Granger Centre
for time series econometrics

Granger Centre Discussion Paper Series

Testing for unit roots in panels with structural changes, spatial and temporal dependence when the time dimension is finite

by

Yiannis Karavias and Elias Tzavalis

Granger Centre Discussion Paper No. 14/03

The University of Nottingham

Testing for Unit Roots in Panels with Structural Changes, Spatial and Temporal Dependence when the Time Dimension is Finite

Yiannis Karavias[a,*] and Elias Tzavalis[b]

July 2014

Abstract

Finite T panel data unit root tests allowing for structural breaks, spatial cross section dependence, heteroscedasticity, serial correlation, heterogeneity and non-linear trends are proposed. The structural breaks can be at known or unknown dates. For the latter, analytic probability density functions of the asymptotic distributions of the tests are provided based on a minimum order statistic. The tests can accommodate general forms of spatial dependence for which the spatial weights matrix does not have to be defined due to the utilization of a non-parametric estimator. A set of sufficient conditions which determines admissible deterministic trend functions is also provided. Finally, extensive Monte Carlo experiments show the usefulness of the new tests.

JEL classification: C22, C23

Keywords: Panel data; Unit roots; Structural breaks; Spatial dependence; Serial correlation; Fixed T

The authors would like to thank David Harvey for assistance with the computations, Stelios Arvanitis, Georges Bresson, Felix Chan, R.B. Arellano Valle, Marc Genton, the participants of the seminar series in Queen's University Belfast and in University of Ioannina, the participants of the Conference in Honor of Professor Hashem Pesaran and of the 11th Conference on Research on Economic Theory and Econometrics, for helpful comments and suggestions.

$*$:Corresponding author

a :School of Economics and Granger Centre for Time Series Econometrics, University of Nottingham. E-mail: ioannis.karavias@nottingham.ac.uk.

b :Department of Economics, Athens University of Economics & Business. E-mail: etzavalis@aueb.gr

1 Introduction

In the last few years there has been an explosion in the number of applications of panel data unit root tests across the many areas of economics. Their main virtues, good size and power properties, stemming from the fact that they exploit both cross section (N) and time (T) dimensions of data sets, and their generality in specification have made them an integral tool of econometric analysis. Important applications of them include the following: the examination of economic growth convergence hypothesis at country, or regional, level (see, e.g., Mello (2011), and Hafner and Mayer-Foulkes (2013)), price convergence in manufacture (see, e.g., Yan et al. (2007), Suvankulov et al. (2012)), persistency in macroeconomic series like per capita consumption, unemployment rate, interest and inflation rates (see, e.g., Chen and Lee (2007), Basker (2005), Romero-Avila and Usabiaga (2007) and Constantini and Lupi (2007)), mean-reversion in firm growth and profitability (see, e.g., Bond et al. (2003) and Canarella et al. (2013)), persistency of oil and energy shocks (see, e.g., (Narayan et al. (2008), Apergis et al. (2010), Lean and Smyth (2013)), the validity of the purchasing power parity hypothesis (see, e.g., Harris et al. (2005) and Murray and Papell (2005)) and income inequality (see, e.g., Lin and Huang (2012)). The list is far from being exhaustive.

Within the above list of applications several important econometric problems, translated in strong assumptions, can arise. Firstly, in most of the above studies (especially the microeconomic ones) the time dimension of the panel data, denoted as T, is small (short). Thus, applications of panel unit root tests assuming large T, instead of fixed (finite), will not lead to good approximations of the small sample distributions of the tests, compared to their fixed-T counterparts. As shown by Harris and Tzavalis (1999) and De Wachter et al. (2007), in this case large-T panel unit root tests will lead to serious size distortions and power reductions.

Secondly, there is always the possibility of structural breaks in the individual effects or deterministic trend components of the panel series employed to test the unit root hypothesis. These breaks can be the outcome of an economic crisis, a credit crunch, a tax policy change, an oil price shock and monetary or fiscal shocks. They are expected to have common side effects to all cross section units of the panel. Allowance for common breaks in panel unit root tests in the finite T framework is made by Tzavalis (2003), Hadri et al. (2012), and Karavias and Tzavalis (2014b) but this is done only for the simple case that the error terms are independent and identically distributed (IID) across both dimensions of the panel. The motivation of these tests was to pursue ideas on how to conduct panel data inference about unit roots in the presence of breaks, when T is fixed. For large-T panels, often used in macroeconomic studies, tests allowing for breaks were proposed by Carrion-i-Silvestre et al. (2005) and Bai and Carrion-i-Silvestre (2009). These tests assume that T increases faster than the cross section dimension of the panel, N, which does not happen in microeconometric studies.

Thirdly, it concerns the assumptions that the error terms of the individual series of the panel are serially and cross sectionally uncorrelated, made by the first generation of finite-T panel unit rot tests (see, e.g., Harris and Tzavalis (1999, 2004), and Hadri et al. (2005)). These assumptions can not be easily relaxed in the framework of finite T inference procedures. Despite the fact that in most microeconomic studies T is very short, there is always the possibility that the errors terms are serially correlated and/or heteroscedastic across the time dimension of the panel. To deal with this problem, De Wachter and Tzavalis (2007) and De Blander and Dhaene (2012) consider finite T panel unit root tests allowing for serial correlation. But this is done for the specific cases that the error terms follow a moving average (MA), or an autoregressive (AR), procedure of lag order one, respectively, which may be proved very restrictive in practice.[1] Furthermore, cross section dependence can be attributed to common stochastic factors of the error terms of the panel (see, e.g., Moon and Perron (2004)) or to spatial dependence based on general types of distances of the error terms. Distance can be defined using specific metrics either in terms of real geographical distances or in terms of economic and social distances (see, e.g., Conley (1999), and Conley and Ligon (2002)).[2] Spatial dependence may be also exacerbated if the existence of structural breaks in the deterministic components of the dynamic panel data model is ignored.

To address the above problems in the context of finite T and large N asymptotics, in this paper we propose panel unit root test statistics allowing for structural breaks, spatial dependence, serial correlation, heteroscedasticity, heterogeneity across individuals and linear and/or non-linear trends of the underlying panel data dynamic model. Furthermore, we allow for multiple structural breaks at known or unknown dates that can be considered under the alternative hypothesis of stationarity or the null of unit roots. Our tests allow for general forms of spatial dependence. They have the interesting feature that the spatial weights matrix does not have to be defined, because a non-parametric estimator of the variance-covariance matrix of the error terms is employed. The lag-order of serial correlation of the error terms allowed by our tests can be larger than one. This can be chosen by the researcher depending on the T-dimension of the panel and the number of breaks, as well as their location considered during the sample. Finally, our tests can be applied to models with individual intercepts, individual trends and other non-linear functions, such as individual quadratic trends. As a by-product of our analysis, partial structural change models can be also considered as well as models with common trends.

In developing the tests, the paper makes a number of contributions to the econometric literature. First, in the case of an unknown date break, the limiting distribution of the suggested test statistics is derived analytically as the minimum order statistic between all alternative sequential statistics that the break can occur at different points in time, during the sample. The limiting distribution of this statistic is obtained as the minimum value of a finite number of correlated variables. This distribution is provided 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 form of this distribution enables us to obtain critical values of the suggested test statistics without having to rely on Monte Carlo simulations, which substantially facilitates application of the tests in practice. To our knowledge, this is the first paper which provides an analytic form of the distribution of a minimum of sequential statistics to test for unit roots in the presence of breaks, since this method was introduced in

[1] Heteroscedasticity in long panels has been recently studied by Westerlund (2014).

[2] Recent contributions in the area of panel unit root tests allowing for cross section dependence comes from Bresson et al. (2007), Chang and Song (2009), Sul (2009), Bai and Ng (2010), Palm et al. (2011), Pesaran et al. (2013) and Meligotsidou et al (2014), inter alia. These studies however assume large T dimension of the panel. Recently, Robertson et al. (2014) proposed a fixed T panel unit root test assuming factors in the errors, based on a GMM estimator which exploits moments available by these factors. For a review of cross section dependence see Sarafidis and Wansbeek (2012).

econometrics by Zivot and Andrews (1992), and Perron and Vogelsang (1992). Note that this distribution can also derived for the case that the number of breaks is more than one.

Second, the tests suggested by the paper are obtained based on the within group (WG) least squares (LS) estimator of the dynamic, AR(1) panel data model adjusted (corrected) for its inconsistency (bias). The latter comes from the within transformation of the individual panel data series considered by the above estimation method, and the serial and spatial correlation of the error terms.[3] This estimation method requires the minimum set of assumptions about the data generating process. In particular, under the null hypothesis of unit roots, it is invariant to the initial observations (conditions) of the panel data. Thus, it does not require mean and covariance stationarity conditions on these observations, as the generalized method of moments (GMM) and the first difference maximum likelihood (FDML) estimators (see, Madsen (1998), and Kruiniger (2008), respectively). These stationary conditions can be proved very restrictive, in practice when testing nonstationarity of data, especially in the presence of structural breaks (see, e.g., Arellano (2003) and De Wachter and Tzavalis (2012)). Furthermore, as noted recently by De Wachter et al. (2007), Han and Phillips (2012, 2013), taking first differences of the individual series of the panel, assumed by the above two estimators to remove individual effects, may lead to identification problems of the parameters of interest for the GMM method and/or to estimation problems for the FDML method. In addition to the initial observations, the LS estimator employed in our tests is invariant to the individual effects of the dynamic panel data model.

The property of our suggested test statistics to rely on a bias adjusted LS estimator draws from two strands of the econometric literature. First, it is similar in spirit to the modified LS estimator of Phillips and Hansen (1990), for single time series analysis. In this way, a non-parametric correction of the estimator is employed which purges the effects of individual, linear and/or quadratic trends and serial and spatial correlation nuisance parameters. However, instead of correcting the limiting distribution of the LS estimator based on estimates of its long run variance-covariance matrix, it does this by exploiting cross section information in estimating the variance-covariance matrix of the error terms. Second, the non-parametric bias correction considered by our tests is based on consistent estimation of the error terms variance-covariance matrix suggested by Abowd and Card (1989) and Arellano (1990, 2003). Under some conditions (e.g., maximum order of serial correlation), the nuisance parameters of this variance-covariance matrix (due, for instance, to short run dynamics, spatial correlation effects and/or incidental trends) can be identified and consistently estimated. This is done by employing appropriately designed selection matrices which exploit the covariance structure of the error terms.

The paper is organized as follows. In Section 2, we derive the limiting distributions of the test statistics under the assumption that the error terms are white noise processes. This analysis will help us to better interpret the limiting distribution of the sequential version of the test statistics, in the case of an unknown date break. In Section 3, we generalize the test statistics to allow for serial correlation in the error terms. In Section 4 we extend the tests to allow also for individual linear trends. In this section, we also show how

[3] Dynamic panel data LS estimators adjusted for their bias have been suggested in the literature by Kiviet (1995), Harris and Tzavalis (1999), Hahn and Kuersteiner (2002), Phillips and Sul (2007), and De Blander and Dhaene (2012), *inter alia.*

to carry out the tests when there is a break in the individual effects of the dynamic panel data model under the null hypothesis of unit roots. Section 5 considers the cases of multiple breaks, partial structural change and more general (quadratic) patterns of linear trends. Section 6 considers the case that the error terms are spatially correlated. Section 7 conducts a Monte Carlo simulation study to examine the small sample performance of the tests. Section 8 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, the proposed panel unit root tests are presented under the assumption that the error terms of the AR(1) panel data model are independent, identically and normally distributed ($NIID$) across the time (t) and cross section units (i) of the model. As said before, this provides a clear illustration of the inference procedure employed to test the null hypothesis of unit roots and to derive the analytic form of the limiting distribution of the suggested test statistics in the case where the break point is treated as unknown. In the next sections, this assumption will be relaxed.

2.1 Known Date of the break

Consider the following AR(1) panel data model:

$$\begin{aligned} y_{it} &= a_i^{(1)} + \zeta_{it} \quad \text{for } t \leq \lambda, \quad t = 1, 2, ..., T \\ y_{it} &= a_i^{(2)} + \zeta_{it} \quad \text{for } t > \lambda \\ \zeta_{it} &= \varphi \zeta_{it-1} + u_{it,}, \quad \text{with } \varphi \in (-1, 1], \end{aligned} \tag{1}$$

where λ denotes the time point of the sample, referred to as break point. The model assumes that a common break in its individual effects, denoted as α_i, occurs, for all cross section units $i = 1, 2, ..., N$. The date of the break is only allowed at times $2, ..., T-1$, i.e. $\lambda \in I = \{2, ..., T-1\}$.

Define the following $N \times 1$ vectors collecting the cross-section observations of the dependent variable y_{it} and error terms ζ_{it} and u_{it} of model (1), for all t: $y_t = (y_{1t}, ..., y_{Nt})'$, $\zeta_t = (\zeta_{1t}, ..., \zeta_{Nt})$ and $u_t = (u_{1t}, ..., u_{Nt})$, respectively. Stacking y_t, ζ_t and u_t into $NT \times 1$ dimension vectors, $y = (y_1', ..., y_T')'$, $\zeta = (\zeta_1', ..., \zeta_T')'$ and $u = (u_1', ..., u_T')'$ respectively, model (1) can be written in a more compact form as follows

$$\begin{aligned} y &= e_T^{(1)} \otimes a^{(1)} + e_T^{(2)} \otimes a^{(2)} + \zeta \\ \zeta &= \varphi \zeta_{-1} + u, \end{aligned} \tag{2}$$

where $\otimes$ denotes the Kronecker product, $e_T^{(1)}$ and $e_T^{(2)}$ are $T \times 1$ vectors whose elements are defined as follows: $e_{Tt}^{(1)} = 1$ if $t \leq \lambda$ and 0 otherwise, and $e_{Tt}^{(2)} = 1$ if $t > \lambda$ and 0 otherwise, $a^{(1)}$and $a^{(2)}$ are $N \times 1$ dimension vectors which collect individual effects α_i before and after break point λ, respectively, and $\zeta_{-1} = (\zeta_0', ..., \zeta_{T-1}')$.

Under the null hypothesis of unit roots (i.e., H_0: $\varphi = 1$), model (1) reduces to the pure random walk model

$y = y_{-1} + u$ while, under the alternative of stationarity (i.e., H_1: $\varphi < 1$), it considers a common structural break in individual effects a_i, for all i. The above specification of the null and the alternative hypotheses is often considered in panel unit root inference procedures (see, e.g., Bai and Carrion-i-Silvestre (2009) and Karavias and Tzavalis (2014b)). 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, as was first pointed out by Perron (1989). The common break assumption across all units of the panel i can be attributed to a monetary regime shift, which is common across all agents (or firms) in the economy, or to a structural economic shock which is independent of error terms u_{it}, like a credit crunch or an exchange rate realignment. As aptly noted by Bai (2010), even if each series of a 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. Note that in model (1), the magnitude of the break change $a_i^{(2)} - a_i^{(1)}$ at point λ can be different across units i, thus allowing for each individual unit of the economy to respond also idiosyncratically to the effects of a structural break.

To derive the limiting distribution of a test statistic of H_0: $\varphi = 1$, we make the following assumption about the sequence of error terms $\{u_{it}\}$.

Assumption A

$\{u_{it}\}$ are normal *IID* random variables (i.e., $NIID$), which are independent and identically distributed across i and t, with mean $E(u_{it}) = 0$, constant variance $\sigma^2 = E(u_{it}^2)$ and finite $2 + \epsilon$ moments.

Assumption A enables us to apply standard asymptotic theory for independent processes under H_0: $\varphi = 1$. It is particularly strong, as it imposes normality, time series and cross-section independence, and homogeneity of error terms u_{it}. To test the above hypothesis based on model (1), allowing for a structural break in individual effects, we rely on the within group (WG) least squares (LS) estimator of φ, defined as

$$\hat{\varphi}^{(\lambda)} = \frac{y'_{-1} Q^{(\lambda)} y}{y'_{-1} Q^{(\lambda)} y_{-1}}, \tag{3}$$

where $y_{-1} = (y'_0, ..., y'_{T-1})'$ and $Q^{(\lambda)} = I_{NT} - X^{(\lambda)}(X^{(\lambda)\prime} X^{(\lambda)})^{-1} X^{(\lambda)\prime}$ is the annihilator matrix which demeans the individual series of the panel y_{it}. $\hat{\varphi}^{(\lambda)}$ is also well known as dummy variables LS (LSDV) estimator. For model (1), matrix $X^{(\lambda)}$ is defined as $X^{(\lambda)} = \left[e^{(1)}, e^{(2)}\right]$, with $e^{(j)} = e_T^{(j)} \otimes I_N$ for $j = 1, 2$. I_{NT} and I_N are identity matrices of dimension $NT \times NT$ and $N \times N$, respectively. The superscript λ indicates dependence of matrices $Q^{(\lambda)}$ and $X^{(\lambda)}$ on break date λ.

Estimator $\hat{\varphi}^{(\lambda)}$ has the interesting property that, under H_0: $\varphi = 1$, is invariant (similar) to the initial conditions of the panel y_{i0} and individual effects $\alpha_i^{(j)}$ because the latter are orthogonal to $Q^{(\lambda)}$. Thus, assumptions on y_{i0}, like mean and covariance stationarity made by the generalized method of moments and conditional and unconditional maximum likelihood estimation procedures (see, e.g., Bond et al. (2005) and Kruiniger (2008)) are no longer required.[4]

[4] LS estimator $\hat{\varphi}^{(\lambda)}$ is also attractive for its small sample properties. De Wachter et al. (2007) and Han and Phillips (2010)

Since $\hat{\varphi}^{(\lambda)}$ is an inconsistent estimator, due to the correlation between vectors y_{-1} and u induced by the within transformation of the individual series of the panel y_{it} (see, e.g. Nickel (1981)), panel unit root test statistics relying on this estimator must correct for its inconsistency (see Harris and Tzavalis (1999)). Under H_0: $\varphi = 1$, the inconsistency (asymptotic bias) of estimator $\hat{\varphi}^{(\lambda)}$ is given by the following relationship:

$$p\lim_{N}\left(\hat{\varphi}^{(\lambda)} - 1 - \frac{b^{(\lambda)}}{d^{(\lambda)}}\right) = p\lim_{N}\left(\hat{\varphi}^{(\lambda)} - 1 - \frac{tr(\Lambda' Q^{(\lambda)}\Gamma)}{tr(\Lambda' Q^{(\lambda)}\Lambda\Gamma)}\right) = 0, \tag{4}$$

(see Lemma A2 for a proof), where $b^{(\lambda)} \equiv E(y_{-1}' Q^{(\lambda)} u) = tr(\Lambda' Q^{(\lambda)}\Gamma)$ and $d^{(\lambda)} \equiv E(y_{-1}' Q^{(\lambda)} y_{-1}) = tr(\Lambda' Q^{(\lambda)}\Lambda\Gamma)$, where Λ is a $NT \times NT$ deterministic matrix defined in the Appendix and $\Gamma = E(uu') = \sigma^2 I_{NT}$ is the variance-covariance matrix of the stacked vector of error terms, u.

To estimate $\Gamma = \sigma^2 I$, entering the bias function of estimator $\hat{\varphi}^{(\lambda)}$, $\frac{b^{(\lambda)}}{d^{(\lambda)}}$, we can rely on a consistent estimator of the variance of u_{it}, σ^2, under H_0: $\varphi = 1$. Lemma A4 shows that such an estimator is given as

$$\hat{\sigma}^2 = \Delta y' \Psi^{(\lambda)} \Delta y / tr(\Psi^{(\lambda)}),$$

where Δ is the difference operator and $\Psi^{(\lambda)}$ is a $NT \times NT$ selection matrix having in its main diagonal the corresponding elements of matrix $\Lambda' Q^{(\lambda)}$ and zeroes elsewhere. The consistency of $\hat{\sigma}^2$ can be easily seen by noticing that, under H_0: $\varphi = 1$, $\Delta y_{it} = u_{it}$, for all i and t. Matrix $\Psi^{(\lambda)}$ is designed so as to select all non-zero elements of variance-covariance matrix Γ and it assigns weights to them according to $\Lambda' Q^{(\lambda)}$ so as $tr(\Lambda' Q^{(\lambda)}) = tr(\Psi^{(\lambda)})$. In doing so, $\Psi^{(\lambda)}$ enables consistent estimation of σ^2 (and ultimately Γ) and it also captures the correlation structure between vectors y_{-1} and u, which is induced by the within transformation matrix $Q^{(\lambda)}$.

