)}\Gamma_N),\ 2tr((K^{(\lambda)}\Gamma_N)^2)\right),\end{aligned}$$

respectively. The last result holds because we have $-\frac{c}{N}\sum_{i=1}^{N} u_i'\Xi' Q^{(\lambda)} u_i \xrightarrow{p} -ctr(\Xi' Q^{(\lambda)}\Gamma_N)$ and $\frac{1}{\sqrt{N}}\sum_{i=1}^{N} u_i'\Lambda' Q^{(\lambda)} u_i \xrightarrow{d} N\left(tr(\Lambda' Q^{(\lambda)}\Gamma_N), 2tr((\Upsilon^{(\lambda)}\Gamma_N)^2)\right)$, using results on quadratic forms of normally distributed variables (see e.g. Schott (1997)). Following analogous steps to the above, we can prove result (B).

Results (A) and (B), proved above, imply that

$$\hat{\delta}_1^{(\lambda)}\sqrt{N}\left(\hat{\varphi}^{(\lambda)} - \frac{\hat{b}_1^{(\lambda)}}{\hat{\delta}_1^{(\lambda)}} - \varphi_N\right) \xrightarrow{d} N\left(-c[tr(\Xi' Q^{(\lambda)}\Gamma_N) - tr(\Psi_1^{(\lambda)}\Lambda\Gamma_N) - tr(\Lambda'\Psi_1^{(\lambda)}\Gamma_N)],\ V_1^{(\lambda)}\right). \qquad (65)$$

The results of Theorem 6 can be proved based on (65) and noticing that $V_1^{(\lambda)}$ and $\hat{\delta}_1^{(\lambda)}$ converge to the same quantities under the null and the sequence of local alternative hypotheses which are free of parameter c. The last enables us to write (65) as follows:

$$\hat{\delta}_1^{(\lambda)}\sqrt{N}\left(\hat{\varphi}^{(\lambda)} - \frac{\hat{b}_1^{(\lambda)}}{\hat{\delta}_1^{(\lambda)}} - 1 + \frac{c}{\sqrt{N}}\right) \xrightarrow{d} N\left(-c[tr(\Xi' Q^{(\lambda)}\Gamma_N) - tr(\Psi_1^{(\lambda)}\Lambda\Gamma_N) - tr(\Lambda'\Psi_1^{(\lambda)}\Gamma_N)],\ V_1^{(\lambda)}\right) \text{ or}$$

$$\hat{\delta}_1^{(\lambda)}\sqrt{N}(\hat{\varphi}^{(\lambda)} - \frac{\hat{b}_1^{(\lambda)}}{\hat{\delta}_1^{(\lambda)}} - 1) + (c\hat{\delta}_1^{(\lambda)}) \xrightarrow{d} N\left(-c[tr(\Xi' Q^{(\lambda)}\Gamma_N) - tr(\Psi_1^{(\lambda)}\Lambda\Gamma_N) - tr(\Lambda'\Psi_1^{(\lambda)}\Gamma_N)],\ V_1^{(\lambda)}\right) \text{ or}$$

$$\hat{\delta}_1^{(\lambda)}\sqrt{N}(\hat{\varphi}^{(\lambda)} - \frac{\hat{b}_1^{(\lambda)}}{\hat{\delta}_1^{(\lambda)}} - 1) \xrightarrow{d} N\left(-c[tr(\Xi' Q^{(\lambda)}\Gamma_N) - tr(\Psi_1^{(\lambda)}\Lambda\Gamma_N) - tr(\Lambda'\Psi_1^{(\lambda)}\Gamma_N) - tr(\Lambda' Q^{(\lambda)}\Lambda\Gamma_N)],\ V_1^{(\lambda)}\right) \text{ or}$$

$$V_1^{(\lambda)-1/2}\hat{\delta}_1^{(\lambda)}\sqrt{N}\left(\hat{\varphi}^{(\lambda)} - \frac{\hat{b}_1^{(\lambda)}}{\hat{\delta}_1^{(\lambda)}} - 1\right)$$