Having defined the above consistent estimator of σ^2, the following bias adjusted estimator of φ can be employed to test H_0: $\varphi = 1$:

$$\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}, \tag{5}$$

where $\hat{d}^{(\lambda)} = \frac{1}{N} y_{-1}' Q^{(\lambda)} y_{-1}$ and $\hat{b}^{(\lambda)} = \frac{1}{N}\hat{\sigma}^2 tr(\Psi^{(\lambda)})$. Note that $\hat{\varphi}^{(\lambda)}$ is adjusted only for the bias of its numerator like in Levin, Lin and Chu (2002) (see also Kruninger and Tzavalis (2002) and Moon and Perron (2004)). Under the null hypothesis, it can be shown that the probability limit of the numerator of the adjusted estimator $\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}$, given as

$$\frac{1}{N} y_{-1}' Q^{(\lambda)} u - \frac{1}{N}\Delta y' \Psi^{(\lambda)} \Delta y = \frac{1}{N} u' \left(\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)}\right) u, \tag{6}$$

becomes zero, i.e., $E\left[u'\left(\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)}\right) u\right] = tr\left[\left(\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)}\right) E(uu')\right] = 0$. This happens because $\left(\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)}\right)$ has zeros in its main diagonal, by the definition of selection matrix $\Psi^{(\lambda)}$ and $E(uu') = \sigma^2 I_{NT}$. From equation (6), it can be clearly seen that, to test H_0: $\varphi = 1$, the bias adjusted estimator $\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}$

have noticed that the performance of the GMM estimator of φ, compared to LS estimator $\hat{\varphi}^{(\lambda)}$, may deteriorate in small samples due to the inaccurate estimation of its weighting matrix. Furthermore, recently Han and Phillips (2013) have find pathologies of the first difference maximum likelihood with a high impact on both small sample and asymptotic performance (see also our discussion in the introduction).

relies on the sample moments of the zero elements of matrix Γ, i.e., $E(u_{it}u_{is}) = 0$, for all $s \neq t$. The following theorem provides the panel data unit root test statistic and its limiting distribution based on $\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}$.

Theorem 1 *Let Assumption A hold. Then, under H_0: $\varphi = 1$ and λ known, we have*

$$Z^{(\lambda)} \equiv V^{(\lambda)-1/2}\hat{d}^{(\lambda)}\sqrt{N}\left(\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}} - 1\right) \xrightarrow{d} N(0,1) \tag{7}$$

as $N \to \infty$, where $\frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}$ is a consistent estimator of the asymptotic bias of $\hat{\varphi}^{(\lambda)}$, with $\hat{b}^{(\lambda)} = \frac{1}{N}\hat{\sigma}^2 tr(\Psi^{(\lambda)})$ and $\hat{d}^{(\lambda)} = \frac{1}{N}y'_{-1}Q^{(\lambda)}y_{-1}$, and

$$V^{(\lambda)} = \frac{1}{N}2\sigma^4 tr(F^{(\lambda)}F^{(\lambda)}) \tag{8}$$

is the variance of the limiting distribution $\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}$, where $F^{(\lambda)} = \frac{1}{2}(\Lambda' Q^{(\lambda)} + Q^{(\lambda)}\Lambda - \Psi^{(\lambda)} - \Psi^{(\lambda)\prime})$.

The test statistic $Z^{(\lambda)}$, given by Theorem 1, can be easily implemented to test H_0: $\varphi = 1$ based on the tables of the standard normal distribution. To this end, variance $V^{(\lambda)}$ can be consistently estimated by replacing σ^2 with $\hat{\sigma}^2 = \frac{\Delta y' \Psi^{(\lambda)} \Delta y}{tr(\Psi^{(\lambda)})}$.

2.2 Unknown Date of the Break

In this section the assumption of known λ is relaxed. As in the single time series literature, the selection of the break point λ is viewed as the outcome of minimizing the standardized test statistic $Z^{(\lambda)}$ over all possible $\lambda \in I$, after trimming out the initial and final parts of the time series observations of the panel. The minimum value of test statistics $Z^{(\lambda)}$, for all $\lambda \in I$, denoted as $\min_{\lambda \in I} Z^{(\lambda)}$, will give the least favourable result on H_0: $\varphi = 1$. Let $\hat{\lambda}_{\min}$ denote break point λ at which the minimum value of $Z^{(\lambda)}$ is obtained. Then, H_0: $\varphi = 1$ will be rejected if

$$Z^{(\hat{\lambda}_{\min})} < z_a^{\min}, \tag{9}$$

where $z_a^{\min}$ denotes the size a left-tail critical value of the limiting distribution of statistic $\min_{\lambda \in I} Z^{(\lambda)}$. The following theorem provides this distribution analytically.

Theorem 2 *Let Assumption A hold. Then, under H_0: $\varphi = 1$ and λ unknown, we have*

$$\min_{\lambda \in I} Z^{(\lambda)} \xrightarrow{d} \psi \equiv \min_{\lambda \in I} N(0, \Sigma) \tag{10}$$

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

$$\sigma_{\mu s} = \frac{tr(F^{(\mu)}F^{(s)})}{\sqrt{tr(F^{(\mu)}F^{(\mu)})}\sqrt{tr(F^{(s)}F^{(s)})}}, \tag{11}$$

where μ and s denote two different break points of the sample that the break can occur.

Theorem 2 implies that critical values of the limiting distribution of the standardized test statistic $\min_{\lambda \in I} Z^{(\lambda)}$, $z_a^{\min}$, 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. This distribution is not directly available. However, given that $\min\{Z^{(2)}, Z^{(3)}..., Z^{(T-1)}\} = \max\{-Z^{(2)}, -Z^{(3)}..., -Z^{(T-1)}\}$, the distribution of the maximum of normal variables $-Z^{(\lambda)}$ provides the critical value $z_a^{\min}$ for a significance level a, i.e.,

$$P(\psi < z_a^{\min}) = P(-\psi > -z_a^{\min}) = a. \tag{12}$$

The integral function $P(\psi > -z_a^{\min}) = a$ can be calculated numerically based on the probability density function (pdf) of $-\psi$. 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 follows:

$$f_\psi(x) = \sum_\lambda \phi(x; \mu_\lambda, \Sigma_{\lambda,\lambda})\Phi(xe_{T-3}; \mu_{-\lambda,\lambda}, \Sigma_{-\lambda-\lambda,\lambda}), \quad x \in R, \tag{13}$$

where e_{T-3} is a $(T-3)$-column vector of unities, $\phi(\cdot)$ and $\Phi(\cdot)$ are respectively 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 $(T-2) \times 1$ dimension vector Z. This vector consists of random variables $Z^{(\lambda)}$, for $\lambda \in I$, partitioned as $Z = (Z^{(-\lambda)} \vdots Z^{(\lambda)})'$, where $Z^{(-\lambda)}$ is a $(T-3) \times 1$ dimension vector consisting of the remaining elements of Z, which exclude $Z^{(\lambda)}$.

The above pdf of random variable $-\psi$, defined as $f_\psi(x)$, is a mixture of normal marginal densities $\phi(x; \mu_\lambda, \Sigma_{\lambda,\lambda})$ corresponding to all possible break points of the sample λ. These densities are weighed with the cdf values of the $(T-3)$-column vector xe_{T-3}, given as $\Phi(xe_{T-3}; \mu_{-\lambda,\lambda}(x), \Sigma_{-\lambda-\lambda,\lambda})$. Intuitively, the pdf formula given by (13) sums up the probabilities that one random variable $-Z^{(\lambda)}$ takes its maximum value x (implying that $Z^{(\lambda)}$ takes its minimum value), while the remaining variables, collected in vector $-Z^{(-\lambda)}$, take values smaller than x. After the pdf has been calculated, critical value $z_a^{\min}$ can be easily obtained, by simple numerical integration.

3 Non-normal, Heterogeneous and Serially Correlated Error Terms

The test statistics presented in the previous section can be extended to allow for non-normal, serially correlated, heteroscedastic and heterogeneous across time (t) error terms u_{it}. Due to the finite-T dimension of model (1) and the allowance for a common structural break in the individual effects $\alpha_i^{(j)}$, $j = 1, 2$, the maximum order of serial correlation of u_{it}, denoted as $p_{\max}$, of the generalized version of the test statistics will be a function of T. To derive the limiting distribution of these statistics, next we relax Assumption A.

Assumption B

(i) $\{u_{it}\}$ are random variables independent across i, with $E(u_{it}) = 0$, for all i and t, and all $4+\epsilon$ mixed moments finite. For all $i = 1, ..., N$, we have $E(u_{it}u_{is}) = 0$ for $t < s$ and $s = t+p+1, ..., T$, where p denotes the order of serial correlation of u_{it}. For p, we have $p \leq p_{\max} = \left[\frac{T}{2} - 2\right]^* < T$, where $[.]^*$ denotes the greatest integer function.

(ii) Define $\Gamma_{iT} = E(u_i^* u_i^{*\prime})$, where $u_i^* = (u_{i1}, ..., u_{iT})'$ and let Γ_{iT} be finite, for all i. Then, define $\Gamma_T = \frac{1}{N}\sum_{i=1}^N \Gamma_{iT}$ for which it holds that $\lim_N (N\Gamma_T)^{-1}\Gamma_{iT} = \lim_N \left(\sum_{i=1}^N \Gamma_{iT}\right)^{-1} \Gamma_{iT} = 0$. Finally, assume that the limit of Γ_T, denoted as $\Gamma_{Tu} = \lim_N \frac{1}{N}\sum_{i=1}^N \Gamma_{iT}$, is finite and positive definite.

Condition (i) of Assumption B is more general than that of Assumption A. It does not require normality of u_{it} and it allows for serial correlation of u_{it}. The maximum order of serial correlation assumed, i.e., $p_{\max} = \left[\frac{T}{2} - 2\right]^*$, enables us to implement our tests independently of break date λ. It is chosen so as to allow for breaks in the beginning, or the end of the sample; if the date of the break is known to be in the middle of the sample, then larger orders of serial correlation can be considered.[5] Although it is assumed that $p_{\max}$ is common for all i, each cross-section unit i can exhibit a different order of serial correlation provided that this does not exceed $p_{\max}$. Condition (ii) of the assumption fulfils the no dominating variance requirement of the Lindeberg-Feller central limit theorem, as now error terms u_{it} are heterogeneous across i. In addition to this, this condition allows u_{it} to follow an unknown pattern of heteroscedasticity across t.

The assumption of serial correlation of error terms u_{it} made above complicates the adjustment of LS estimator $\hat{\varphi}^{(\lambda)}$ for its inconsistency. This happens because the variance-covariance matrix Γ is no longer diagonal, i.e., $\Gamma \neq \sigma^2 I_{NT}$; Lemma A5 shows that the inconsistency of the estimator is still given by (4), i.e., $p\lim_N \left(\hat{\varphi}^{(\lambda)} - 1 - \frac{tr(\Lambda' Q^{(\lambda)}\Gamma)}{tr(\Lambda' Q^{(\lambda)}\Lambda\Gamma)}\right) = 0$. If errors u_{it} are heteroscedastic and heterogeneous across units i, the diagonal elements of Γ are not equal with each other. In addition, if u_{it} are serially correlated, then there will be also non-zero off diagonal elements of Γ. All these nuisance parameters of matrix Γ make a consistent estimation of it impossible, if T is finite.

However, because to correct for the inconsistency of estimator $\hat{\varphi}^{(\lambda)}$ a consistent estimator of $b^{(\lambda)} = tr(\Lambda' Q^{(\lambda)}\Gamma)$ is required and not of Γ itself, as shown by (4) and (5), the above problem can be resolved by selecting the non-zero elements of variance-covariance matrix $\Gamma_T = \frac{1}{N}\sum_{i=1}^N E(u_i^* u_i^{*\prime})$. These elements are defined as follows: $E(u_{it}u_{is}) \neq 0$, for all $s = 1, 2, ..., p$ and $t < s$. The matrix which selects the non-zero elements of Γ_T, denoted as $\Psi_T^{(\lambda)}$, will be defined as one which has in its main diagonal, and its p-lower and $p-$upper diagonals the corresponding elements of matrix $\Lambda' Q^{(\lambda)}$, and zero otherwise. Then, inference about unit roots can be conducted based on the remaining, zero-mean elements of Γ_T, i.e., $E(u_{it}u_{is}) = 0$, for all $s = t+p+1, ..., T$, where $p \leq p_{\max}$. These elements exist by the assumption that there is an upper bound in the order of correlation of u_{it}, given by condition $p_{\max} < T$. To consistently estimate matrix Γ_T, we can rely on the following nonparametric estimator (see Lemma A7 for a proof):

$$\hat{\Gamma}_T = \frac{1}{N}\sum_{i=1}^N \Delta y_i^* \Delta y_i^{*\prime},$$

[5] Note that, for single time series unit root tests, $p_{\max}$ is assumed to increase with T with an order of $o(T^{1/2})$, see Chang and Park (2002).

where $\Delta y_i^* = (\Delta y_{i1}, ..., \Delta y_{iT})'$ and $\Delta y_{it} = u_{it}$ under H_0: $\varphi = 1$. Using the property that the trace function $tr : R^{NT \times NT} \to R$ is a linear map which is not $1-1$, it can be shown that, under H_0: $\varphi = 1$, we have

$$p\lim_N \left[\frac{1}{N} tr(\Psi^{(\lambda)} \hat{\Gamma}) - \frac{1}{N} tr(\Psi^{(\lambda)} \Gamma) \right] = 0, \tag{14}$$

where $\hat{\Gamma} = \hat{\Gamma}_T \otimes I_N$ and $\Psi^{(\lambda)} = \Psi_T^{(\lambda)} \otimes I_N$.[6] Note that this result holds despite the fact that $\hat{\Gamma}$ an inconsistent estimator of Γ (see Remarks A3 and A4 in the appendix for a further discussion). The above justify the use of

$$\hat{b}^{(\lambda)} = \frac{1}{N} tr(\Psi^{(\lambda)} \hat{\Gamma}) \tag{15}$$

as a consistent estimator of $b^{(\lambda)} = tr(\Lambda' Q^{(\lambda)} \Gamma)$, when adjusting $\hat{\varphi}^{(\lambda)}$ for its inconsistency. The following theorem provides the asymptotic distribution of the adjusted for its bias estimator $\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}$.

Theorem 3 *Let Assumption B hold. Then, under H_0: $\varphi = 1$ and λ known, we have*

$$Z^{(\lambda)} \equiv V^{(\lambda)-1/2} \hat{d}^{(\lambda)} \sqrt{N} \left(\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}} - 1 \right) \xrightarrow{d} N(0,1) \tag{16}$$

as $N \to \infty$, where $\hat{d}^{(\lambda)} = \frac{1}{N} y_{-1}' Q^{(\lambda)} y_{-1}$, $\frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}$ is a consistent estimator of the asymptotic bias of $\hat{\varphi}^{(\lambda)}$, with $\hat{b}^{(\lambda)} = \frac{1}{N} tr(\Psi^{(\lambda)} \hat{\Gamma})$ where $\hat{\Gamma} = \hat{\Gamma}_T \otimes I_N$ and $\Psi^{(\lambda)} = \Psi_T^{(\lambda)} \otimes I_N$, and

$$V^{(\lambda)} = \frac{1}{N} 2tr(F^{(\lambda)} \Gamma F^{(\lambda)} \Gamma) \tag{17}$$

is the variance function of the limiting distribution of the adjusted for its inconsistency estimator $\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}$, with $F^{(\lambda)} = \frac{1}{2}(\Lambda' Q^{(\lambda)} + Q^{(\lambda)} \Lambda - \Psi^{(\lambda)} - \Psi^{(\lambda)\prime})$.

When the date of the brake λ is unknown, implementation of the test statistic $Z^{(\lambda)}$ given by Theorem 2 follows similar steps to those described in the previous section, for the case that $u_{it} \sim NIID(0, \sigma^2)$. The elements of the variance-covariance matrix of sequential statistic $\min_{\lambda \in I} Z^{(\lambda)}$, $\Sigma \equiv [\sigma_{\mu s}]$, can be calculated based on the following formula:

$$\sigma_{\mu s} = \frac{tr(F^{(\mu)} \Gamma F^{(s)} \Gamma)}{\sqrt{tr(F^{(\mu)} \Gamma F^{(\mu)} \Gamma)} \sqrt{tr(F^{(s)} \Gamma F^{(s)} \Gamma)}}, \tag{18}$$

where Γ is replaced by its estimate $\hat{\Gamma} = \hat{\Gamma}_T \otimes I_N$. Critical values of the distribution of random variable ψ can be calculated by replacing the values of $\sigma_{\mu s}$ in pdf formula (13) with those of $\sigma_{\mu s}$, given by (18).

4 Linear Deterministic Trends

In this section, we suggest unit root test statistics based on extensions of model (1) allowing for individual linear trends, referred to as incidental trends. Two linear trend specifications of model (1) are considered. The first assumes a common break in these trends is present only under alternative hypothesis H_1: $\varphi < 1$ (see,

[6] Note that, for $\Gamma = \sigma^2 I_{NT}$, matrix $\Psi^{(\lambda)}$ reduces to that defined by Theorem 1.

e.g., Zivot and Andews (1992), Lumsdaine and Papell (1997) and Karavias and Tzavalis (2014b)), while the second assumes that the break is present under both H_0: $\varphi = 1$ and H_1: $\varphi < 1$ (see, e.g., Lee and Strazicich (2003) and Kim and Perron (2009)). The first of the above specifications of model (1) is more appropriate in distinguishing between non-stationary panel data series which exhibit persistence and stationary series which evolve around broken linear trends. The second case is more suitable when considering more explosive panel data models under H_0: $\varphi = 1$, which can exhibit both deterministic and random shifts from their linear trends which are persistent.

4.1 Broken Trends Under the Alternative Hypothesis of Stationarity

Consider the following extension of model (1) under H_0: $\varphi = 1$:

$$y_{it} = a_i + \beta_i t + \zeta_{it}, \quad \text{for } t = 1, ..., T \tag{19}$$

where ζ_{it} is defined in (1) as $\zeta_{it} = \zeta_{it-1} + u_{it}$, and under H_1: $\varphi < 1$

$$\begin{aligned} y_{it} &= a_i^{(1)} + \beta_i^{(1)} t + \zeta_{it}, \quad \text{for } t \leq \lambda \\ y_{it} &= a_i^{(2)} + \beta_i^{(2)} t + \zeta_{it}, \quad \text{for } t > \lambda, \end{aligned}$$

where $\beta_i^{(1)}$ and $\beta_i^{(2)}$ are the slope coefficients of individual linear trends before and after the break point λ, respectively. Under H_0: $\varphi = 1$, the above model assumes that $a_i^{(1)} = a_i^{(2)} = a_i$ and $\beta_i^{(1)} = \beta_i^{(2)} = \beta_i$. Using stacked vector notation, model (19) can be written as

$$y = e_T \otimes a + \tau_T \otimes \beta + \zeta$$

while under H_1: $\varphi < 1$

$$y = e_T^{(1)} \otimes a^{(1)} + e_T^{(2)} \otimes a^{(2)} + \tau_T^{(1)} \otimes \beta^{(1)} + \tau_T^{(2)} \otimes \beta^{(2)} + \zeta,$$

where e_T is a $T \times 1$ vector of unities, $\tau_{Tt} = t$ for $t = 1, ..., T$, $a^{(1)}$ and $\alpha^{(2)}$ are the vectors of individual effects defined in (2), $\beta^{(1)}$ and $\beta^{(2)}$ are $N \times 1$ vectors collecting slope coefficients $\beta_i^{(1)}$ and $\beta_i^{(2)}$, respectively, and $\tau_T^{(1)}$ and $\tau_T^{(2)}$ are $T \times 1$ vectors whose elements are defined as follows: $\tau_{Tt}^{(1)} = t$ if $t \leq \lambda$ and 0 otherwise, and $\tau_{Tt}^{(2)} = t$ if $t > \lambda$ and 0 otherwise.

For model (19), the annihilator matrix of estimator $\hat{\varphi}^{(\lambda)}$ becomes $Q^{(\lambda)} = I_{NT} - X^{(\lambda)}(X^{(\lambda)\prime}X^{(\lambda)})^{-1}X^{(\lambda)}$, where matrix $X^{(\lambda)}$, in addition to vectors $e^{(1)}$ and $e^{(2)}$, also includes the vectors of the linear trends components $\tau^{(1)}$ and $\tau^{(2)}$, with $\tau^{(j)} = \tau_T^{(j)} \otimes I_N$, for $j = 1, 2$, i.e., $X^{(\lambda)} = \left[e^{(1)}, e^{(2)}, \tau^{(1)}, \tau^{(2)}\right]$. Under H_0: $\varphi = 1$, the above specification of matrix $Q^{(\lambda)}$, apart from the initial observations of the panel, y_{i0}, renders estimator $\hat{\varphi}^{(\lambda)}$ invariant to individual effects β_i. As shown in Lemma A8, the inconsistency of this estimator is still given by relationship (4). However, variance-covariance matrix Γ_T can no longer be estimated consistently based on nonparametric estimator $\hat{\Gamma}_T = \frac{1}{N}\sum_{i=1}^{N} \Delta y_i^* \Delta y_i^{*\prime}$, since now Δy_{it} contains individual effects β_i under H_0: $\varphi =$

1, i.e., $\Delta y_{it} = \beta_i + u_{it}$. In this case, as shown in Lemma A9, we have $p\lim_N \left[\hat{\Gamma}_T - \Gamma_T - \beta_T^2 e_T e_T'\right] = 0$, where $\beta_T^2 = \frac{1}{N}\sum_{i=1}^N E(\beta_i^2)$ and consequently $p\lim_N \left[\frac{1}{N}tr(\Psi^{(\lambda)}\hat{\Gamma}) - \frac{1}{N}tr(\Psi^{(\lambda)}\Gamma)\right] \neq 0$ because $tr(\Psi_T^{(\lambda)} e_T e_T') \neq 0$.

To adjust $\hat{\varphi}^{(\lambda)}$ for its bias (following the reasoning of our previous section) we need to find a consistent estimator of $b^{(\lambda)} = tr(\Lambda' Q^{(\lambda)}\Gamma)$, based on a selection matrix of the diagonal and serially correlated elements of Γ_T. Let us denote this selection matrix $\Phi_T^{(\lambda)}$. Since now Γ_T can not be estimated consistently by $\hat{\Gamma}_T$, due to individual effects β_T^2, this selection matrix must also render the limiting distribution of the adjusted for its estimator $\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}$ net of effects β_T^2 (where $\Psi_T^{(\lambda)}$ fails). Thus, it must satisfy the following relationship:

$$p\lim_N \left[\frac{1}{N}tr(\Phi^{(\lambda)}\hat{\Gamma}) - \frac{1}{N}tr(\Lambda' Q^{(\lambda)}\Gamma)\right] = 0.$$

where $\Phi^{(\lambda)} = \Phi_T^{(\lambda)} \otimes I_N$ and matrix $\Phi_T^{(\lambda)}$ is analytically given as follows:

$$\Phi_T^{(\lambda)} = \Psi_T^{(\lambda)} - tr(\Psi_T^{(\lambda)} e_T e_T')\frac{M_T}{tr(M_T e_T e_T')},$$

where M_T is $T \times T$ selection matrix with elements $m_{ts} = 0$ if $\gamma_{tsT} \neq 0$ and $m_{ts} = 1$ if $\gamma_{tsT} = 0$. Note that matrix M_T selects the elements of matrix $(\Gamma_T + e_T e_T' \beta_T^2)$ which contain only $E(\beta_i^2)$. Based on M_T, we can derive a consistent estimator of β_T^2. This is given as $\frac{tr\left(M_T\hat{\Gamma}_T\right)}{tr(M_T e_T e_T')}$, i.e., $p\lim_N \left(\frac{tr\left(M_T\hat{\Gamma}_T\right)}{tr(M_T e_T e_T')} - \beta_T^2\right) = 0$ (see Lemma A10 for a proof).

In the above formula of selection matrix $\Phi_T^{(\lambda)}$, matrix $\Psi_T^{(\lambda)}$ selects the nonzero elements of matrix Γ_T to correct for the inconsistency of the numerator of estimator $\hat{\varphi}^{(\lambda)}$ due to the serial correlation effects and the within group transformation of the panel series y_{it} (as in the previous section). The second component of matrix $\Phi_T^{(\lambda)}$, i.e., $tr(\Psi_T^{(\lambda)} e_T e_T')\frac{M_T}{tr(M_T e_T e_T')}$, substitutes out nuisance parameter effects β_T^2 entering the limiting distribution of the adjusted estimator $\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}$, based on their consistent estimator $\frac{tr(M_T\hat{\Gamma}_T)}{tr(M_T e_T e_T')}$. Given the above formulas of selection matrices $\Phi^{(\lambda)}$ and $\Phi_T^{(\lambda)}$, next we derive the limiting distribution of a unit root test based on bias adjusted estimator $\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}$. This is based on Assumption B, made before, and Assumption C about β_i, given below. Assumption C is necessary because the individual effects appear in the statistic through the $\hat{\Gamma}$ and therefore, they must also obey the Lindeberg-Feller CLT.

Assumption C

$\beta_i^{(1)}$ and $\beta_i^{(2)}$ are random variables which are independent of u_{it} and across i, and have with finite $4+\epsilon$ moments. Also, we have $\lim_N \frac{\max(E(\beta_i^{(j)2}))}{N\beta_T^{(j)2}} = 0$ and $p\lim_N \beta_T^{(j)2} = p\lim_N \frac{1}{N}\sum_{i=1}^N E(\beta_i^{(j)2}) = \beta_{Tu}^{(j)2}$ is finite, for $j = 1, 2$.

Theorem 4 *Let Assumptions B and C hold. Then, under H_0: $\varphi = 1$ and λ known, we have*

$$Z^{(\lambda)} \equiv V^{(\lambda)-1/2}\hat{d}^{(\lambda)}\sqrt{N}\left(\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}} - 1\right) \xrightarrow{d} N(0,1) \tag{20}$$

as $N \to \infty$, where $\hat{d}^{(\lambda)} = \frac{1}{N}y_{-1}' Q^{(\lambda)} y_{-1}$, $\frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}$ is a consistent estimator of the asymptotic bias of $\hat{\varphi}^{(\lambda)}$, with

$\hat{b}^{(\lambda)} = \frac{1}{N} tr(\Phi^{(\lambda)}\hat{\Gamma})$ *where* $\hat{\Gamma} = \hat{\Gamma}_T \otimes I_N$ *and* $\Phi^{(\lambda)} = \Phi_T^{(\lambda)} \otimes I_N$, *and*

$$V^{(\lambda)} = \frac{1}{N}\tilde{F}^{(\lambda)\prime}\Theta\tilde{F}^{(\lambda)}, \tag{21}$$

where $\tilde{F}^{(\lambda)} = vec(Q^{(\lambda)}\Lambda - \Phi^{(\lambda)\prime})$ *and* $\Theta = Var(\Delta y \Delta y^{\prime})$.

Since model (19) under alternative hypothesis H_1: $\varphi < 1$ includes, in addition to individual effects $a^{(j)}$, $j = 1, 2$, linear trends, application of the test statistic $Z^{(\lambda)}$ given by Theorem 4 requires trimming out two time series observations from the end of the sample, i.e. $\lambda \in I = \{2, ..., T-2\}$. A useful expression for the estimation of variance $V^{(\lambda)}$ comes by noticing that $V^{(\lambda)} = \frac{1}{N}\tilde{F}^{(\lambda)\prime}\Theta\tilde{F}^{(\lambda)} = \tilde{F}_T^{(\lambda)\prime}\Theta_T\tilde{F}_T^{(\lambda)}$, where $\tilde{F}_T^{(\lambda)} = vec(Q_T^{(\lambda)}\Lambda_T - \Phi_T^{(\lambda)\prime})$, can be written as follows: $vec(Q_T^{(\lambda)}\Lambda_T - \Phi_T^{(\lambda)\prime})^{\prime}\Theta_T vec(Q_T^{(\lambda)}\Lambda_T - \Phi_T^{(\lambda)\prime})$, where $\Theta_T = \frac{1}{N}\sum_{i=1}^{N} Var(vec(\Delta y_i^{*}\Delta y_i^{*\prime})vec(\Delta y_i^{*}\Delta y_i^{*\prime})^{\prime})$. An estimator of variance-covariance matrix Θ_T is given as

$$\hat{\Theta}_T = \frac{1}{N}\sum_{i=1}^{N} vec(\Delta y_i^{*}\Delta y_i^{*\prime})vec(\Delta y_i^{*}\Delta y_i^{*\prime})^{\prime}. \tag{22}$$

Substituting $\hat{\Theta}_T$ into $vec(Q_T^{(\lambda)}\Lambda_T - \Phi_T^{(\lambda)\prime})^{\prime}\Theta_T vec(Q_T^{(\lambda)}\Lambda_T - \Phi_T^{(\lambda)\prime})$ yields a consistent estimator of $V^{(\lambda)}$ (see Lemma A12 and Remark A5). To implement the above test statistic in the case of unknown λ, we can follow an analogous procedure to that suggested in the previous sections. The elements of matrix $\Sigma \equiv [\sigma_{\mu s}]$ can be calculated based on following formula:

$$\sigma_{\mu s} = \frac{\tilde{F}_T^{(\mu)\prime}\Theta_T\tilde{F}_T^{(s)}}{\sqrt{\tilde{F}_T^{(\mu)\prime}\Theta_T\tilde{F}_T^{(\mu)}}\sqrt{\tilde{F}_T^{(s)\prime}\Theta_T\tilde{F}_T^{(s)}}},$$

by replacing Θ_T with $\hat{\Theta}_T$, given in (22).

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 H_0: $\varphi = 1$, consider the following extension of $AR(1)$ model (19):

$$\begin{aligned} y_{it} &= a_i^{(1)} + \beta_i^{(1)} t + \zeta_{it}, \quad \text{for } t \leq \lambda \\ y_{it} &= a_i^{(2)} + \beta_i^{(2)} t + \zeta_{it}, \quad \text{for } t > \lambda. \end{aligned} \tag{23}$$

In stacked vector notation, this model becomes

$$y = e_T^{(1)} \otimes a^{(1)} + e_T^{(2)} \otimes a^{(2)} + \tau_T^{(1)} \otimes \beta^{(1)} + \tau_T^{(2)} \otimes \beta^{(2)} + \zeta,$$

under both H_0: $\varphi = 1$ and H_1: $\varphi < 1$. It constitutes a more general specification than (19) because under H_0: $\varphi = 1$ restrictions $a_i^{(1)} = a_i^{(2)}$ and $\beta_i^{(1)} = \beta_i^{(2)}$ no longer apply. The presence of a break under H_0: $\varphi = 1$

reduces the maximum order of serial correlation of the error terms u_{it}, $p_{\max}$. This now is given as follows:[7]

$$p_{\max} = \begin{cases} \frac{T}{2} - 3, & \text{if } T \text{ is even and } \lambda = \frac{T}{2} \\ \min\{\lambda - 2, T - \lambda - 2\} & \text{in all other cases of } T \text{ or } \lambda \end{cases} \tag{24}$$

For model (23), LS estimator $\hat{\varphi}^{(\lambda)}$ is the same with that of model (19), i.e., matrix $Q^{(\lambda)}$ is defined a $Q^{(\lambda)} = I_{NT} - X^{(\lambda)}(X^{(\lambda)\prime}X^{(\lambda)})^{-1}X^{(\lambda)}$, with $X^{(\lambda)} = \left[e^{(1)}, e^{(2)}, \tau^{(1)}, \tau^{(2)}\right]$. Also, its bias function is given as before, i.e., $p\lim_N \left(\hat{\varphi}^{(\lambda)} - 1 - \frac{tr(\Lambda' Q^{(\lambda)}\Gamma)}{tr(\Lambda' Q^{(\lambda)}\Lambda\Gamma)}\right)$. One complication which arises with the break specification of the above AR(1) panel data model is that now we need to correct the liming distribution of the adjusted for its bias LS estimator $\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}$ for the following nuisance parameter effects: $\beta_T^{(1)2} = \frac{1}{N}\sum_{i=1}^{N} E(\beta_i^{(1)2})$ and $\beta_T^{(2)2} = \frac{1}{N}\sum_{i=1}^{N} E(\beta_i^{(2)2})$, which affect this limiting distribution before and after break point λ. This happens because, under H_0: $\varphi = 1$, model (23) implies $\Delta y_{it} = \beta_i^{(j)} + u_{it}$, $j = 1, 2$. The presence of effects $\beta_i^{(j)}$ renders $\hat{\Gamma}_T$ an inconsistent estimator of variance-covariance matrix Γ_T, $p\lim_N \left[\hat{\Gamma}_T - \Gamma_T - \beta_T^{(1)2} e_T^{(1)} e_T^{(1)\prime} - \beta_T^{(2)2} e_T^{(2)} e_T^{(2)\prime}\right] = 0$. To substitute out $\beta_T^{(1)2}$ and $\beta_T^{(2)2}$ from the limiting distribution of $\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}$, based on estimator $\hat{\Gamma}_T$, we need to employ the following selection matrix:

$$\Omega_T^{(\lambda)} = \Psi_T^{(\lambda)} - tr(\Psi_T^{(\lambda)} e_T^{(1)} e_T^{(1)\prime}) \frac{M_T^{(1)}}{tr(M_T^{(1)} e_T^{(1)} e_T^{(1)\prime})} - tr(\Psi_T^{(\lambda)} e_T^{(2)} e_T^{(2)\prime}) \frac{M_T^{(2)}}{tr(M_T^{(2)} e_T^{(2)} e_T^{(2)\prime})},$$

instead of $\Phi_T^{(\lambda)}$, where matrices $M_T^{(1)}$ and $M_T^{(2)}$ select the elements of matrix $(\Gamma_T + e_T^{(1)} e_T^{(1)\prime} \beta_T^{(1)2} + e_T^{(2)} e_T^{(2)\prime} \beta_T^{(2)2})$ consisting of effects $\beta_T^{(1)2}$ and $\beta_T^{(2)2}$, respectively. That is, the elements of matrix $M^{(1)}$ are defined as follows: $m_{ts}^{(1)} = 0$ if $\gamma_{tsT} \neq 0$, $m_{ts}^{(1)} = 1$ if $\gamma_{tsT} = 0$ for $t, s \leq \lambda$ and $m_{ts}^{(1)} = 0$ everywhere for t or $s > \lambda$, while those of $M_T^{(2)}$ as: $m_{ts}^{(2)} = 0$ if $\gamma_{tsT} \neq 0$, and $m_{2ts} = 1$ if $\gamma_{tsT} = 0$ for $t, s > \lambda$ and $m_{ts}^{(2)} = 0$ for t or $s \leq \lambda$. These two selection matrices provide the tools for estimating nuisance parameter effects $\beta_T^{(1)2}$ and $\beta_T^{(2)2}$ consistently, i.e., $p\lim_N \left(\frac{tr\left(M_T^{(1)}\hat{\Gamma}_T\right)}{tr(M_T^{(1)} e_T^{(1)} e_T^{(1)\prime})} - \beta_T^{(1)2}\right) = 0$ and $p\lim_N \left(\frac{tr\left(M_T^{(2)}\hat{\Gamma}_T\right)}{tr(M_T^{(2)} e_T^{(2)} e_T^{(2)\prime})} - \beta_T^{(2)2}\right) = 0$, respectively, and then substituting them out from the limiting distribution of the bias adjusted estimator $\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}$. Lemmas A13 - A16 prove that the above definition of matrix $\Omega_T^{(\lambda)}$ implies that

$$p\lim_N \left[\frac{1}{N} tr(\Omega^{(\lambda)}\hat{\Gamma}) - \frac{1}{N} tr(\Lambda' Q^{(\lambda)}\Gamma)\right] = 0,$$

where $\Omega^{(\lambda)} = \Omega_T^{(\lambda)} \otimes I_N$.

Based on the above formulas of selection matrices $\Omega^{(\lambda)}$ and $\Omega_T^{(\lambda)}$, the next theorem gives the limiting distribution of a unit root test statistic for model (23) based on $\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}$. This is derived under Assumptions B and C.

[7] Again, $p_{\max}$ is chosen so as our testing procedure to allow for some elements of variance-covariance matrix Γ_T to be zero, i.e., $E(u_{it}u_{is}) = 0$, for all $s = t + p_{\max} + 1, ..., T$. This condition means that variance function $V^{(\lambda)}$ will be different than zero. If T is even, then $p_{\max} = \min\{\lambda - 2, T - T_0 - 2\}$, with the exception the case that $\lambda = \frac{T}{2}$ where $p_{\max} = \frac{T}{2} - 3$. To see this more clearly, consider the following examples. First, $T = 10$ and $T_0 = 3$, then we have that $p_{\max} = \min\{\lambda - 2, T - \lambda - 2\} = \min\{1, 5\} = 1$. If $\lambda = \frac{T}{2} = 5$, then $p_{\max}$ becomes $p_{\max} = \frac{T}{2} - 3 = 2$. Note that, instead of the above, if we use the results of (24) to determine $p_{\max}$, implying $p_{\max} = \min\{\lambda - 2, T - \lambda - 2\} = \min\{3, 3\} = 3$, then $Z^{(\lambda)}$ could not be applied since $V^{(\lambda)} = 0$. If $T = 15$, then $p_{\max}$ becomes $p_{\max} = \min\{\lambda - 2, T - \lambda - 2\}$. For $\lambda = 7$, this becomes $p_{\max} = \min\{5, 6\} = 5$.

Theorem 5 *Let Assumptions B and C hold. Then, under H_0: $\varphi = 1$ and λ known, we have*

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

as $N \to \infty$, where $\hat{d}^{(\lambda)} = \frac{1}{N}y'_{-1}Q^{(\lambda)}y_{-1}$, $\frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}$ is a consistent estimator of the asymptotic bias of $\hat{\varphi}^{(\lambda)}$, with $\hat{b}^{(\lambda)} = \frac{1}{N}tr(\Omega^{(\lambda)}\hat{\Gamma})$ where $\hat{\Gamma} = \hat{\Gamma}_T \otimes I_N$ and $\Omega^{(\lambda)} = \Omega_T^{(\lambda)} \otimes I_N$, and

$$V^{(\lambda)} = \frac{1}{N}\tilde{F}^{(\lambda)\prime}\Theta\tilde{F}^{(\lambda)},$$

where $\tilde{F}^{(\lambda)} = vec(Q^{(\lambda)}\Lambda - \Omega^{(\lambda)\prime})$ and $\Theta = Var(\Delta y \Delta y')$.

The variance $V^{(\lambda)}$ involved in test statistic $Z^{(\lambda)}$ given by Theorem 5 can be consistently estimated following analogous steps to those discussed after Theorem 4, for model (19). If break point λ is unknown, then implementation of this test statistic requires a consistent estimator of λ, in a first step. This can be obtained under H_0: $\varphi = 1$ based on the first differences of panel data series y_{it}, i.e., $\Delta y_{it} = \beta_i^{(j)} + u_{it}$, for $j = 1, 2$ (see Bai (2010)). This procedure provides consistent estimates λ converging at $o(\sqrt{N})$ rate.[8]

5 Non-linear Trends and Multiple Breaks

In this section, we suggest extensions of our tests to allow for the presence of quadratic trends (see, e.g., Harvey et al. (2011)) and multiple breaks (see, e.g., Bai and Carrion (2009)) in AR(1) panel data model (1). Existing panel unit root tests allowing for structural breaks do not consider non-linear trends, like the quadratic one often assumed in explosive macroeconomic series, in the data generating process. As the time dimension of the panel T is finite, the number of breaks, the form of non-linearity and the maximum order of serial correlation allowed by our model (1) will depend on T. Thus, to implement the above extensions of our tests, a set of regulatory sufficient conditions are needed.

5.1 Non-linear Individual Trends

Consider the following version of model (1) with linear and quadratic trends and a break under H_0: $\varphi = 1$:

$$\begin{aligned} y_{it} &= a_i^{(1)} + \beta_i^{(1)}t + \delta_i^{(1)}t^2 + \zeta_{it}, \quad \text{for } t \leq \lambda \\ y_{it} &= a_i^{(2)} + \beta_i^{(2)}t + \delta_i^{(2)}t^2 + \zeta_{it}, \quad \text{for } t > \lambda. \end{aligned} \tag{25}$$

In stacked vector notation, the above model can be written as

$$y = \sum_{j=1}^{2} e_T^{(j)} \otimes a^{(j)} + \sum_{j=1}^{2} \tau_T^{(j)} \otimes \beta^{(j)} + \sum_{j=1}^{2} \tau_{2T}^{(j)} \otimes \delta^{(j)} + \zeta,$$

[8] Alternatively, we can also exploit break point (date) estimation methods recently proposed in the single time series literature by Harvey and Leybourne (2013).

where $\tau^{(j)}_{2Tt} = (\tau^{(j)}_{Tt})^2$, for all t, and $\delta^{(j)} = (\delta^{(j)}_1, ..., \delta^{(j)}_N)'$, $j = 1, 2$. For the above model, the annihilator matrix of estimator $\hat{\varphi}^{(\lambda)}$ becomes $Q^{(\lambda)} = I_{NT} - X^{(\lambda)}(X^{(\lambda)\prime}X^{(\lambda)})^{-1}X^{(\lambda)}$, where $X^{(\lambda)} = [e^{(1)}, e^{(2)}, \tau^{(1)}, \tau^{(2)}, \tau^{(1)}_2, \tau^{(2)}_2]$, with $\tau^{(j)}_2 = \tau^{(j)}_{2T} \otimes I_N$, $j = 1, 2$, and $\tau^{(j)}_{2T}$ are $T \times 1$ dimension vectors with elements $\tau^{(j)}_{2Tt}$.