$$\xrightarrow{d} N\left(\frac{-c[tr(\Xi' Q^{(\lambda)}\Gamma_N) - tr(\Psi_1^{(\lambda)}\Lambda\Gamma_N) - tr(\Lambda'\Psi_1^{(\lambda)}\Gamma_N) - tr(\Lambda' Q^{(\lambda)}\Lambda\Gamma_N)]}{V_1^{(\lambda)1/2}},\ 1\right),$$

which implies

$$Z_1^{(\lambda)} \equiv V_1^{(\lambda)-1/2}\hat{\delta}_1^{(\lambda)}\sqrt{N}\left(\hat{\varphi}^{(\lambda)} - \frac{\hat{b}_1^{(\lambda)}}{\hat{\delta}_1^{(\lambda)}} - 1\right) \xrightarrow{d} N\left(-c\frac{tr(\Xi' Q^{(\lambda)}\Gamma_N) + tr(\Lambda' Q^{(\lambda)}\Lambda\Gamma_N) - tr(\Psi_1^{(\lambda)}\Lambda\Gamma_N) - tr(\Lambda'\Psi_1^{(\lambda)}\Gamma_N)}{\sqrt{F^{(\lambda)\prime}(K_{T^2} + I_{T^2})(\Gamma_N \otimes \Gamma_N)F^{(\lambda)}}}, 1\right), \tag{66}$$

since $V_1^{(\lambda)} = F^{(\lambda)\prime}(K_{T^2} + I_{T^2})(\Gamma_N \otimes \Gamma_N)F^{(\lambda)}$ under the assumptions of Theorem 6.

The limiting distribution given by Theorem 3 can be obtained from (66) under the following assumptions: $\Gamma_N = \sigma_u^2 I_T$ and $\Psi_1^{(\lambda)} = \Psi^{(\lambda)}$, which hold under the conditions of Assumption 1. Then, the limiting distribution of test statistic $Z^{(\lambda)}$ under the sequence of local alternatives φ_N can be written as

$$Z_1^{(\lambda)} \equiv V_1^{(\lambda)-1/2}\hat{\delta}_1^{(\lambda)}\sqrt{N}\left(\hat{\varphi}^{(\lambda)} - \frac{\widehat{\sigma}_u^2 tr(\Lambda' Q^{(\lambda)})}{\hat{\delta}_1^{(\lambda)}} - 1\right) \xrightarrow{d} N\left(-c\frac{tr(\Xi' Q^{(\lambda)}) + tr(\Lambda' Q^{(\lambda)}\Lambda)}{\sqrt{F^{(\lambda)\prime}(K_{T^2} + I_{T^2})F^{(\lambda)}}}, 1\right). \tag{67}$$

The result of Theorem 3 can be proved based on the limiting distribution given by (67) and using the following results:

$F^{(\lambda)\prime}(K_{T^2} + I_{T^2})F^{(\lambda)} = 2tr((\Upsilon^{(\lambda)})^2) - 2tr((Y_\psi^{(\lambda)})^2)$, where now $Y_\psi^{(\lambda)}$ is defined as $Y_\psi^{(\lambda)} = \frac{1}{2}(\Psi^{(\lambda)\prime} + \Psi^{(\lambda)})$.
$tr(\Lambda' Q^{(\lambda)}\Lambda) = \frac{T^2}{6}(2\lambda^2 - 2\lambda + 1) - \frac{2}{6}$
$tr(\Xi' Q^{(\lambda)}) = -\frac{T^2}{6}(2\lambda^2 - 2\lambda + 1) + \frac{T}{2} - \frac{2}{3}$
$tr((Y^{(\lambda)})^2) = \frac{1}{2}tr((\Lambda' Q^{(\lambda)})^2) + \frac{1}{2}tr(\Lambda' Q^{(\lambda)}\Lambda)$
$tr((\Lambda' Q^{(\lambda)})^2) = -\frac{T^2}{12}(2\lambda^2 - 2\lambda + 1) + \frac{T}{2} - \frac{5}{6}$ and $tr((\Upsilon_\psi^{(\lambda)})^2) = \frac{-\frac{1}{T} + 2(\lambda-1)\lambda T}{6(\lambda-1)\lambda} - 1$.

Proof of Theorem 7: It can be proved following analogous steps to those followed for the proof of Theorem 1. Under the null hypothesis $\varphi = 1$, vector $y_{i,-1}$ can be decomposed as

$$y_{i,-1} = y_{i0}e + \Lambda e\beta_i + \Lambda u_i.$$

Multiplying both sides of the last relationship by $Q_*^{(\lambda)}$ yields

$$Q_*^{(\lambda)} y_{i,-1} = Q_*^{(\lambda)}\Lambda u_i,$$

since $Q_*^{(\lambda)}e = 0$ and $Q_*^{(\lambda)}\Lambda e = 0$. Also, note that, under $\varphi = 1$, the following relationships hold:

$$\Delta y_i = u_i + e\beta_i$$

and

$$Q_*^{(\lambda)}\Delta y_i = Q_*^{(\lambda)} u_i \quad \text{and} \quad Q_*^{(\lambda)}\Lambda\Delta y_i = Q_*^{(\lambda)}\Lambda u_i. \tag{68}$$

Using (68), the numerator and denominator of $\hat{\varphi}_*^{(\lambda)} - 1$ become

$$y'_{i,-1}Q_*^{(\lambda)}u_i = u_i'\Lambda' Q_*^{(\lambda)}u_i = \Delta y_i'\Lambda' Q_*^{(\lambda)}\Delta y_i \text{ and}$$

$$y'_{i,-1}Q_*^{(\lambda)}y_{i,-1} = u'_i\Lambda' Q_*^{(\lambda)}\Lambda u_i = \Delta y'_i\Lambda' Q_*^{(\lambda)}\Lambda\Delta y_i,$$

respectively. By Kitchin's LLN, it can be shown that the inconsistency of estimator $\hat{\varphi}_*^{(\lambda)}$ is given as

$$\hat{\varphi}_*^{(\lambda)} - 1 = \frac{\sum_{i=1}^N y'_{i,-1}Q_*^{(\lambda)}u_i}{\sum_{i=1}^N y'_{i,-1}Q_*^{(\lambda)}y_{i,-1}} \xrightarrow{p} \frac{b_2^{(\lambda)}}{\delta_2^{(\lambda)}} = \frac{tr(\Lambda' Q_*^{(\lambda)}\mathbf{\Gamma}_N)}{tr(\Lambda' Q_*^{(\lambda)}\Lambda\mathbf{\Gamma}_N)}. \tag{69}$$

The last result holds because, as $N \to +\infty$, we have

$$\frac{1}{N}\sum_{i=1}^N y'_{i,-1}Q_*^{(\lambda)}u_i - tr(\Lambda' Q_*^{(\lambda)}(\Gamma_N + \beta_N^2 J_T)) \xrightarrow{p} 0, \text{ where } \beta_N^2 = \frac{1}{N}\sum_{i=1}^n E((\beta_i)^2),$$
$$\text{or } \frac{1}{N}\sum_{i=1}^N y'_{i,-1}Q_*^{(\lambda)}u_i - tr(\Lambda' Q_*^{(\lambda)}\Gamma_N) \xrightarrow{p} 0,$$

since $tr(\Lambda' Q_*^{(\lambda)}J_T) = 0$, and

$$\frac{1}{N}\sum_{i=1}^N y'_{i,-1}Q_*^{(\lambda)}y_{i,-1} - tr(\Lambda' Q_*^{(\lambda)}\Lambda(\Gamma_N + \beta_N^2 J_T)) \xrightarrow{p} 0,$$

since $tr(\Lambda' Q_*^{(\lambda)}\Lambda J_T) = 0$.

The remaining of the proof follows the same steps with those of the proof of Theorem 1. That is, subtract the consistent estimator of $\frac{b_2^{(\lambda)}}{\delta_2^{(\lambda)}}$, given by (42), from $\hat{\varphi}_*^{(\lambda)} - 1$ and, then, apply standard asymptotic theory.