The presence of quadratic trends t^2 in panel data model (25) requires, in addition to nuisance parameters $\beta^{(j)}_i$, correction of the limiting distribution of LS estimator $\hat{\varphi}^{(\lambda)}$ for the presence of nuisance parameters associated with these quadratic trends, i.e., $\delta^{(j)}_i$. As for the model with linear trends (23), the bias of estimator $\hat{\varphi}^{(\lambda)}$ is given by $p\lim_N \left(\hat{\varphi}^{(\lambda)} - 1 - \frac{tr(\Lambda' Q^{(\lambda)}\Gamma)}{tr(\Lambda' Q^{(\lambda)}\Lambda\Gamma)}\right)$(see Lemma A17) and $\hat{\Gamma}_T$ constitutes an inconsistent estimator of Γ_T, since $\Delta y_{it} = \beta^{(j)}_i + (2t-1)\delta^{(j)}_i$ under H_0: $\varphi = 1$, for $j = 1, 2$. The inconsistency of $\hat{\Gamma}_T$ is given by $p\lim_N \left[\hat{\Gamma}_T - \Gamma_T - \sum_{j=1}^2 \beta^{(j)2}_T e^{(j)}_T e^{(j)\prime}_T - \sum_{j=1}^2 \delta^{(j)2}_T \tilde{e}^{(j)}_T \tilde{e}^{(j)\prime}_T\right] = 0$ (see Lemma A18), where $\tilde{e}^{(j)}_T = (2\tau^{(j)}_T - e^{(j)}_T)$ and $\delta^{(j)2}_T = \frac{1}{N}\sum_{i=1}^N E(\delta^{(j)2}_i)$, for $j = 1, 2$.

Consistent estimators of the nuisance parameter effects $\beta^{(j)2}_T$ and $\delta^{(j)2}_T$, defined above, can be obtained by the following relationships:

$$p\lim_N \left[\frac{tr\left(J^{(j)}_T \hat{\Gamma}_T\right)}{tr(M^{(j)}_T e^{(j)}_T e^{(j)\prime}_T)} - \beta^{(j)2}_T\right] = 0 \quad \text{and} \quad p\lim_N \left[\frac{tr(L^{(j)}_T \hat{\Gamma}_T)}{tr(L^{(j)}_T \tilde{e}^{(j)}_T \tilde{e}^{(j)\prime}_T)} - \delta^{(j)2}_T\right] = 0, \text{ for } j = 1, 2,$$

respectively, where $J^{(j)}_T = M^{(j)}_T - \frac{tr(M^{(j)}_T \tilde{e}^{(j)}_T \tilde{e}^{(j)\prime}_T)}{tr\left(L^{(j)}_T \tilde{e}^{(j)}_T \tilde{e}^{(j)\prime}_T\right)} L^{(j)}_T$. $L^{(j)}_T$ is a selection matrix whose its *k-th* line has values that interchange between -1 and 1 for the non zero elements of matrix $M^{(j)}_T$ (i.e., $m^{(j)}_{ts} \neq 0$) and as long as both values -1 and 1 appear consecutively. This also applies for the *k-th* column of $L^{(j)}_T$. Matrix $L^{(j)}_T$ enables us to identify the nuisance parameter effects $\delta^{(j)2}_T$, due to the presence of quadratic trends in (25), from those coming from the presence of the linear trends, i.e., $\beta^{(j)2}_T$. Both of these sets of parameters appear simultaneously in the zero off-diagonal elements of matrix $\Gamma_T p\lim_N \hat{\Gamma}_T$. To identify them, matrix $L^{(j)}_T$ exploits the fact that they appear with different weights in $p\lim_N \hat{\Gamma}_T$.[9] These weights are given by matrices $e^{(j)}_T$ and $\tilde{e}^{(j)}_T = (2\tau^{(j)}_T - e^{(j)}_T)$, respectively.[10]

---

[9] To see more clearly how $\delta^{(j)2}_T$ and $\beta^{(j)2}_T$ can be identified based on the zero off-diagonal elements of matrix $p\lim_N \hat{\Gamma}_T$, consider the following example. For $\lambda = 5$ and, for simplicity, $p = 0$, the upper 5×5 block of $p\lim_N \hat{\Gamma}_T$ is given as

$$\begin{pmatrix}
\gamma_{11T} + \delta^{(1)2}_T + \beta^{(1)2}_T & 3\delta^{(1)2}_T + \beta^{(1)2}_T & 5\delta^{(1)2}_T + \beta^{(1)2}_T & 7\delta^{(1)2}_T + \beta^{(1)2}_T & 9\delta^{(1)2}_T + \beta^{(1)2}_T \\
3\delta^{(1)2}_T + \beta^{(1)2}_T & \gamma_{22T} + 9\delta^{(1)2}_T + \beta^{(1)2}_T & 15\delta^{(1)2}_T + \beta^{(1)2}_T & 21\delta^{(1)2}_T + \beta^{(1)2}_T & 27\delta^{(1)2}_T + \beta^{(1)2}_T \\
5\delta^{(1)2}_T + \beta^{(1)2}_T & 15\delta^{(1)2}_T + \beta^{(1)2}_T & \gamma_{33T} + 25\delta^{(1)2}_T + \beta^{(1)2}_T & 35\delta^{(1)2}_T + \beta^{(1)2}_T & 45\delta^{(1)2}_T + \beta^{(1)2}_T \\
7\delta^{(1)2}_T + \beta^{(1)2}_T & 21\delta^{(1)2}_T + \beta^{(1)2}_T & 35\delta^{(1)2}_T + \beta^{(1)2}_T & \gamma_{44T} + 49\delta^{(1)2}_T + \beta^{(1)2}_T & 63\delta^{(1)2}_T + \beta^{(1)2}_T \\
9\delta^{(1)2}_T + \beta^{(1)2}_T & 27\delta^{(1)2}_T + \beta^{(1)2}_T & 45\delta^{(1)2}_T + \beta^{(1)2}_T & 63\delta^{(1)2}_T + \beta^{(1)2}_T & \gamma_{55T} + 81\delta^{(1)2}_T + \beta^{(1)2}_T
\end{pmatrix}$$

From this, it can be easily seen that parameters $\delta^{(1)2}_T$ and $\beta^{(1)2}_T$ have different weights given by the matrices $e^{(1)}_T$ and $\tilde{e}^{(1)}_T = (2\tau^{(1)}_T - e^{(1)}_T)$. If elements $(1,2)$ and $(1,3)$ of the above matrix are selected and subtracted from each other, then we obtain

$$(1)\left(5\delta^{(1)2}_T + \beta^{(1)2}_T\right) + (-1)\left(3\delta^{(1)2}_T + \beta^{(1)2}_T\right) = 2\delta^{(1)2}_T. \tag{26}$$

Thus, $\delta^{(1)2}_T$ can be identified by exploiting linear combinations of moments. This can be done through selection matrix $L^{(1)}_T$. Given $\delta^{(1)2}_T$, $\beta^{(1)2}_T$ can be found from the off diagonal elements of Γ_T.

[10] As an illustrative example of matrix $L^{(1)}_T$, assume that $\lambda = 4$ and $p = 1$. Then, the upper left block of $M^{(1)}_T$ is given as $\begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}$ while its lower right block is zero. Then, the upper left block matrix $L^{(1)}_T$ is given as $\begin{pmatrix} 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$.

Given the above estimators of nuisance parameter effects $\beta_T^{(j)2}$ and $\delta_T^{(j)2}$, the selection matrix which can be employed to adjust estimator $\hat{\varphi}^{(\lambda)}$ for its inconsistency due to the with transformation of the data, the serial correlation effects in error terms u_{it} and the presence of the linear and quadratic trends is given as follows:

$$\Xi_T^{(\lambda)} = \Psi_T^{(\lambda)} - \sum_{j=1}^{2} \frac{tr\left(\Psi_T^{(\lambda)} e_T^{(j)} e_T^{(j)\prime}\right)}{tr\left(M_T^{(j)} e_T^{(j)} e_T^{(j)\prime}\right)} J_T^{(j)} - \sum_{j=1}^{2} \frac{tr(\Psi_T^{(\lambda)} \tilde{e}_T^{(j)} \tilde{e}_T^{(j)\prime})}{tr(L_T^{(j)} \tilde{e}_T^{(j)} \tilde{e}_T^{(j)\prime})} L_T^{(j)}.$$

Under H_0: $\varphi = 1$, it can be easily shown (see Lemma A19) that matrix $\Xi_T^{(\lambda)}$ implies

$$p\lim_N \left[\frac{1}{N} tr(\Xi^{(\lambda)}\hat{\Gamma}) - \frac{1}{N} tr(\Lambda' Q^{(\lambda)}\Gamma)\right] = 0,$$

where $\Xi^{(\lambda)} = \Xi_T^{(\lambda)} \otimes I_N$, which means that $\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}$, with $\hat{b}^{(\lambda)} = \frac{1}{N} tr(\Xi^{(\lambda)}\hat{\Gamma})$, constitutes a consistent estimator of φ. Given the above formulas of matrices $\Xi_T^{(\lambda)}$ and $\Xi^{(\lambda)}$, the next theorem derives the limiting distribution of a unit root test statistic based on $\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}$. In addition to assumptions B and C, this derivation also requires the following assumption.

Assumption D

$\delta_i^{(1)}$and $\delta_i^{(2)}$ are random variables which are independent of u_{it} and across i, and have finite $4+\epsilon$ moments. Also, we have $\lim_N \frac{\max(E(\delta_i^{(j)2}))}{N\delta_T^{(j)2}} = 0$ and $p\lim_N \delta_T^{(j)2} = p\lim_N \frac{1}{N}\sum_{i=1}^N E(\delta_i^{(j)2}) = \delta_{Tu}^{(j)2}$ which is finite, for $j = 1, 2$.

Theorem 6 *Let Assumptions B, C and D hold. Then, under H_0: $\varphi = 1$ and λ known, we have*

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

as $N \to \infty$, where $\hat{d}^{(\lambda)} = \frac{1}{N} y_{-1}' Q^{(\lambda)} y_{-1}$, $\frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}$ is a consistent estimator of the asymptotic bias of $\hat{\varphi}^{(\lambda)}$, with $\hat{b}^{(\lambda)} = \frac{1}{N} tr(\Xi^{(\lambda)}\hat{\Gamma})$ where $\hat{\Gamma} = \hat{\Gamma}_T \otimes I_N$ and $\Xi^{(\lambda)} = \Xi_T^{(\lambda)} \otimes I_N$, and

$$V^{(\lambda)} = \frac{1}{N} \tilde{F}^{(\lambda)\prime} \Theta \tilde{F}^{(\lambda)},$$

where $\tilde{F}^{(\lambda)} = vec(Q^{(\lambda)}\Lambda - \Xi^{(\lambda)\prime})$ and $\Theta = Var(\Delta y \Delta y')$.

To implement the test statistic $Z^{(\lambda)}$ given by Theorem 6, variance $V^{(\lambda)}$ can be estimated following the same procedure with that of the test statistic given by Theorem 4. The case of test statistic $Z^{(\lambda)}$ with no break under H_0: $\varphi = 1$ (see model (19)) can be treated similarly, based on consistent estimators of nuisance

Note that the elements -1 and 1 of matrix $L_T^{(1)}$ appear always together so that they subtract the proper moments to identify nuisance parameters $\beta_i^{(j)2}$ and $\delta_i^{(j)2}$, for $j = 1, 2$ (see also (26)). The elements (2,4) and (4,2) of the upper blocks are non-zero in $M_T^{(1)}$, but they cannot be non-zero in $L_T^{(1)}$ as there are not enough moments to pair 1 and -1. If $\lambda = 5$, then $M_T^{(1)}$ becomes $\begin{pmatrix} 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{pmatrix}$ and $L_T^{(1)}$ is given as $\begin{pmatrix} 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ -1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$.

parameter effects $\beta_T^2 = \frac{1}{N}\sum_{i=1}^{N} E(\beta_i^2)$.and $\delta_T^2 = \frac{1}{N}\sum_{i=1}^{N} E(\delta_i^2)$. If the break is unknown, then it must be estimated under H_0: $\varphi = 1$, as in the case of model (23). If a break occurs only under H_1: $\varphi < 1$, then the sequential testing procedure suggested for model (19) can be applied. This can be done after trimming out four time series observations from the end of the sample and three from the start, i.e., $\lambda \in I = \{3, ..., T-4\}$.

5.2 Multiple Breaks

There is always the possibility that more than one breaks occur in a time span, even if the time dimension of the panel T is small. Structural changes also may not affect all the parameters of the model (see, e.g., Bai and Perron (1998)). This section provides extensions of our tests to the above directions. Let S_b be the number of structural breaks occurring during our sample, then models (1), (23) and (25) can be respectively rewritten as

$$
\begin{aligned}
y &= \sum_{j=1}^{S_b+1} e_T^{(j)} \otimes a^{(j)} + \zeta \\
y &= \sum_{j=1}^{S_b+1} e_T^{(j)} \otimes a^{(j)} + \sum_{j=1}^{S_b+1} \tau_T^{(j)} \otimes \beta^{(j)} + \zeta \quad \text{and} \\
y &= \sum_{j=1}^{S_b+1} e_T^{(j)} \otimes a^{(j)} + \sum_{j=1}^{S_b+1} \tau_T^{(j)} \otimes \beta^{(j)} + \sum_{j=1}^{S_b+1} \tau_{2T}^{(j)} \otimes \delta^{(j)} + \zeta.
\end{aligned}
$$

Given S_b, our sequential testing procedure for panel data unit roots described in the previous sections can be easily extended to the case of mutlibreaks, by appropriately specifying annihilator matrix $Q^{(\lambda)}$ to allow for $S_b > 1$ break points. This matrix will be henceforth denoted as $Q^{(\lambda_1,...,\lambda_{S_b})}$. For instance, if $S_b = 2$ and model (1) constitutes the correct data generating process, then $Q^{(\lambda_1,...,\lambda_{S_b})}$ can be defined as $Q^{(\lambda_1,\lambda_2)} = I_{NT} - X^{(\lambda_1,\lambda_2)}(X^{(\lambda_1,\lambda_2)\prime}X^{(\lambda_1,\lambda_2)})^{-1}X^{(\lambda_1,\lambda_2)}$, with $X^{(\lambda_1,\lambda_2)} = [e_T^{(1)}, e_T^{(2)}, e_T^{(3)}] \otimes I_N$, $e_{Tt}^{(1)} = 1$ if $t \leq \lambda_1$ and 0 otherwise, $e_{Tt}^{(2)} = 1$ if $\lambda_1 < t < \lambda_2$ and 0 otherwise, and $e_{Tt}^{(3)} = 1$ if $\lambda_2 < t$, and 0 otherwise. Note that, in the above definitions, the different breaks can happen at any time point of the sample $\lambda_1, ..., \lambda_{S_b}$, as long as matrix $Q^{(\lambda_1,...,\lambda_{S_b})}$ exists, or equivalently $X^{(\lambda_1,...,\lambda_{S_b})\prime}X^{(\lambda_1,...,\lambda_{S_b})}$ is invertible. In the model with intercepts, this assumption permits for consecutive breaks, while in the model with individual linear trends it does not.[11]

The above specification of matrix $X^{(\lambda_1,...,\lambda_{S_b})}$ wipes off the deterministic components of the panel data series and the theorems of the previous sections for the case of a single break point can be applied. The within group LS estimator of φ will now be defined as

$$
\hat{\varphi}^{(\lambda_1,...,\lambda_{S_b})} = \left[y_{-1}' Q^{(\lambda_1,...,\lambda_{S_b})} y_{-1}\right]^{-1} \left[y_{-1}' Q^{(\lambda_1,...,\lambda_{S_b})} y\right],
$$

while test statistic $Z^{(\lambda)}$ as $Z^{(\lambda_1,...,\lambda_{S_b})}$. A notable difference is when the dates of the S_b break points are unknown. Then, minimization of $Z^{(\lambda_1,...,\lambda_{S_b})}$ happens over all possible combinations of break points,

[11] This has been also found by Lumsdaine and Papell (1997) for their single time series unit root tests.

$\lambda_1, ..., \lambda_{S_b} \in I$. The limiting distribution of the minimum value of $Z^{(\lambda_1,...,\lambda_{S_b})}$ is given as

$$\min_{\lambda_1,...,\lambda_{S_b} \in I} Z^{(\lambda_1,...,\lambda_{S_b})} \xrightarrow{d} \psi \equiv \min_{\lambda_1,...,\lambda_{S_b} \in I} N(0, \Sigma).$$

To calculate statistic $\min\limits_{\lambda_1,...,\lambda_{S_b} \in I} Z^{(\lambda_1,...,\lambda_{S_b})}$ for model (1), the number of all possible test statistics $Z^{(\lambda_1,...,\lambda_{S_b})}$ required will no longer be $T-2$, but it will be equal to the number of combinations without repetition $\binom{T-2}{S_b}$. For all these statistics, pairwise correlations can be computed as before and their asymptotic distribution can be described by expression (13).

In addition to the invertibility condition of $X^{(\lambda_1,...,\lambda_{S_b})\prime} X^{(\lambda_1,...,\lambda_{S_b})}$, mentioned above, implementation of test statistic $\min_{\lambda_1,...,\lambda_{S_b} \in I} Z^{(\lambda_1,...,\lambda_{S_b})}$ requires that the variance of the limiting distribution of the adjusted for its bias LS estimator $\hat{\varphi}^{(\lambda_1,...,\lambda_{S_b})}$, i.e., $V^{(\lambda_1,...,\lambda_{S_b})}$ exists and is different from zero. The last condition reflects all serial correlation restrictions and trend induced nuisance parameter identification restrictions affecting the estimator. Both of the above conditions can be easily checked, since they are based on deterministic matrices which can be defined before inference is conducted. Note however that these conditions must be checked for all sets of possible break points, separately. Existence of $V^{(\lambda_1,...,\lambda_{S_b})}$ is mostly important for specifications of the AR(1) panel data model including trends, i.e., (19), (23) and (25). If there are not enough moments for identification of the nuisance parameters of these specifications of the dynamic panel data model, then some of the denominators of selection matrices $\Phi_T^{(\lambda)}$, $\Omega_T^{(\lambda)}$ and $\Xi_T^{(\lambda)}$ will become 0. Thus, these matrices and, hence, variance $V^{(\lambda_1,...,\lambda_{S_b})}$ will no longer exist. If only intercepts are included in the model (see, e.g., model (1)), then $V^{(\lambda_1,...,\lambda_{S_b})}$ can be zero only in the case that the degree of serial correlation is assumed to be very large to make $\Lambda_T' Q_T^{(\lambda_1,...,\lambda_{S_b})} - \Psi_T^{(\lambda_1,...,\lambda_{S_b})} = 0$.

6 Spatial Dependence

In this section the assumption of cross-sectional independence is relaxed. Besides its significance in regional data, spatial dependence is also frequently used to capture weaker forms of cross sectional dependence by considering economic type distances.[12] The spill over effects across countries, states and regions can be captured through a spatial dependence structure. In a motivating paper, Baltagi et al. (2007) conduct extensive Monte Carlo experiments to show that panel unit root tests which do not take spatial dependence into account have considerable size distortions. This phenomenon will be further aggravated by ignoring the existence of structural breaks in the data panel data model.

Widely used forms of spatial dependence include the Spatial Autoregressive model (SAR), which is considered as a global dependence model, and the local dependence models of Spatial Moving Averages

[12] Applying methods which assume strong factors in the errors when dependence is weak can lead to misleading results, see e.g. Sarafidis and Wansbeek (2012). The same is true for the case of structural breaks in the factor loadings which increase the dimension of the factor space and which also result in the common factors becoming less important (see, e.g., Breitung and Eickmeier (2011)).

(SMA) and Spatial Error Components (SEC) (see Anselin (2003, 2007)) defined, respectively, as

$$\text{SAR:} \qquad u_{it} = \mu \sum_{j=1}^{N} w_{ij} u_{jt} + \varepsilon_{it}, \tag{27}$$

$$\text{SMA:} \qquad u_{it} = \mu \sum_{j=1}^{N} w_{ij} \varepsilon_{jt}, + \varepsilon_{it}, \tag{28}$$

$$\text{SEC:} \qquad u_{it} = \mu \sum_{j=1}^{N} w_{ij} \xi_{jt}, + \varepsilon_{it} \tag{29}$$

where w_{ij} are known parameters reflecting economic or geographic distances which can be bundled together in a $N \times N$ weighting matrix, denoted W.[13] Stacking the errors in equations (27)-(29) in the time dimension results in the following:

$$u_t = \mu W u_t + \varepsilon_t = (I_N - \mu W)^{-1} \varepsilon_t \tag{30}$$

$$u_t = \mu W \varepsilon_t + \varepsilon_t = (I_N + \mu W) \varepsilon_t \tag{31}$$

$$u_t = \mu W \xi_t + \varepsilon_t. \tag{32}$$

For exposition purposes, our analysis will be focused on SAR and SMA models, as the SEC model cannot be written in a similar way. However the proposed statistics are valid for this case as well. A major advantage of our panel unit root tests for the above models is that they are robust to the type of spatial dependence and to the form of matrix W considered due to the non-parametric estimator $\hat{\Gamma}$. Define matrix $\Pi_N = (I_N - \mu W)^{-1}$ for the SAR model and $\Pi_N = (I_N + \mu W)$ for the SMA model. For weights w_{ij} and spatial correlation parameter μ consider the following assumption:

Assumption E

(i) $\{\varepsilon_{it}\}$ are independent random variables across i and t, with $E(u_{it}) = 0$ and uniformly bounded $4 + \epsilon$ moments.

(ii) ξ_{jt} are IID error terms with variance σ_ξ^2 independent of ε_{it}, for all i and t.

(iii) The weighting matrix W has zeros in its main diagonal.

(iv) The spatial correlation parameter satisfies $\mu \in (-c_{1,\mu}, c_{2,\mu})$ with $-\infty < -c_\mu < -c_{1,\mu}, c_{2,\mu} < c_\mu < \mu$.

(v) The $N \times N$ matrix Π_N exists and is non-singular for all $\mu \in (-c_{1,\mu}, c_{2,\mu})$.

(vi) The row and column sums of matrices Π_N, for the SAR and SMA models, are bounded uniformly in absolute value.

Condition (i) implies that error terms ε_{it} are independent across i and t, but they are allowed to be heteroscedastic and heterogeneous. Condition (ii) is standard in the spatial dependence literature (see also

[13] To see how ignoring a structural break can exacerbate spatial dependence, consider doing so in the model (1) with SAR errors. Then u_{it} become $\breve{u}_{it} = (1-\varphi)a_i^{(2)} + \mu \sum_{j=1}^{N} w_{ij} \breve{u}_{jt} + \varepsilon_{it} = (1-\varphi)a_i^{(2)} + \mu \sum_{j=1}^{N} w_{ij}(1-\varphi)a_j^{(2)} + \mu \sum_{j=1}^{N} w_{ij} u_{jt} + \varepsilon_{it}$. The term $\mu \sum_{j=1}^{N} w_{ij}(1-\varphi)a_j^{(2)}$ captures the spatial transmission of the break into neighbouring units.

Kelejian and Prucha (2010)). Condition (iii) gives a specific normalization to weight matrix W. Conditions (iv) and (v) provide conditions for invertibility of Π_N. Condition (vi) implies that there is no dominant cross section unit in the sample, i.e., an individual unit which is correlated with all remaining units (see also Sarafidis (2009)).

Under the conditions of Assumption E, it can be seen that $E(uu') = E(\Pi\varepsilon\varepsilon'\Pi') = \Pi\Gamma^{\varepsilon}\Pi'$, where $\Gamma^{\varepsilon} = E(\varepsilon\varepsilon')$ and $\Pi = I_T \otimes \Pi_N$. The inconsistency of LS estimator $\hat{\varphi}^{(\lambda)}$ is given as

$$p\lim_{N}\left[\hat{\varphi}^{(\lambda)} - 1 - \frac{tr\left(\Lambda' Q^{(\lambda)}\Pi\Gamma^{\varepsilon}\Pi'\right)}{tr\left(\Lambda' Q^{(\lambda)}\Lambda\Pi\Gamma^{\varepsilon}\Pi'\right)}\right] = 0 \tag{33}$$

This reduces to that given by (4), if $\mu = 0$. Spatial dependence enters the bias function of $\hat{\varphi}^{(\lambda)}$ through matrix Π. To see more clearly how estimator $\hat{\varphi}^{(\lambda)}$ can be adjusted for its bias, consider model (1) under the assumption of no serial correlation of error terms u_{it}. The more general specifications of this model, presented in the previous sections (see, (19), (23) and (25)), can be analyzed along the same lines.

For model (1), first note that the non-parametric variance-covariance estimator $\hat{\Gamma}$, defined in the previous sections as $\hat{\Gamma} = \hat{\Gamma}_T \otimes I_N$, has the following property:

$$p\lim_{N}\left[\frac{1}{N}tr(\hat{\Gamma}) - \frac{1}{N}tr\left(\Pi\Gamma^{\varepsilon}\Pi'\right)\right] = 0. \tag{34}$$

This means that, under H_0: $\varphi = 1$, the bias coming from the spatially dependent error terms u_{it} can be captured by the first difference of y_{it}, given as $\Delta y_{it} = u_{it}$, for model (1). This happens because $E(\Delta y \Delta y') = E(uu') = \Pi\Gamma^{\varepsilon}\Pi$. This result implies that selection matrix $\Psi^{(\lambda)}$ (which has non-zero elements in its main diagonal) can be also employed to adjust LS estimator $\hat{\varphi}^{(\lambda)}$ for its inconsistency due to spatial correlation effects. It is straightforward to show that

$$p\lim_{N}\left[\frac{1}{N}tr(\Psi^{(\lambda)}\hat{\Gamma}) - \frac{1}{N}tr\left(\Lambda' Q^{(\lambda)}\Pi\Gamma^{\varepsilon}\Pi'\right)\right] = 0.$$

Analogous results hold for our more general specifications of model (1), mentioned above. For instance, consider model (19), which allows for serial correlation effects. For this model, we have

$$E(\Delta y \Delta y') = E[(u + e_T \otimes \beta)(u + e_T \otimes \beta)'] = E[uu' + e_T e_T' \otimes \beta\beta'] = \Pi\Gamma^{\varepsilon}\Pi' + e_T e_T' \otimes E(\beta\beta')$$

and thus,

$$p\lim_{N}\left[\frac{1}{N}tr(\hat{\Gamma}) - \frac{1}{N}tr\left(\Pi\Gamma^{\varepsilon}\Pi\right) - \frac{1}{N}tr(e_T e_T' \otimes E(\beta\beta'))\right] = 0.$$

The last result indicates that, by employing selection matrix $\Phi^{(\lambda)}$, which annihilates nuisance parameter effects $e_T e_T' \otimes E(\beta\beta')$, we can have that

$$p\lim_{N}\left[\frac{1}{N}tr(\Phi^{(\lambda)}\hat{\Gamma}) - \frac{1}{N}tr\left(\Lambda' Q^{(\lambda)}\Pi\Gamma^{\varepsilon}\Pi\right)\right] = 0.$$

This means that, for model (19), the mean of the limiting distribution of adjusted LS estimator $\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}$ based on selection matrix $\Phi^{(\lambda)}$ is not only net of incidental parameter effects, but also of spatial correlation effects.

The next theorem provides the limiting distribution of test statistic $Z^{(\lambda)}$ for the simple version of model (1), without serially correlated errors, and the case of a known break point λ. As our analysis above shows, analogous formulas of this test statistic can be obtained for the more general specifications of model (1), by choosing appropriately the selection matrix annihilating the nuisance parameters of these models.

Theorem 7 *Let Assumption E hold. Then, under H_0: $\varphi = 1$ and λ known, we have*

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

as $N \to \infty$, where $\frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}} \equiv \frac{\frac{1}{N}tr\left(\Psi^{(\lambda)}\hat{\Gamma}\right)}{\frac{1}{N}y_{-1}'Q^{(\lambda)}y_{-1}}$ is a consistent estimator of the asymptotic bias of $\hat{\varphi}^{(\lambda)}$, with $\hat{b}^{(\lambda)} = \frac{1}{N}tr\left(\Psi^{(\lambda)}\hat{\Gamma}\right)$ and $\hat{d}^{(\lambda)} = \frac{1}{N}y_{-1}'Q^{(\lambda)}y_{-1}$, and

$$V^{(\lambda)} = \frac{1}{N}2tr(\tilde{F}^{(\lambda)}\Gamma^{\varepsilon}\tilde{F}^{(\lambda)}\Gamma^{\varepsilon}),$$

where $\tilde{F}^{(\lambda)} = \frac{1}{2}\left(\Pi'\left(\Lambda'Q^{(\lambda)} - \Psi^{(\lambda)}\right)\Pi + \Pi'\left(Q^{(\lambda)}\Lambda - \Psi^{(\lambda)\prime}\right)\Pi\right)$.

The results of Theorem 7 imply that the implementation of test statistic $Z^{(\lambda)}$ is not straightforward in practice. This happens because the variance function of this limiting distribution, $V^{(\lambda)}$, can not be consistently estimated based on estimator $\hat{\Gamma}$. As can be seen from (34), $\hat{\Gamma}$ cannot estimate Γ^{ε}, separately, but the following product of matrices: $\Pi\Gamma^{\varepsilon}\Pi$. Furthermore, $\tilde{F}^{(\lambda)}$ is not known as in the previous sections, because it includes Π which contains the spatial correlation parameter μ.

For the above reasons, we suggest implementing test statistic $Z^{(\lambda)}$ based on the bootstrap method. In particular, given the cross sectional dependence of error terms u_{it}, a block bootstrap should be applied. The idea is to resample blocks of units of u_{it} taking into account the spatial dependence between them. There is a large literature concerning the block bootstrap and its variations, i.e. the block size, whether it is fixed or random, and whether blocks are overlapping or not (see, e.g., Hall (1985) and Anselin (1990) for spatial dependence and Basawa et al. (1991)) for unit root tests).

To prove the consistency of the bootstrap method for our test statistic $Z^{(\lambda)}$, given by Theorem 7, we follow Horowitz (2001). Define F_0 the $T-$dimensional multivariate cdf from which the data come and let F_0 belong to a family of distributions J. Let $G_N(\pi, F_0) \equiv P(Z^{(\lambda)} \leq \pi)$ denote the exact, finite sample cdf of $Z^{(\lambda)}$ and let $G_{\infty}(\cdot, F_0)$ denote its asymptotic distribution. The bootstrap replaces F_0 with its estimator F_N (such as the empirical distribution function or a parametric estimator) so the bootstrap estimator of $G_N(\cdot, F_0)$ becomes $G_N(\cdot, F_N)$. This estimation procedure would be consistent if $p\lim_N G_N(\cdot, F_N) = G_{\infty}(\cdot, F_0)$. Then, $G_{\infty}(\cdot, F_0)$ can be used to approximate $G_N(\cdot, F_0)$.

Theorem 8 *Let the sequence $\{y_{it}\}$ be generated according to model (1) and Assumption E hold. Let P_N denote the joint probability distribution of the sample. Then, under H_0: $\varphi = 1$ and λ known, the bootstrap* *estimator $G_N(\cdot, F_N)$ is consistent, i.e., for each $\delta > 0$ and $F_0 \in J$, we have*

$$\lim_{N\to\infty} P_N \left[\sup_{\pi} |G_N(\pi, F_N) - G_\infty(\pi, F_0)| > \delta\right] = 0.$$

There are important theoretical advances regarding the specification of the bootstrap (see Horowitz (2001) for a review). Hall et al. (1995) show that overlapping blocks are more efficient than non-overlapping. They also show that, under certain assumptions, the optimal block length when estimating a one sided distribution function (e.g., $P_N(Z^{(\lambda)} \leq \pi)$) is $O(N^{1/4})$. Lahiri (1999) shows that the fixed blocks are preferable compared to random blocks (see Politis and Romano (1994)).

The block bootstrap preserves the spatial structure of the data. First a $b \times T$ block is selected, then a second $b \times T$ block and so on, until the final jth block is selected, where $j \cdot b = N$. All these blocks, put together, create a bootstrap sample.[14] The full bootstrap procedure can be outlined as follows:

1. Use the data to compute test statistic $Z^{(\lambda)}$, defined by Theorem 7.

2. Generate B bootstrap samples of size $N \times T$ where each bootstrap sample is composed by smaller blocks. Sampling is done with replacement from the residuals u^B where $u^B = \Delta y$. This applies for all specifications of the AR(1) panel data model (1) considered.[15] Then, generate the bootstrap samples as

$$\begin{aligned} y_{-1}^B &= e_T \otimes y_0 + \Lambda u^B, \\ y^B &= y_{-1}^B + u^B. \end{aligned}$$

   where y_0 contains the actual initial observations.

3. For each bootstrap sample calculate the following statistic: $\left(Z^{B(\lambda)} - Z^{(\lambda)}\right)$ and, based on the repetitions of Step 2, compute the empirical probability of the event $\left(Z^{B(\lambda)} - Z^{(\lambda)}\right) \leq \pi$.

When the date of λ is unknown, our previous sections' results on the distribution of the minimum of $Z^{(\lambda)}$ still hold. The elements of the variance-covariance matrix between two statistics $Z^{(\mu)}$ and $Z^{(s)}$, $\sigma_{\mu s}$, will be given as

$$\sigma_{\mu s} = \frac{tr(\tilde{F}^{(\mu)}\Gamma^{\varepsilon}\tilde{F}^{(s)}\Gamma^{\varepsilon})}{\sqrt{tr(\tilde{F}^{(\mu)}\Gamma^{\varepsilon}\tilde{F}^{(\mu)}\Gamma^{\varepsilon})}\sqrt{tr(\tilde{F}^{(s)}\Gamma^{\varepsilon}\tilde{F}^{(s)}\Gamma^{\varepsilon})}}. \tag{35}$$

The latter however can not be estimated as in previous sections for the reasons mentioned above. The bootstrap method proposed previously is valid for the following test statistic: $\min_{\lambda\in I}\left(Z^{B(\lambda)} - Z^{(\lambda)}\right)$. This

[14] Resampling can be done across both the unit and the time dimension as long as it respects the date of the break for the model with a break under the null. This means that there should be a separate resampling for observations before the break and separate for observations after the break.

[15] Notice that $u^B = u$ for the version of the model with intercepts, $u^B = e_T \otimes \beta + u$ for that with linear trends and no break under H_0: $\varphi = 1$ and, finally, $u^B = e_T^{(1)} \otimes \beta^{(1)} + e_T^{(2)} \otimes \beta^{(2)} + u$ for the version of the model with a break under H_0: $\varphi = 1$. Thus, the individual effect information is not lost. The samples are generated under H_0: $\varphi = 1$ in order to maintain their unit root process behaviour (Basawa et al. (1991)).

comes from the continuity of the minimum function and the fact that the asymptotics are taken with respect to N and not T. The bootstrap procedure can be outlined as follows:

1. Use the data to compute $Z^{(\lambda)}$, for all $\lambda \in I$.

2. Same as Step 2 for the case of known λ. This step remains the same as before because the bootstrap samples are based on errors which are estimated by taking the first difference of y_{it}, Δy_{it}, and not by using a break dependent estimator. Therefore, the errors u^B contain the break information.

3. For each bootstrap sample, calculate test statistic $\left(Z^{B(\lambda)} - Z^{(\lambda)}\right)$, for all $\lambda \in I$. Then, select the min of it, i.e., $\min_{\lambda \in I}\left(Z^{B(\lambda)} - Z^{(\lambda)}\right)$. Based on the repetitions of Step 2, then compute the empirical probability of the event $\min_{\lambda \in I}\left(Z^{B(\lambda)} - Z^{(\lambda)}\right)$.

7 Simulation Results

In this section we present the results of a Monte Carlo study investigating the small sample performance of the proposed test statistics. For reasons of space, we present results only for the case of unknown break point λ, as this is more relevant in practice. Sample sizes for N and T are chosen to be $N = \{50, 100, 200\}$ and $T = \{8, 10, 15\}$, respectively. We consider the following fractions of sample that the break occurs: $\lambda/T = \{0.25, 0.5, 0.75\}$. All experiments are conducted based on 1000 iterations.

We present size and power performance results for models (1) and (19) allowing for serial correlation of error terms u_{it}, spatial dependence and two break points, respectively. The nominal size is set at 5% and the power of the tests is calculated based on this level of size. The extension of the models with serial correlation assumes that error terms u_{it} follow MA(1) process: $u_{it} = \varepsilon_{it} + \theta\varepsilon_{it-1}$ with $\varepsilon_{it} \sim NIID(0, 1)$, for all i and t, and $\theta = \{-0.5, 0.0, 0.5\}$. Spatial dependence is modelled through the global and local dependence models SAR and SMA, respectively, for $\mu = \{0.4, 0.8\}$. The spatial dependence weighting matrix W has zeroes in its main diagonal and is labelled as "2 ahead and 2 behind", with the non-zero elements being equal to 1/4 (see also Baltagi et al. (2007)). The number of bootstrap samples is set to 199 and, for simplicity, a fixed block length of 5 is chosen where the blocks are allowed to overlap.[16] The values of the nuisance parameters of models (1) and (19) considered in our study, namely the individual effects and/or the slope coefficients of incindental trends are assumed that are driven from the following distributions: $\alpha_i^{(1)} \sim U(-0.5, 0)$, $\alpha_i^{(2)} \sim U(0, 0.5)$, $\alpha_i^{(3)} \sim U(0, 1.5)$, $\beta_i \sim U(0, 0.05)$, $\beta_i^{(1)} \sim U(0, 0.025)$, $\beta_i^{(2)} \sim U(0.025, 0.05)$, $\beta_i^{(3)} \sim U(0.05, 0.75)$, where $U(\cdot)$ stands for the uniform distribution, and $y_{i0} = 0$, for all i. These magnitudes of $\alpha_i^{(j)}$ and $\beta_i^{(j)}$, for $j = 1, 2$, correspond to evidence provided in the empirical literature, see e.g., Hall and Mairesse (2005).

Tables 1 and 2 present the results of our Monte Carlo study for sequential test statistic $\min_{\lambda \in I} Z^{(\lambda)}$ corresponding to models (1) and (19) allowing for serially correlated error terms. These indicate that $\min_{\lambda \in I} Z^{(\lambda)}$ has size which is close to its nominal level 5%, for both models considered. This is true for all combinations of N and T considered. It is also true even for the case that the MA parameter θ takes a large

[16] Note that a more sophisticated application of the block bootstrap can be also considered, but it is not pursued, here, as the main focus is to study the overall performance of the tests.

negative value, i.e. $\theta = -0.5$. Note that, for this case, single time series unit root tests are critically oversized (see, e.g., Schwert (1989)). The size of statistic $\min_{\lambda \in I} Z^{(\lambda)}$ improves as N increases relative to T. This can be obviously attributed to the fact that the variance-covariance matrix Γ is more precisely estimated by estimator $\hat{\Gamma}$, as N increases. The above results hold independently on the break point of the sample λ.

Regarding the power of our tests, the results of the table indicate that $\min_{\lambda \in I} Z^{(\lambda)}$ has its highest power for the dynamic panel data model which consists only of individual effects, i.e., model (1). As was expected from the literature (see, e.g., Karavias and Tzavalis (2014b)), the power of test statistic $\min_{\lambda \in I} Z^{(\lambda)}$ for model (19), considering also incidental trends, is much less than that of model (1). However, for both these models, the power of the test increases faster with T rather than N. The value of MA parameter θ has significant impact on the power of statistic $\min_{\lambda \in I} Z^{(\lambda)}$, especially for model (1). For this model, the power of the test increases if $\theta \geq 0$ and the break is in the middle or towards the end of the sample. Consistently with the theory, the power of the tests increases also as the value of φ moves away from unity.