Proof of Theorem 8: To prove the theorem we will follow analogous steps to those for the proof of Theorem 6. In our analysis, we will employ relationships (55) and (56), where now vector $d_i^{(\lambda)}$ is extended and is defined as:

$$d_{*,i}^{(\lambda)} = \begin{pmatrix} (1-\varphi_N)a_i^{(1)} + \varphi_N\beta_i^{(1)} \\ (1-\varphi_N)a_i^{(2)} + \varphi_N\beta_i^{(2)} \\ (1-\varphi_N)\beta_i^{(1)} \\ (1-\varphi_N)\beta_i^{(2)} \end{pmatrix} = e_*\nu_i + (1-\varphi_N)\mu_i,$$

due to the presence of individual trends under the sequence of local alternatives φ_N. The non standardized test statistic $Z_2^{(\lambda)}$ can be written as

$$\begin{aligned} \hat{\delta}_2^{(\lambda)}\sqrt{N}\left(\hat{\varphi}_*^{(\lambda)} - \frac{\hat{b}_2^{(\lambda)}}{\hat{\delta}_2^{(\lambda)}} - \varphi_N\right) &= \hat{\delta}_2^{(\lambda)}\sqrt{N}\left(\hat{\varphi}_*^{(\lambda)} - \varphi_N - \frac{tr(\Psi_2^{(\lambda)}\hat{\Gamma}_N)}{\hat{\delta}_2^{(\lambda)}}\right) \\ &= \frac{1}{\sqrt{N}}\sum_{i=1}^N y'_{i,-1}Q_*^{(\lambda)}u_i - \frac{1}{\sqrt{N}}\sum_{i=1}^N tr(\Delta y'_i\Psi_2^{(\lambda)}\Delta y_i). \end{aligned} \tag{70}$$

Following analogous steps to those for the proof of Theorem 6, we can show that

$$Z_2^{(\lambda)} = V_2^{(\lambda)-1/2}\hat{\delta}_2^{(\lambda)}\sqrt{N}\left(\hat{\varphi}_*^{(\lambda)} - 1 - \frac{\hat{b}_2^{(\lambda)}}{\hat{\delta}_2^{(\lambda)}}\right) \xrightarrow{d} N(k_2, 1),$$

where $k_2 = \frac{-\Pi_1 - c[\Pi_2 + \Pi_3]}{\sqrt{F_2^{(\lambda)\prime}(K_{T^2} + I_{T^2})(\Gamma_N \otimes \Gamma_N)F_2^{(\lambda)}}}$. Under the assumption that disturbance terms are homoscedastic across i and t, and $p = 0$ it can be seen

$$tr(\Lambda' Q_*^{(\lambda)} \Lambda \Gamma_N) + tr(\Xi Q_*^{(\lambda)} \Gamma_N) - tr(\Psi_2^{(\lambda)} \Lambda \Gamma_N) - tr(\Lambda' \Psi_2^{(\lambda)} \Gamma_N) = 0,$$

which means that $\Pi_2 = 0$.

Proof of Theorem 9: See proof of Theorem 7.

References

[1] Andrews, Donald W K, 1993. Tests for Parameter Instability and Structural Change with Unknown Change Point. Econometrica, Econometric Society, vol. 61(4), pages 821-56, July.

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

[3] 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.

[4] Arellano-Valle, R. B., and Genton, M. G. (2008). On the exact distribution of the maximum of absolutely continuous dependent random variables. Statistics and Probability Letters, 78, 27-35.

[5] Bai J.(2010). Common breaks in means and variances for panel data. Journal of Econometrics 157, 78-92.

[6] 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.

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

[8] 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.

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

[10] Blundell R. and Bond S. 1998. Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics, Elsevier, vol. 87(1), pages 115-143, August.

[11] Binder M., Hsiao C. and Pesaran H. (2005). Estimation and inference in short panel vector autoregression with unit roots and cointegration. Econometric Theory, 21, 795-837.