The size and power results of our simulation exercise for the case that error terms u_{it} are spatially dependent are presented in Tables 3-6. Tables 3 and 4 present results for model (1) for the cases that u_{it} follow models SMA and SAR, respectively. For model (19), the corresponding results are presented by Tables 5 and 6, respectively. Overall, the results of this exercise indicate that $\min_{\lambda \in I} Z^{(\lambda)}$ allowing for spatial dependence has very good size and power performance. For both the above panel data models examined, the performance of $\min_{\lambda \in I} Z^{(\lambda)}$ is better for the SMA model of spatial dependence rather than the SAR. For the last model of spatial dependence, $\min_{\lambda \in I} Z^{(\lambda)}$ has both very good size and power performance for the smaller value of μ examined, i.e., $\mu = 0.4$. For $\mu = 0.8$, it is oversized. As also argued by Baltagi et al. (2007), in this case the SAR model assumes a very high degree of dependence. However, our results indicate that even in this case adjusting unit root test statistics for spatial dependence improves both their size and power performance. As for the case of serially correlated errors, the performance of statistic $\min_{\lambda \in I} Z^{(\lambda)}$ increases with N and T, but faster with T. The location of the break is not found to affect the size and power performance of the test significantly.

Finally, Table 7 reports the results of our simulation exercise for the case that there two-breaks in models (1) and (19). In this exercise, we consider different combinations of break locations and, for exposition reasons, we assume no serial and cross section dependence. The results of the table indicate that the performance our test statistic in this case, denoted as $\min_{\lambda \in I} Z^{(\lambda_1,\lambda_2)}$, is similar to that with one break point, reported in the previous tables. This happens even for very short time-dimension of the panel, e.g., $T = 8$. Also, the different locations of the breaks do not seem to affect the size and power of our test statistic. Allowing for serial or spatial correlation does not change these conclusions either.

8 Empirical Application

Below, we illustrate the use of our proposed tests in answering the question if the net real income per fiscal household for French administrative communes ($N = 1000$ in number) contains a unit root in its autoregressive component. Using the same data set, Baltagi et al. (2007) study the impact of spatial dependence on panel unit root tests and find that it leads to size distortions if not accounted for. Studying the net real income per fiscal household per commune, these authors show that the individual panel data series involved are cross-sectionally correlated. Based on various large-T panel unit root tests, they get conflicting evidence on whether the individual series of this panel data set constitute unit root processes. It is notable that in the most relevant case where the panel data model considered allows for linear trends (see Figure 1), none of the applied tests can reject the null hypothesis of a unit root, whether they account for cross section dependence, or not. This lacks an economic intuition and can be attributed to the short time dimension of the panel data involved, i.e., $T = 14$ yearly observations covering the period 1985-1998, and/or the existence of a break point in the data generating process.

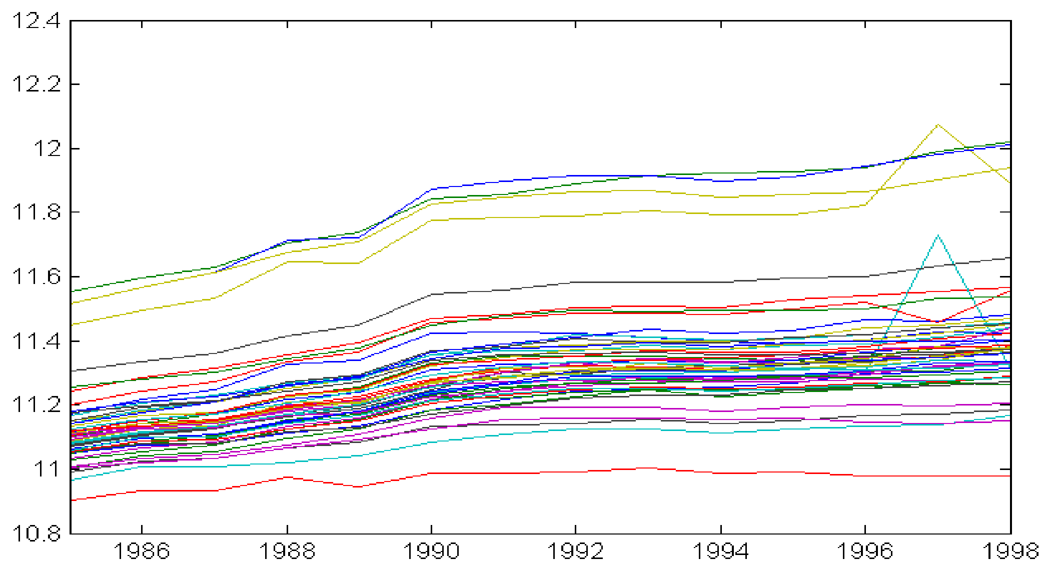

Figure 1. Net real income per fiscal household for French administrative communes

To address the above question, we implement our test statistic $\min_{\lambda \in I} \left(Z^{B(\lambda)} - Z^{(\lambda)} \right)$, which allows for spatial dependence across the units of our panel data set. This is done for the auxiliary AR(1) panel data regression model which considers incidental trends in the data generating process and a break under the alternative hypothesis. In the implementation of our test (denoted as WGSP), we have chosen the block length to be 5; this is according to Hall et al. (1995) and to the fact that the block length has to divide N. The bootstrap samples are chosen to be 1999. For comparison, we also consider the fixed-T panel unit root tests of Harris and Tzavalis (1999) (denoted HT) and Breitung's (2000) (denoted BRT), as was extended by Karavias and Tzavalis (2014) for finite T samples, as well as our test statistic WGSP, but without allowing for a common break (denoted $WGSP^{NO\ BREAK}$). The latter can be implemented by the previous theorems, after appropriately designing the annihilator and selection matrices without broken deterministic components. For the case without spatial correlation the limiting distribution is a standard normal while for the case with spatial dependence the bootstrap must be applied. The results of these tests are presented below:

$$HT = 12.06,\ BRT = 5.042,\ WGSP^{NO\ BREAK} = -0.006,\ WGSP = -0.0403^{***}\ \text{(date of break: 1988)}$$

where "***" indicates significance at 1%. These results clearly indicate that only the test which allows for a break point in the panel data model (see $WGSP$) can clearly reject the null hypothesis of unit roots. All the other tests considered can not reject the this hypothesis, including the version of our test that allows for cross-dependence, i.e., $WGSP^{NO\ BREAK}$. Consistently with the pictorial results of Figure 1, our test indicates that the break point occurs in year 1988. These results support the view that real income per fiscal household constitutes stationary series, for all French communes.

9 Concluding Remarks

In this paper new panel unit root tests are proposed for finite (fixed) T panel data models. They allow for multiple structural breaks, linear and/or nonlinear trends, spatial and temporal (serial correlation) dependence in the error terms of the dynamic panel data model. The finite T assumption of the tests make them appropriate for short panels, with small time dimensions often employed in microeconomic studies. The tests do not rely on any distributional assumptions about the initial conditions of the panel, which may be proved restrictive in practice, and they can be implemented to the case of unknown date breaks. In the last case, the paper derives the limiting distribution of the tests, analytically, based on recent results on the distribution of the minimum order statistic. This distribution is a mixture of normals and considerably facilitates calculation of the critical values of the tests.

The heteroscedasticity, heterogeneity and short term dependence considered by the tests can be of unknown form. This is due to the fact that asymptotics are taken across the cross section (N) dimension of the panel. The order of serial correlation is bounded by T. Also, spatial dependence can be considered without having to specify the weighting matrix of the economic or geographic distance among the cross section units of the panel. To carry out the tests in the case of spatial dependence, we recommend application of the block bootstrap method.

To examine the small sample performance of the tests, the paper conducts a Monte Carlo study. The results of this study clearly demonstrate that the suggested tests have size very close to their nominal level and very satisfactory power. This happens even under spatial and/or serial correlation of the error terms. These properties of the tests are valid even for very short panels of $T = \{8, 10\}$ observations and they also hold for the case of multiple breaks. When testing the null hypothesis of a unit root in the net real income per fiscal household for the 1000 largest French administrative communes, we find that only when we consider both spatial dependence and a structural break there is evidence of stationarity, as is expected by the economic theory.

References

[1] Abowd, J.M., Card, D., 1989. On the Covariance Structure of Earnings and Hours Changes. Econometrica. Econometric Society, vol. 57(2), pages 411-45.

[2] Anselin, L., 1990. Some robust approaches to testing and estimation in spatial econometrics. Regional Science and Urban Economics, 20(2), 141-163.

[3] Anselin, L., 2003. Spatial Externalities, Spatial Multipliers, And Spatial Econometrics. International Regional Science Review, 26: 153.

[4] Anselin, L., 2007. Spatial econometrics in RSUE: Retrospect and prospect. Regional Science and Urban Economics, 37(4), 450-456.

[5] Apergis, N., Loomis, D., Payne, J.E., 2010. Are fluctuations in coal consumption transitory or permanent? Evidence from a panel of US states. Applied Energy, 87, 2424–2426.

[6] Arellano, M., 1990. Testing for Autocorrelation in Dynamic Random Effects Models. Review of Economic Studies, Wiley Blackwell, vol. 57(1), pages 127-34.

[7] Arellano, M., 2003. Panel data econometrics. Oxford University Press.

[8] 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.

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

[10] 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.

[11] Bai, J., Ng, S., 2010. Panel Unit Root Tests With Cross-Section Dependence: A Further Investigation. Econometric Theory, 26(04), 1088-1114.

[12] Bai, J., Perron, P., 1998. Estimating and Testing Linear Models with Multiple Structural Changes. Econometrica, 66(1), 47-78.

[13] Baltagi, B.H., Bresson, G., Pirotte, A., 2007. Panel unit root tests and spatial dependence. Journal of Applied Econometrics, 22(2), 339-360.

[14] Basawa, I. V., Mallik, A. K., McCormick, W. P., Reeves, J. H., Taylor, R. L., 1991. Bootstrapping unstable first-order autoregressive processes. Annals of Statistics. 19, 1098-1101.

[15] Basker, E., 2005. Job Creation or Destruction? Labor Market Effects of Wal-Mart Expansion. The Review of Economics and Statistics, 87(1), 174-183.

[16] Beran, R., Ducharme, G.R., 1991. Asymptotic Theory for Bootstrap Methods in Statistics, Les Publications CRM, Centre de recherches mathematiques, Universite de Montreal, Montreal, Canada.

[17] Bond, S., Nauges, C., and Windmeijer, F., 2005. Unit roots: Identification and testing in micropanels. Cemmap Working Paper CWP07/05, The Institute for Fiscal Studies, UCL.

[18] Bond, S., Elston, J.A., Mairesse, J., Mulkay, B., 2003. Financial Factors and Investment in Belgium, France, Germany, and the United Kingdom: A comparison Using Company Panel Data. The Review of Economics and Statistics, 85(1), 153-165.

[19] Breitung J., 2000. The local power of some unit root tests for panel data. In Badi H. Baltagi, Thomas B. Fomby, R. Carter Hill (ed.) Nonstationary Panels, Panel Cointegration, and Dynamic Panels (Advances in Econometrics, Volume 15), Emerald Group Publishing Limited, pp.161-177.

[20] Breitung, J., Eickmeier, S., 2011. Testing for structural breaks in dynamic factor models. Journal of Econometrics, Elsevier, vol. 163(1), pages 71-84, July.

[21] Canarella, G., Miller, S.M., Nourayi, M.M., 2013. Firm profitability: Mean-reverting or random-walk behavior? Journal of Economics and Business 66, 76– 97.

[22] 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.

[23] Chen, P., Lee, C.C., 2007. Is energy consumption per capita broken stationary? New evidence from regional-based panels. Energy Policy 35, 3526–3540.

[24] Chang, Y., Song, W., 2009. Testing for Unit Roots in Small Panels with Short-run and Long-run Cross-sectional Dependencies. Review of Economic Studies 76, 903–935.

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

[26] Conley, T.G., 1999. GMM estimation with cross sectional dependence. Journal of Econometrics, 92, 1-45.

[27] Conley, T.G., Ligon, E., 2002. Economic Distance and Cross-Country Spillovers, Journal of Economic Growth, 7(2), 157-87.

[28] Constantini, M., Lupi, C., 2007. An analysis of inflation and interest rates. New panel unit root results in the presence of structural breaks. Economics Letters 95, 408–414.

[29] De Blander, R., Dhaene, G., 2012. Unit root tests for panel data with AR(1) errors and small T. Econometrics Journal, 15(1), 101-124.

[30] 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.

[31] De Wachter S., and Tzavalis E. (2012). Detection of structural breaks in linear panel data models, Computational Statistics & Data Analysis, 56, 3020-3034.

[32] Hadri, K., Larsson, R., Rao, Y., 2005. Testing for stationarity in heterogeneous panel data where the time dimension is finite. Econometrics Journal, 8(1), 55-69.

[33] Hadri, K., Larsson, R., Rao, Y., 2012. Testing For Stationarity With A Break In Panels Where The Time Dimension Is Finite. Bulletin of Economic Research, 64, 123-148.

[34] Hafner, K., Mayer-Foulkes, D., 2013. Fertility, economic growth, and human development causal determinants of the developed lifestyle. Journal of Macroeconomics, forthcoming.

[35] Hahn, J., 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.

[36] 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.

[37] Hall, P., 1985. Resampling a coverage pattern. Stochastic Processes and their Applications 20, 231-246.

[38] Hall, P., Horowitz, J.L., Bing-Yi, J., 1995. On blocking rules for the bootstrap with dependent data. Biometrika, 82,3:561–574.

[39] Han, C and Phillips P.C.B (2012). GMM estimation for dynamic panels with fixed effects and strong instruments at unity. Econometric Theory, 26, 119–151.

[40] Han, C and Phillips P.C.B (2013). First Difference Maximum Likelihood and Dynamic Panel Estimation, Journal of Econometrics, 175, 35-45.

[41] Harris, D., Leybourne, S., McCabe, B., 2005. Panel stationarity tests for Purchasing Power Parity with cross-sectional dependence. Journal of Business and Economic Statistics, 23(4) (395-409).

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

[43] Harris, R.D.F., 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.

[44] Harvey, D.I., Leybourne, S.J., 2013. Break date estimation for models with deterministic structural change. Granger Centre Discussion Paper Series, 13/02.

[45] Harvey, D.I., Leybourne, S.J., Taylor, A.M.R., 2011. Testing For Unit Roots And The Impact Of Quadratic Trends, With An Application To Relative Primary Commodity Prices Econometric Reviews. 30(5), 514-547.

[46] Horowitz, J., 2001. The Bootstrap, in Handbook of Econometrics, Vol. 5, ed. by J. J. Heckman and E. E. Leamer. Amsterdam: Elsevier, 3159-3228.

[47] Kapoor, M., Kelejian, H.H., Prucha, I., 2007. Panel Data Models with Spatially Correlated Error Components. Journal of Econometrics 140, 2007, 97-130.

[48] Karavias Y., and Tzavalis E., 2014a. A fixed-T version of Breitung's panel data unit root test. Economics Letters, 124(1), 83-87.

[49] Karavias Y., and Tzavalis E., 2014b. Testing for unit roots in short panels allowing for structural breaks. Computational Statistics and Data Analysis, 76, 391-407.

[50] Kelejian, H., Prucha I.R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Journal of Econometrics 157, 53-67.

[51] Kim, D., Perron, P., 2009. Unit root tests allowing for a break in the trend function at an unknown time under both the null and alternative hypotheses. Journal of Econometrics, 148, 1-13.

[52] Kiviet, J.F., 1995. On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. Journal of Econometrics, 68, 53-78.

[53] Kruiniger, H., 2008. Maximum likelihood estimation and inference methods for the covariance stationary panel AR(1)/unit root model. Journal of Econometrics 144, 447-464.

[54] Kruiniger, H., and E., Tzavalis, 2002. Testing for unit roots in short dynamic panels with serially correlated and heteroscedastic disturbance terms. Working Papers 459, Department of Economics, Queen Mary, University of London, London.

[55] Lahiri, S.N., 1999. Theoretical comparisons of block bootstrap methods. Annals of Statistics, 7(1), 386-404.

[56] Lean, H.H., Smyth, R., 2013. Will policies to promote renewable electricity generation be effective? Evidence from panel stationarity and unit root tests for 115 countries. Renewable and Sustainable Energy Reviews, 22, 371–379.

[57] Levin, A., C. Lin, C., J., Chu, 2002. Unit Root Tests in Panel Data: Asymptotic and Finite-sample Properties. Journal of Econometrics 108, 1–24.

[58] Lin, P., Huang, H., 2012. Inequality convergence revisited: Evidence from stationarity panel tests with breaks and cross correlation. Economic Modelling 29, 316–325.

[59] Lee, J., Strazicich, M.C., 2003. Minimum LM Unit Root Test with Two Structural Breaks. Review of Economics and Statistics, 85(4), 1082-1089.

[60] Lumsdaine, R.L., Papell, D.H., 1997. Multiple trend breaks and the unit root hypothesis. The Review of Economics and Statistics, 79(2), 212-218.

[61] Madsen, E., 2008. GMM estimators and unit root tests in the AR(1) panel data model. Centre for Applied Microeconometrics. University of Copenhagen

[62] Meligotsidou, L., Tzavalis E. and Y. Vrontos (2014). A Bayesian Method Detecting Unit Root from Stationary Processes Based on Panel Data Models with Cross-Sectional Dependence, Statistics and Computing, forthcoming.

[63] Mello, M., 2011. Stochastic convergence across US states. Macroeconomic Dynamics, 15, 160-183.

[64] Moon, H.R., Perron, B., 2004. Testing for a unit root in panels with dynamic factors, Journal of Econometrics, 122(1), 81-126.

[65] Murray, J., Papell, D.H., 2005. Do Panels Help Solve the Purchasing Power Parity Puzzle? Journal of Business and Economic Statistics, Vol. 23(4), pp. 410-415.

[66] Narayan, P.K., Narayan, S., Smyth, R., 2008. Are oil shocks permanent or temporary? Panel data evidence from crude oil and NGL production in 60 countries. Energy Economics, 30, 919–936.

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

[68] Palm, F.C., Smeekes, S., Urbain, J.P., 2011. Cross-sectional dependence robust block bootstrap panel unit root tests. Journal of Econometrics, 163, 85–104.

[69] Perron, P., 1989. The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis. Econometrica, 57(6), 1361-1401.

[70] Perron, P., Vogelsang, T., 1992. Nonstationarity and Level Shifts with an Application to Purchasing Power Parity. Journal of Business & Economic Statistics, 10, 301-320.

[71] Pesaran, M.H., Smith, L., Yamagata, T., 2013. Panel unit root tests in the presence of a multifactor error structure. Journal of Econometrics, 175(2), 94-115.

[72] Phillips, Peter C. B., and Hansen, Bruce E., 1990. Statistical Inference in Instrumental Variables Regression with I(1) Processes. Review of Economic Studies, Wiley Blackwell, vol. 57(1), pages 99-125.

[73] Phillips, P.C.B., Sul, D. 2007. Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence, Journal of Econometrics, 137, 162-188.

[74] Politis, D.N. and Romano, J.P. 1994. The Stationary Bootstrap. Journal of the American Statistical Association. 89:1303-1313.

[75] Robertson, D., Sarafidis, V., Westerlund, J., 2014. GMM Unit Root Inference in Generally Trending and Cross-Correlated Dynamic Panels. MPRA Paper 53419, University Library of Munich, Germany.

[76] Romero-Avila, D., Usabiaga, C., 2007. Unit Root Tests, Persistence, and the Unemployment Rate of the U.S. States. Southern Economic Journal, 73(3), 698-716.

[77] Sarafidis, V., 2009. GMM Estimation of Short Dynamic Panel Data Models With Error Cross-Sectional Dependence. MPRA Paper 25176, University Library of Munich, Germany.

[78] Sarafidis, V., Wansbeek, T., 2012. Cross-Sectional Dependence in Panel Data Analysis. Econometric Reviews, 31(5), 483-531.

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

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

[81] Sul, D., 2009. Panel unit root tests under cross section dependence with recursive mean adjustment. Economics Letters,105, 123–126.

[82] Suvankulov, F., Chi Keung Lau, M., Ogucu, F., 2012. Price regulation and relative price convergence: Evidence from the retail gasoline market in Canada. Energy Policy, 40, 325-334.

[83] Westerlund, J., 2014. Heteroscedasticity robust panel unit root tests. Journal of Business and Economic Statistics, 32:1, 112-135.

[84] Yan, B., Bernard, A., Warre, P., 2007. Industry characteristics and the law of one price. North American Journal of Economics and Finance, 18, 93-105.

[85] 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.