[12] Breitung, J. and Pesaran H. 2005. Unit Roots and Cointegration in Panels. Cambridge Working Papers in Economics 0535, Faculty of Economics, University of Cambridge.

[13] Carrion-i-Silvestre J.L., Del Barrio-Castro, and Lopez-Bazo E. (2002). 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.

[14] 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.

[15] Chang Y. and Park J., 2002. On The Asymptotics Of Adf Tests For Unit Roots. Econometric Reviews, vol. 21(4), pages 431-447.

[16] Culver S.E. and Papell D.H. (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.

[17] de la Fuente, Angel, (1997). The empirics of growth and convergence: A selective review. Journal of Economic Dynamics and Control, Elsevier, vol. 21(1), pages 23-73, January.

[18] De Wachter S., Harris R.D.F., Tzavalis E. (2007). Panel unit root tests: the role of time dimension and serial correlation. Journal of Statistical Inference and Planning, 137, 230-244.

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

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

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

[22] Hadri K.& Rao Y., (2008). Panel Stationarity Test with Structural Breaks. Oxford Bulletin of Economics and Statistics, Department of Economics, University of Oxford, vol. 70(2), pages 245-269.

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

[24] Hall, B. H. and Mairesse J. (2005). Testing for unit roots in panel data: an exploration using real and simulated data. In D. Andrews and J. Stock (Eds.), Identification and Inference in Econometric Models: Essays in Honor of Thomas J. Rothenberg. Cambridge: Cambridge University Press.

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

[26] Harris R.D.F and Tzavalis E. (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 23, 149-166.

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

[28] Hsiao C., Pesaran H. and Tahmiscioglu K., (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. Journal of Econometrics, Elsevier, vol. 109(1), pages 107-150, July.

[29] Im, K.S., Pesaran H., and Shin Y. (2003). Testing for Unit Roots in Heterogenous Panels. Journal of Econometrics, 115, 53-74.

[30] Karavias Y. and Tzavalis E. (2012a). Testing for unit roots in short panels allowing for structural breaks. Unpublished.

[31] Karavias Y. and Tzavalis E. (2012b). On the local power of fixed T panel unit root tests under serial correlation. Unpublished.

[32] Kim, D (2011). Estimating a common deterministic time break in large panels with cross-sectional dependence. Journal of Econometrics, 164 (2), 310-330.

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

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

[35] Madsen E. (2008). GMM Estimators and Unit Root Tests in the AR(1) Panel Data Model. unpublished.

[36] Moon, R. and Perron B. (2004). Testing for Unit Root in Panels with Dynamic Factors. Journal of Econometrics, 122, 81 p.126.

[37] Moon H.R., Perron B. & Phillips P.C.B., (2007). Incidental trends and the power of panel unit root tests. Journal of Econometrics, 141(2), 416-459.

[38] Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. Econometrica 49, 1417-1426.

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

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

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

[42] Perron, P. (1997). Further Evidence on Breaking Trend Functions in Macroeconomic Variables. Journal of Econometrics, 80 (2),355-385.

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

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

[45] Phillips P.C.B., (1987). Time series regression with a unit root. Econometrica 55, 277–301.

[46] Ploberger, W. and Phillips P.C.B. (2002). Optimal Testing for Unit Roots in Panel Data. Unpublished.

[47] Said, S.E., Dickey, D.A., (1984). Testing for unit roots in autoregressive-moving average processes of unknown order. Biometrika 71, 599–607.

[48] Schott J.R. (1996). Matrix Analysis for Statistics, Wiley-Interscience.

[49] 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.

[50] Schwert, G.W., (1989). Tests for unit roots:a Monte Carlo investigation. Journal of Business and Economic Statistics 7, 147–160.

[51] Tzavalis E. (2002). Structural breaks and unit root tests for short panels. ESRC Conference, Cass Business School, City University London (2002).

[52] Vogelsang, T. and Perron, P. (1998). Additional tests for a unit root allowing for a break in the trend function at an unknown time. International Economic Review, 39(4):1073–1100.

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

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

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