10 Appendix

In this appendix, we provide proofs of the theorems presented in the main text of the paper. To prove these theorems, we rely on a number of lemmas and remarks. The appendix is organised as follows. In section A, we provide some preliminary results (remarks and lemmas) which apply throughout the paper. Sections B, C, D, E and F provide lemmas and proofs of the theorems of the paper corresponding to Sections 2, 3, 4, 5 and 6, respectively.

Section A (Preliminary matrix algebra results):

Remark A1 (Properties of matrices $Q^{(\lambda)}$ and Λ_T).

i) Annihilator matrix $Q^{(\lambda)}$, defined in Section 2 (see equation (3)), is a $NT \times NT$ matrix which can be written as $Q^{(\lambda)} = Q_T^{(\lambda)} \otimes I_N$ where $Q_T^{(\lambda)}$ is a $T \times T$ matrix defined as $Q_T^{(\lambda)} = I_T - X_T^{(\lambda)}(X_T^{(\lambda)\prime} X_T^{(\lambda)})^{-1} X_T^{(\lambda)}$. I_T is an identity matrix of dimension $T \times T$ and $X_T^{(\lambda)} = \left[e_T^{(1)}, e_T^{(2)}\right]$ for model (1). It also holds $X^{(\lambda)} = X_T^{(\lambda)} \otimes I_N$.

Matrix $Q_T^{(\lambda)}$ is idempotent and othrogonal to the vectors of matrix $X_T^{(\lambda)}$, i.e., $Q_T^{(\lambda)}e_T^{(1)} = Q_T^{(\lambda)}e_T^{(2)} = 0$. For models (19) and (23), $X_T^{(\lambda)}$ is given as $X_T^{(\lambda)} = \left[e_T^{(1)}, e_T^{(2)}, \tau_T^{(1)}, \tau_T^{(2)}\right]$ and also satisfies the following orthogonality conditions: $Q_T^{(\lambda)}e_T^{(1)} = Q_T^{(\lambda)}e_T^{(2)} = Q_T^{(\lambda)}\tau_T^{(1)} = Q_T^{(\lambda)}\tau_T^{(2)} = 0$. These properties of $Q_T^{(\lambda)}$ also hold for $X_T^{(\lambda)} = [e_T^{(1)}, e_T^{(2)}, \tau_T^{(1)}, \tau_T^{(2)}, \tau_{2T}^{(1)}, \tau_{2T}^{(2)}]$ (see model 25), as well as for $X^{(\lambda_1,\dots,\lambda_{S_b})}$ employed in the panel model allowing for multiple breaks. Using the Kronecker product properties, it can shown that

$$\begin{aligned} Q^{(\lambda)}(e_T^{(1)} \otimes a^{(1)}) &= (Q_T^{(\lambda)} \otimes I_N)(e_T^{(1)} \otimes a^{(1)}) = Q_T^{(\lambda)}e_T^{(1)} \otimes I_N a^{(1)} = 0 \quad \text{and} \\ Q^{(\lambda)}X^{(\lambda)}(a^{(1)\prime}, a^{(2)\prime})' &= 0. \end{aligned}$$

(ii) Matrix Λ_T, which naturally arises in presentations of stacked vectors of AR(1) models, is defined as $(\Lambda_T)_{r,c} = 1$, if $r > c$ and 0 otherwise. Define $\Lambda = \Lambda_T \otimes I_N$. Then, the following properties hold for Λ:

$tr(\Lambda_T) = 0$,

$Q_T^{(\lambda)}\Lambda_T e_T = Q_T^{(\lambda)}\Lambda_T e_T^{(1)} = Q_T^{(\lambda)}\Lambda_T e_T^{(2)} = 0$, for $X_T^{(\lambda)} = \left[e_T^{(1)}, e_T^{(2)}, \tau_T^{(1)}, \tau_T^{(2)}\right]$, or equivalently,

$Q^{(\lambda)}\Lambda\left(e_T \otimes \beta\right) = Q^{(\lambda)}\Lambda\left(e_T^{(1)} \otimes \beta^{(1)}\right) = Q^{(\lambda)}\Lambda\left(e_T^{(2)} \otimes \beta^{(2)}\right) = 0$, for $X^{(\lambda)} = \left[e^{(1)}, e^{(2)}, \tau^{(1)}, \tau^{(2)}\right]$,

$Q_T^{(\lambda)}\Lambda_T\left(\sum_{j=1}^{2}\beta_i^{(j)}e_T^{(j)} + \sum_{j=1}^{2}\delta_i^{(j)}(2\tau_T^{(j)} - e_T^{(j)})\right) = 0$, for $X_T^{(\lambda)} = [e_T^{(1)}, e_T^{(2)}, \tau_T^{(1)}, \tau_T^{(2)}, \tau_{2T}^{(1)}, \tau_{2T}^{(2)}]$.

Remark A2. This remark justifies the use of variance-covariance matrix formulas employed in proofs of the paper. Based on Lemma A1 of Kelejian and Prucha (2010), it can be shown that, for a zero mean random vector $\zeta = (\zeta_1, ..., \zeta_{NT})'$ with positive definite variance-covariance matrix $\Gamma_\zeta = SS'$ and for a $NT \times NT$ non-stochastic matrix A_ζ for which the elements $(S'A_\zeta S)_{jj}$ are equal to 0, the following results hold:

$$E(\zeta' A_\zeta \zeta) = tr(S'A_\zeta S) = tr(A_\zeta \Gamma_\zeta) = 0 \quad \text{and} \quad Var(\zeta' A_\zeta \zeta) = 2tr(A_\zeta \Gamma_\zeta A_\zeta \Gamma_\zeta).$$

The important difference of these results from those on standard quadratic forms (see e.g. Schott (1996)) is that the form of a variance-covariance matrix does not contain higher than second order terms despite the fact that ζ may not be normally distributed.

If $(S'A_\zeta S)_{jj} \neq 0$, for some j, then define $\eta = S^{-1}\zeta$. Assuming that the elements of η are independently distributed with finite fourth moments $E(\eta_j^4)$, the following results holds for $S'A_\zeta S$:

$$Var(\zeta' A_\zeta \zeta) = 2tr(A_\zeta \Gamma_\zeta A_\zeta \Gamma_\zeta) + \sum_{j=1}^{NT}(S'A_\zeta S)_{jj}\left[E(\eta_j^4) - 3\right]. \tag{36}$$

The latter is also provided by Lemma A1 of Kelejian and Prucha (2010).

The following two lemmas provide the relationship between quadratic forms employing $NT \times 1$ and $T \times 1$ vectors, respectively. This relationship is frequently used in the proofs of lemmas presented below and, to our knowledge, has not been previously used in the literature.

Lemma A1. If $A_\zeta = A_\zeta^* \otimes I_N$ where A_ζ^* is a $T \times T$ matrix, $\zeta = (\zeta_1', ..., \zeta_T')'$ where $\zeta_t = (\zeta_{1t}, ..., \zeta_{Nt})'$, then we have

$$\zeta' A_\zeta \zeta = \sum_{i=1}^{N} \zeta_i^{*\prime} A_\zeta^* \zeta_i^* \tag{37}$$

where $\zeta_i^* = (\zeta_{i1}, ..., \zeta_{iT})'$ are $T \times 1$ random vectors.

Proof: Define the $N \times 1$ vector $e_N^{(i)}$ which has 1 at place i and zeros everywhere else. Then, we have $\zeta = \sum_{i=1}^{N} \zeta_i^* \otimes e_N^{(i)}$ and, hence,

$$\begin{aligned}
\zeta' A_\zeta \zeta &= \left(\sum_{i=1}^{N} \zeta_i^* \otimes e_N^{(i)} \right)' (A_\zeta^* \otimes I_N) \left(\sum_{i=1}^{N} \zeta_i^* \otimes e_N^{(i)} \right) \\
&= \left(\sum_{i=1}^{N} \zeta_i^{*\prime} A_\zeta^* \otimes e_N^{(i)\prime} \right) \left(\sum_{i=1}^{N} \zeta_i^* \otimes e_N^{(i)} \right) = \sum_{i=1}^{N} \zeta_i^{*\prime} A_\zeta^* \zeta_i^*,
\end{aligned}$$

since $e_N^{(i)\prime} e_N^{(j)} = 1$, for $i = j$, and 0, for $i \neq j$.

Lemma A2. If $A_\zeta = A_\zeta^* \otimes \Pi$ where A_ζ^* is a $T \times T$ matrix and Π is a $N \times N$ matrix, then we have

$$\zeta' A_\zeta \zeta = \sum_{i=1}^{N} \pi_{ii}^2 \zeta_i^{*\prime} A_\zeta^* \zeta_i^* + \sum_{i \neq j} \pi_{ij}^2 \zeta_i^{*\prime} A_\zeta^* \zeta_i^* \tag{38}$$

where $\zeta_i^* = (\zeta_{i1}, ..., \zeta_{iT})'$ are $T \times 1$ random vectors.

Proof: This can be proved based on the arguments of the proof of Lemma A1 and noticing that $e_N^{(i)\prime} \Pi' \Pi e_N^{(i)} = \pi_{ii}^2$ and $e_N^{(i)\prime} \Pi' \Pi e_N^{(j)} = \pi_{ij}^2$.

The following two remarks discuss some key properties between selection matrices of dimension $NT \times NT$ and $T \times T$ employed in our test statistics. These matrices apply to variance-covariance matrices $\Gamma = E(uu')$ and $\Gamma_T = \frac{1}{N} \sum_{i=1}^{N} E(u_i^* u_i^{*\prime})$, or their estimators given as $\hat{\Gamma}$ and $\hat{\Gamma}_T$, respectively, to select the elements of them which affect the bias of estimator $\hat{\varphi}^{(\lambda)}$.

Remark A3. The selection matrices employed in our test statistics are defined, first, as $T \times T$ dimension matrices to gain intuition. Their functioning is the same even though they are considered in their $NT \times NT$ dimension forms, applied to $\hat{\Gamma}$. For instance, matrix $\Psi_T^{(\lambda)}$ selects all elements of $\hat{\Gamma}_T$ in the main and the p-upper and p-lower diagonals and assigns them weights in accordance to $\Lambda_T' Q_T^{(\lambda)}$. The elements in these diagonals are non-zero, and thus can capture the inconsistency of LS estimator $\varphi^{(\lambda)}$. The $NT \times NT$ dimension form of matrix $\Psi_T^{(\lambda)}$ is given as $\Psi^{(\lambda)} = \Psi_T^{(\lambda)} \otimes I_N$. It selects all the elements of $\hat{\Gamma}$ which exist on the main Np-upper and Np-lower diagonals and assigns them weights in accordance to $\Lambda' Q^{(\lambda)}$. This result holds because matrices $\Gamma = E(uu')$ and $\Gamma_T \otimes I_N$ have their non-zero elements at the same places, even though they are not equal (see also Remark A4, below). Similar arguments hold for selection matrices M_T and M, or $F^{(\lambda)}$ and $F_T^{(\lambda)}$**,** etc.

Remark A4. For model (1), it can be shown that

$$p\lim_{N} \left[\hat{\Gamma} - (\Gamma_T \otimes I_N) \right] = 0_T,$$

where $\hat{\Gamma} = \hat{\Gamma}_T \otimes I_N$ and $\hat{\Gamma}_T = \frac{1}{N}\sum_{i=1}^{N} \Delta y_i^* \Delta y_i^{*\prime}$, since $p\lim_N(\hat{\Gamma}_T - \Gamma_T) = 0_T$. However, as noted before, $\Gamma = E(uu') \neq (\Gamma_T \otimes I_N)$ meaning that $p\lim_N \left[\hat{\Gamma} - \Gamma\right] \neq 0_{NT}$ and, thus, $\hat{\Gamma}$ is not a consistent estimator of Γ. This can be easily seen by deriving an analytic form of $\Gamma = E(uu')$ using $u = \sum_{i=1}^{N} u_i^* \otimes e_N^{(i)}$, where $E(uu') = E\left[(\sum_{i=1}^{N} u_i^* \otimes e_N^{(i)})(\sum_{i=1}^{N} u_i^{*\prime} \otimes e_N^{(i)\prime})\right] = E\left(\sum_{i=1}^{N} u_i^* u_i^{*\prime} \otimes e_{NN}^{(i)(i)\prime}\right) = \sum_{i=1}^{N} \Gamma_{iT} \otimes e_{NN}^{(i)(i)\prime}$. At the same time, note that $(\Gamma_T \otimes I_N) = \frac{1}{N}\sum_{i=1}^{N} \Gamma_{iT} \otimes I_N$, but $e_{NN}^{(i)(i)\prime} \neq I_N$.

Despite the fact that $\Gamma = E(uu') \neq (\Gamma_T \otimes I_N)$, matrices Γ and Γ_T have non-zero elements at the same places, as mentioned before. This implies that $tr(\Gamma_T) = \frac{1}{N}tr(\Gamma)$ and $tr(\hat{\Gamma}_T) = \frac{1}{N}tr(\Delta y \Delta y')$, and thus

$$tr(\Psi^{(\lambda)}\hat{\Gamma}) = tr(\Psi^{(\lambda)}\Delta y \Delta y').$$

To show this, write $tr(\Psi^{(\lambda)}\hat{\Gamma}) = Ntr(\Psi_T^{(\lambda)}\hat{\Gamma}_T) = N\frac{1}{N}\sum_{i=1}^{N} \Delta y_i^{*\prime}\Psi_T^{(\lambda)}\Delta y_i^*$. By Lemma A1, $\sum_{i=1}^{N} \Delta y_i^{*\prime}\Psi_T^{(\lambda)}\Delta y_i^*$ can be written as $\sum_{i=1}^{N} \Delta y_i^{*\prime}\Psi_T^{(\lambda)}\Delta y_i^* = \Delta y'\Psi^{(\lambda)}\Delta y = tr(\Psi^{(\lambda)}\Delta y \Delta y')$, which proves the above result.

Section B (Lemmas and Theorem Proofs for Section 2): The following lemmas prove various claims made in the text and they are needed for the proofs of Theorems 1 and 2.

Lemma A3. Under Assumption A, the within group LS estimator of φ, $\hat{\varphi}^{(\lambda)}$, for model (1) under H_0: $\varphi = 1$ is inconsistent, with its inconsistency given by $p\lim_N \left(\hat{\varphi}^{(\lambda)} - 1 - \frac{tr(\Lambda' Q^{(\lambda)})}{tr(\Lambda' Q^{(\lambda)}\Lambda)}\right) = 0$.

Proof: Model (1) implies

$$y = e_T^{(1)} \otimes a^{(1)} + e_T^{(2)} \otimes a^{(2)} + \zeta \quad \text{and} \quad y_{-1} = e_T^{(1)} \otimes a^{(1)} + e_T^{(2)} \otimes a^{(2)} + \zeta_{-1}.$$

By multiplying the second equation with φ and subtracting it from the first equation yields

$$y = \varphi y_{-1} + (1-\varphi)(e_T^{(1)} \otimes a^{(1)}) + (1-\varphi)(e_T^{(2)} \otimes a^{(2)}) + u. \tag{39}$$

Consider the last equation under H_0: $\varphi = 1$, i.e.,

$$y = y_{-1} + u, \tag{40}$$

Substituting this relationship backwards yields

$$y_{-1} = e_T \otimes y_0 + \Lambda u. \tag{41}$$

Using the above relationships, we can write

$$\begin{aligned}
\hat{\varphi}^{(\lambda)} - 1 &= \frac{y_{-1}' Q^{(\lambda)} y}{y_{-1}' Q^{(\lambda)} y_{-1}} - 1 = \frac{y_{-1}' Q^{(\lambda)}\left(y_{-1} + u\right)}{y_{-1}' Q^{(\lambda)} y_{-1}} - 1 = \frac{y_{-1}' Q^{(\lambda)} u}{y_{-1}' Q^{(\lambda)} y_{-1}} \\
&= \frac{(u'\Lambda' + e_T' \otimes y_0') Q^{(\lambda)} u}{(u'\Lambda' + e_T' \otimes y_0') Q^{(\lambda)}\left(e_T \otimes y_0 + \Lambda u\right)} = \frac{\frac{1}{N} u'\Lambda' Q^{(\lambda)} u}{\frac{1}{N} u'\Lambda' Q^{(\lambda)} \Lambda u}
\end{aligned}$$

By applying standard properties for quadratic forms, the numerator of the last relationship has $E(\frac{1}{N}u'\Lambda'Q^{(\lambda)}u) = \frac{1}{N}\sigma^2 tr(\Lambda'Q^{(\lambda)})$ and $Var(\frac{1}{N}u'\Lambda'Q^{(\lambda)}u) = 2\sigma^4 tr(A_n A_n)$, where $A_n = \frac{1}{2}(\Lambda'Q^{(\lambda)} + Q^{(\lambda)}\Lambda)$. The trace of matrix $A_n A_n$ is given as

$$\begin{aligned}
tr(A_n A_n) &= \tfrac{1}{4}tr((\Lambda'Q^{(\lambda)} + Q^{(\lambda)}\Lambda)(\Lambda'Q^{(\lambda)} + Q^{(\lambda)}\Lambda)) \\
&= \tfrac{1}{4}tr(\Lambda'Q^{(\lambda)}\Lambda'Q^{(\lambda)} + Q^{(\lambda)}\Lambda\Lambda'Q^{(\lambda)} + \Lambda'Q^{(\lambda)}Q^{(\lambda)}\Lambda + Q^{(\lambda)}\Lambda Q^{(\lambda)}\Lambda) \\
&= \tfrac{1}{4}tr(\Lambda_T'Q_T^{(\lambda)}\Lambda_T'Q_T^{(\lambda)} \otimes I_N + Q_T^{(\lambda)}\Lambda_T\Lambda_T'Q_T^{(\lambda)} \otimes I_N + \Lambda_T'Q_T^{(\lambda)}Q_T^{(\lambda)}\Lambda_T \otimes I_N + Q_T^{(\lambda)}\Lambda_T Q_T^{(\lambda)}\Lambda_T \otimes I_N) \\
&= \tfrac{1}{4}tr\left[\left(\Lambda_T'Q_T^{(\lambda)}\Lambda_T'Q_T^{(\lambda)} + Q_T^{(\lambda)}\Lambda_T\Lambda_T'Q_T^{(\lambda)} + \Lambda_T'Q_T^{(\lambda)}Q_T^{(\lambda)}\Lambda_T + Q_T^{(\lambda)}\Lambda_T Q_T^{(\lambda)}\Lambda_T\right) \otimes I_N\right] \\
&= \tfrac{1}{4}tr(\Lambda_T'Q_T^{(\lambda)}\Lambda_T'Q_T^{(\lambda)} + Q_T^{(\lambda)}\Lambda_T\Lambda_T'Q_T^{(\lambda)} + \Lambda_T'Q_T^{(\lambda)}Q_T^{(\lambda)}\Lambda_T + Q_T^{(\lambda)}\Lambda_T Q_T^{(\lambda)}\Lambda_T)tr(I_N).
\end{aligned}$$

Using the following results:

$tr(\Lambda_T'Q_T^{(\lambda)}\Lambda_T'Q_T^{(\lambda)} + Q_T^{(\lambda)}\Lambda_T\Lambda_T'Q_T^{(\lambda)} + \Lambda_T'Q_T^{(\lambda)}Q_T^{(\lambda)}\Lambda_T + Q_T^{(\lambda)}\Lambda_T Q_T^{(\lambda)}\Lambda_T) = O(T)$,

$tr(I_N) = O(N)$ and, hence,

$tr(A_n A_n) = O(TN)$,

it can be easily seen that $Var(\frac{1}{N}u'\Lambda'Q^{(\lambda)}u) = O(\frac{1}{N^2})O(TN) = O(\frac{T}{N}) = o(1)$. By Chebyshev's inequality $P(\left|\frac{1}{N}u'\Lambda'Q^{(\lambda)}u - E(\frac{1}{N}u'\Lambda'Q^{(\lambda)}u)\right| > \varepsilon) \le \frac{Var(\frac{1}{N}u'\Lambda'Q^{(\lambda)}u)}{\varepsilon^2} \to 0$, we have that $\frac{1}{N}u'\Lambda'Q^{(\lambda)}u \xrightarrow{p} \frac{1}{N}\sigma^2 tr(\Lambda'Q^{(\lambda)})$, as $N \to \infty$.

Following similar steps to the above, we can show that the numerator of $\hat{\varphi}^{(\lambda)} - 1$ scaled by N, given as $\frac{1}{N}u'\Lambda'Q^{(\lambda)}\Lambda u$, has $E(\frac{1}{N}u'\Lambda'Q^{(\lambda)}\Lambda u) = \frac{1}{N}\sigma^2 tr(\Lambda'Q^{(\lambda)}\Lambda)$ and $Var(\frac{1}{N}u'\Lambda'Q^{(\lambda)}\Lambda u) = \frac{1}{N^2}2\sigma^4 tr(A_{dn}A_{dn})$, where $A_{dn} = \frac{1}{2}(\Lambda'Q^{(\lambda)}\Lambda + \Lambda'Q^{(\lambda)}\Lambda)$ and $\frac{1}{N}u'\Lambda'Q^{(\lambda)}u \xrightarrow{p} \frac{1}{N}\sigma^2 tr(\Lambda'Q^{(\lambda)}\Lambda)$. Combining the above results on the numerator and denominator of $\hat{\varphi}^{(\lambda)} - 1$ implies that $p\lim_N(\hat{\varphi}^{(\lambda)} - 1 - \frac{\sigma^2 tr(\Lambda'Q^{(\lambda)})}{\sigma^2 tr(\Lambda'Q^{(\lambda)}\Lambda)}) = 0$, which proves the inconsistency of $\hat{\varphi}^{(\lambda)}$ under H_0: $\varphi = 1$.

Lemma A4. Under Assumption A and H_0: $\varphi = 1$, $\hat{\sigma}^2 = \frac{\Delta y'\Psi^{(\lambda)}\Delta y}{tr(\Psi^{(\lambda)})}$ is consistent estimator of σ^2.

Proof: To prove this, write $\hat{\sigma}^2$ as

$$\hat{\sigma}^2 = \frac{1}{tr(\Psi^{(\lambda)})}\Delta y'\Psi^{(\lambda)}\Delta y = \frac{1}{tr(\Psi^{(\lambda)})}tr(\Delta y'\Psi^{(\lambda)}\Delta y) = \frac{1}{tr(\Psi^{(\lambda)})}tr(\Psi^{(\lambda)}uu').$$

Based on Lemma A1, u can be written as $u = \sum_{i=1}^N u_i^* \otimes e_N^{(i)}$ and uu' as

$$uu' = \left(\textstyle\sum_{i=1}^N u_i^* \otimes e_N^{(i)}\right)\left(\sum_{i=1}^N u_i^{*\prime} \otimes e_N^{(i)\prime}\right) = \sum_{i=1}^N u_i^* u_i^{*\prime} \otimes e_N^{(i)}e_N^{(i)\prime} + \sum_{i\neq j} u_i^* u_j^{*\prime} \otimes e_N^{(i)}e_N^{(j)\prime}.$$

Using this relationship, it can be shown that

$$\begin{aligned}
tr(\Psi^{(\lambda)}uu') &= tr\left[\left(\Psi_T^{(\lambda)} \otimes I_N\right)\left(\textstyle\sum_{i=1}^N u_i^* u_i^{*\prime} \otimes e_N^{(i)}e_N^{(i)\prime} + \sum_{i\neq j} u_i^* u_j^{*\prime} \otimes e_N^{(i)}e_N^{(j)\prime}\right)\right] \\
&= tr\left(\textstyle\sum_{i=1}^N \Psi_T^{(\lambda)}u_i^* u_i^{*\prime} \otimes e_N^{(i)}e_N^{(i)\prime} + \sum_{i\neq j} \Psi_T^{(\lambda)}u_i^* u_j^{*\prime} \otimes e_N^{(i)}e_N^{(j)\prime}\right) \\
&= tr(\textstyle\sum_{i=1}^N \Psi_T^{(\lambda)}u_i^* u_i^{*\prime} \otimes e_N^{(i)}e_N^{(i)\prime}) + tr(\sum_{i\neq j} \Psi_T^{(\lambda)}u_i^* u_j^{*\prime} \otimes e_N^{(i)}e_N^{(j)\prime}).
\end{aligned}$$

Using properties of trace, the last relationship gives

$tr(\sum_{i\neq j} \Psi_T^{(\lambda)}u_i^* u_j^{*\prime} \otimes e_N^{(i)}e_N^{(j)\prime}) = \sum_{i\neq j} tr(\Psi_T^{(\lambda)}u_i^* u_j^{*\prime})tr(e_N^{(i)}e_N^{(j)\prime}) = 0$, since $tr(e_N^{(i)}e_N^{(j)\prime}) = 0$ for $i \neq j$,

and

$tr(\sum_{i=1}^{N} \Psi_T^{(\lambda)} u_i^* u_i^{*\prime} \otimes e_N^{(i)} e_N^{(i)\prime}) = \sum_{i=1}^{N} tr\left(\Psi_T^{(\lambda)} u_i^* u_i^{*\prime}\right) tr\left(e_N^{(i)} e_N^{(i)\prime}\right) = \sum_{i=1}^{N} tr\left(\Psi_T^{(\lambda)} u_i^* u_i^{*\prime}\right)$, since $tr\left(e_N^{(i)} e_N^{(i)\prime}\right) = 1$ for all i.

Based on the above results, it can be shown that

$$\hat{\sigma}^2 = \frac{1}{tr(\Psi^{(\lambda)})} \Delta y' \Psi^{(\lambda)} \Delta y = \frac{1}{tr(\Psi^{(\lambda)})} \sum_{i=1}^{N} tr\left(\Psi_T^{(\lambda)} u_i^* u_i^{*\prime}\right) = \frac{1}{tr(\Psi_T^{(\lambda)}) N} \sum_{i=1}^{N} u_i^{*\prime} \Psi_T^{(\lambda)} u_i^*.$$

For T finite, Assumption A implies $E(u_i^{*\prime} \Psi_T^{(\lambda)} u_i^*) = tr(\Psi_T^{(\lambda)} E(u_i^* u_i^{*\prime})) = \sigma^2 tr(\Psi_T^{(\lambda)}) < \infty$, for all i, and $Var(u_i^{*\prime} \Psi_T^{(\lambda)} u_i^*) = 2\sigma^4 tr(A_\Psi A_\Psi) < \infty$, where $A_\Psi = \frac{1}{2}(\Psi_T^{(\lambda)} + \Psi_T^{(\lambda)\prime})$. Then, by applying Khinchine's Weak Law of Large Numbers follows: $p\lim_N \frac{1}{tr(\Psi_T^{(\lambda)})N} \sum_{i=1}^{N} u_i^{*\prime} \Psi_T^{(\lambda)} u_i^* = \sigma^2$, which proves the consistency of $\hat{\sigma}^2$.

Proof of Theorem 1: For model (1), test statistic $Z^{(\lambda)}$ can be written under H_0: $\varphi = 1$ as follows:

$$\begin{aligned}
& \hat{d}^{(\lambda)} \sqrt{N} \left(\hat{\varphi}^{(\lambda)} - 1 - \frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}\right) \\
= & \left(\frac{1}{N} y_{-1}' Q^{(\lambda)} y_{-1}\right) \sqrt{N} \left(\frac{y_{-1}' Q^{(\lambda)} u}{y_{-1}' Q^{(\lambda)} y_{-1}} - \frac{\frac{1}{N} \Delta y' \Psi^{(\lambda)} \Delta y}{\frac{1}{N} y_{-1}' Q^{(\lambda)} y_{-1}}\right) \\
= & \sqrt{N} \left(\frac{1}{N} y_{-1}' Q^{(\lambda)} u - \frac{1}{N} \Delta y' \Psi^{(\lambda)} \Delta y\right) = \frac{1}{\sqrt{N}} u' (\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)}) u,
\end{aligned}$$

using (40) and (41). By Lemma A1, we have the following results:

$$u'(\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)})u = \sum_{i=1}^{N} u_i^* (\Lambda_T' Q_T^{(\lambda)} - \Psi_T^{(\lambda)}) u_i^* \quad \text{and}$$

$$\begin{aligned}
E\left[u_i^* (\Lambda_T' Q_T^{(\lambda)} - \Psi_T^{(\lambda)}) u_i^*\right] &= tr\left[(\Lambda_T' Q_T^{(\lambda)} - \Psi_T^{(\lambda)}) E\left(u_i^* u_i^{*\prime}\right)\right] = tr\left[(\Lambda_T' Q_T^{(\lambda)} - \Psi_T^{(\lambda)}) \sigma^2 I_T\right] \\
&= \sigma^2 tr\left(\Lambda_T' Q_T^{(\lambda)} - \Psi_T^{(\lambda)}\right) = 0.
\end{aligned}$$

Also by Remark A2, we have $Var\left[u_i^* (\Lambda_T' Q_T^{(\lambda)} - \Psi_T^{(\lambda)}) u_i^*\right] = 2\sigma^4 tr(F_T^{(\lambda)} F_T^{(\lambda)}) = \frac{1}{N} 2\sigma^4 tr(F^{(\lambda)} F^{(\lambda)})$. Since T is finite, Assumption A implies that $2\sigma^4 tr(F_T^{(\lambda)} F_T^{(\lambda)}) < \infty$, for all i. Using the above results, we can show that

$$\frac{1}{\sqrt{N}} u' (\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)}) u = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} u_i^* (\Lambda_T' Q_T^{(\lambda)} - \Psi_T^{(\lambda)}) u_i^* \xrightarrow{d} N\left(0, 2\sigma^4 tr(F_T^{(\lambda)} F_T^{(\lambda)})\right),$$

by the Lindeberg-Levy Central Limit Theorem. Note that the above limiting distribution can also be written in terms of $tr(F^{(\lambda)} F^{(\lambda)})$, i.e., using selection matrix $F^{(\lambda)}$. This can be done by noticing that $\frac{1}{N} tr(F^{(\lambda)} F^{(\lambda)}) = \frac{1}{N} tr(F_T^{(\lambda)} F_T^{(\lambda)} \otimes I_N) = tr(F_T^{(\lambda)} F_T^{(\lambda)})$.

Proof of Theorem 2: 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 variance-covariance matrix of random variables $Z^{(\mu)}$ and $Z^{(s)}$, for all $\mu \neq s$

(denoted as $\Sigma \equiv [\sigma_{\mu s}]$), can be derived by writing

$$\begin{aligned} Z^{(\mu)}Z^{(s)} &= V^{(\mu)-1/2}\hat{d}^{(\mu)}\sqrt{N}\left(\hat{\varphi}^{(\mu)}-1-\frac{\hat{b}^{(\mu)}}{\hat{d}^{(\mu)}}\right)V^{(s)-1/2}\hat{d}^{(s)}\sqrt{N}\left(\hat{\varphi}^{(s)}-1-\frac{\hat{b}^{(s)}}{\hat{d}^{(s)}}\right) \\ &= \frac{\hat{d}^{(\mu)}\hat{d}^{(s)}}{\sqrt{V^{(\mu)}V^{(s)}}}N\left(\hat{\varphi}^{(\mu)}-1-\frac{\hat{b}^{(\mu)}}{\hat{d}^{(\mu)}}\right)\left(\hat{\varphi}^{(s)}-1-\frac{\hat{b}^{(s)}}{\hat{d}^{(s)}}\right) \\ &= \frac{1}{\sqrt{V^{(\mu)}V^{(s)}}}\frac{1}{N}u'(\Lambda'Q^{(\mu)}-\Psi^{(\mu)})uu'(\Lambda'Q^{(s)}-\Psi^{(s)})u \end{aligned}$$

using results $\hat{d}^{(\mu)}\left(\hat{\varphi}^{(\mu)}-1-\frac{\hat{b}^{(\mu)}}{\hat{d}^{(\mu)}}\right)=\frac{1}{N}u'(\Lambda'Q^{(\mu)}-\Psi^{(\mu)})u$ and $\hat{d}^{(s)}\left(\hat{\varphi}^{(s)}-1-\frac{\hat{b}^{(s)}}{\hat{d}^{(s)}}\right)=\frac{1}{N}u'(\Lambda'Q^{(s)}-\Psi^{(s)})u$ from the proof of Theorem 1. Since $E\left(u'(\Lambda'Q^{(\mu)}-\Psi^{(\mu)})u\right)=0$ and $E\left(u'(\Lambda'Q^{(s)}-\Psi^{(s)})u\right)=0$ (see also proof of Theorem 1), the following result holds:

$$\begin{aligned} \frac{1}{N}E\left(u'(\Lambda'Q^{(\mu)}-\Psi^{(\mu)})uu'(\Lambda'Q^{(s)}-\Psi^{(s)})u\right) &= \frac{1}{N}Cov(u'(\Lambda'Q^{(\mu)}-\Psi^{(\mu)})u;u'(\Lambda'Q^{(s)}-\Psi^{(s)})u) \\ &= \frac{1}{N}2\sigma^4 tr(F^{(\mu)}F^{(s)}) \end{aligned}$$

Hence, the elements of $\Sigma \equiv [\sigma_{\mu s}]$ can be written analytically as

$$\sigma_{\mu s}=\frac{1}{\sqrt{V^{(\mu)}V^{(s)}}}E\left(\frac{1}{N}u'(\Lambda'Q^{(\mu)}-\Psi^{(\mu)})uu'(\Lambda'Q^{(s)}-\Psi^{(s)})u\right)=\frac{tr(F^{(\mu)}F^{(s)})}{\sqrt{tr(F^{(\mu)}F^{(\mu)})}\sqrt{tr(F^{(s)}F^{(s)})}}.$$

Section C (Lemmas and Theorem Proofs for Section 3): The following lemmas are needed for the proof of Theorem 3.

Lemma A5. Under Assumption B, the inconsistency of $\hat{\varphi}^{(\lambda)}$ for model (1) under H_0: $\varphi=1$ is given by

$$p\lim_N(\hat{\varphi}^{(\lambda)}-1-\tfrac{tr(\Lambda'Q^{(\lambda)}\Gamma)}{tr(\Lambda'Q^{(\lambda)}\Lambda\Gamma)})=0.$$

Proof: Based on Lemma A3, it can be shown that $\hat{\varphi}^{(\lambda)}-1=\frac{\frac{1}{N}u'\Lambda'Q^{(\lambda)}u}{\frac{1}{N}u'\Lambda'Q^{(\lambda)}\Lambda u}$, with $\frac{1}{N}u'\Lambda'Q^{(\lambda)}u=\frac{1}{N}\sum_{i=1}^{N}u_i^{*\prime}\Lambda_T'Q_T^{(\lambda)}u_i^*$ (see Lemma A1 and Remark A1). The numerator of the last relationship has

$$E(\tfrac{1}{N}u'\Lambda'Q^{(\lambda)}u)=\tfrac{1}{N}tr\left[\Lambda'Q^{(\lambda)}E(uu')\right]=\tfrac{1}{N}tr\left[\Lambda'Q^{(\lambda)}\Gamma\right] \text{ and}$$
$$Var(\tfrac{1}{N}u'\Lambda'Q^{(\lambda)}u)=\tfrac{1}{N^2}\sum_{i=1}^{N}Var(u_i^{*\prime}\Lambda_T'Q_T^{(\lambda)}u_i^*)=o(1),$$

by Condition (i) of Assumption B. Similarly, we can find the mean and variance of the denominator of $\hat{\varphi}^{(\lambda)}-1$, $\frac{1}{N}u'\Lambda'Q^{(\lambda)}\Lambda u$. The inconsistency of $\hat{\varphi}^{(\lambda)}$ can be proved by applying Chebyshev's inequality, as in the proof of Lemma A3. Note that, under Assumption B, error terms u_{it} are not normal and Remark A2 does not hold for $\Lambda_T'Q_T^{(\lambda)}$. Therefore, $Var(u_i^{*\prime}\Lambda_T'Q_T^{(\lambda)}u_i^*)$ is a function of the fourth moments of u_{it}.

Lemma A6. For model (1) with $\Gamma=E(uu')\neq\sigma^2 I_{NT}$, the following result holds:

$$\tfrac{1}{N}tr(\Psi^{(\lambda)}\Gamma)=tr(\Psi_T^{(\lambda)}\Gamma_T).$$

Proof: Write $\frac{1}{N}tr(\Psi^{(\lambda)}\Gamma) = \frac{1}{N}tr(\Psi^{(\lambda)}E(uu')) = E\left[\frac{1}{N}tr(u'\Psi^{(\lambda)}u)\right]$. Using Lemma A1, the last relationship can be written as follows:

$$\begin{aligned} E\left[\frac{1}{N}tr(\sum_{i=1}^{N} u_i^{*\prime}\Psi_T^{(\lambda)}u_i^*)\right] &= \frac{1}{N}tr\left[\sum_{i=1}^{N} E(u_i^{*\prime}\Psi_T^{(\lambda)}u_i^*)\right] = \frac{1}{N}\sum_{i=1}^{N} tr(\Psi_T^{(\lambda)}E(u_i^*u_i^{*\prime})) \\ &= \frac{1}{N}\sum_{i=1}^{N} tr(\Psi_T^{(\lambda)}\Gamma_{iT}) = tr(\Psi_T^{(\lambda)}\Gamma_T). \end{aligned}$$

Lemma A7. For model (1) with $\Gamma = E(uu')$, we have $p\lim_N(\hat{\Gamma}_T - \Gamma_T) = 0_T$, where 0_T is a $T \times T$ matrix of zeros.

Proof: Under H_0: $\varphi = 1$, model (1) implies $\Delta y_i^* = u_i^*$ and thus, $\hat{\Gamma}_T$ can be written as

$$\hat{\Gamma}_T = \frac{1}{N}\sum_{i=1}^{N}\Delta y_i^*\Delta y_i^{*\prime} = \frac{1}{N}\sum_{i=1}^{N} u_i^*u_i^{*\prime}.$$

This is matrix which has elements of the form $\frac{1}{N}\sum_{i=1}^{N} u_{i\mu}u_{i\nu}$, for $\mu, \nu = 1, ..., T$, with $E(u_{i\mu}u_{i\nu})=\gamma_{i\mu\nu T}$, where $\gamma_{i\mu\nu T}$ is the (μ, ν) element of matrix $\Gamma_{iT} = E(u_i^*u_i^{*\prime})$. Also by Assumption B, $Var(u_{i\mu}u_{i\nu})$ is finite. Then, by Chebyshev's Weak Law of Large Numbers it can be shown that $p\lim_N\left[\frac{1}{N}\sum_{i=1}^{N}(u_{i\mu}u_{i\nu} - \gamma_{i\mu\nu T})\right] = 0$, for all $\mu, \nu = 1, ..., T$, which means that $p\lim_N(\hat{\Gamma}_T - \Gamma_T) = 0_T$.

Proof of Theorem 3: As in Theorem 1, write $Z^{(\lambda)}$ under H_0: $\varphi = 1$ as

$$\begin{aligned} \hat{d}^{(\lambda)}\sqrt{N}\left(\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}} - 1\right) &= \left(\frac{1}{N}y_{-1}'Q^{(\lambda)}y_{-1}\right)\sqrt{N}\left(\frac{y_{-1}'Q^{(\lambda)}u}{y_{-1}'Q^{(\lambda)}y_{-1}} - \frac{\frac{1}{N}tr(\Psi^{(\lambda)}\hat{\Gamma})}{\frac{1}{N}y_{-1}'Q^{(\lambda)}y_{-1}}\right) \\ &= \frac{1}{\sqrt{N}}\left[u'\Lambda'Q^{(\lambda)}u - tr(\Psi^{(\lambda)}\hat{\Gamma})\right] \end{aligned}$$

By Remark A4, $tr(\Psi^{(\lambda)}\hat{\Gamma})$ can be written as $tr(\Psi^{(\lambda)}\hat{\Gamma}) = tr(\Psi^{(\lambda)}\Delta y\Delta y') = \Delta y'\Psi^{(\lambda)}\Delta y$. Since under H_0 : $\varphi = 1$ we have $\Delta y = u$, $\frac{1}{\sqrt{N}}\left[u'\Lambda'Q^{(\lambda)}u - tr(\Psi^{(\lambda)}\hat{\Gamma})\right]$ can be written as follows:

$$\frac{1}{\sqrt{N}}\left[u'\Lambda'Q^{(\lambda)}u - tr(\Psi^{(\lambda)}\hat{\Gamma})\right] = \frac{1}{\sqrt{N}}u'(\Lambda'Q^{(\lambda)} - \Psi^{(\lambda)})u = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} u_i^*(\Lambda_T'Q_T^{(\lambda)} - \Psi_T^{(\lambda)})u_i^*,$$

with $E\left[u_i^*(\Lambda_T'Q_T^{(\lambda)} - \Psi_T^{(\lambda)})u_i^*\right] = tr\left[\left(\Lambda_T'Q_T^{(\lambda)} - \Psi_T^{(\lambda)}\right)\Gamma_{iT}\right] = 0$ by construction. By Remark A2, we have $Var\left[u_i^*(\Lambda_T'Q_T^{(\lambda)} - \Psi_T^{(\lambda)})u_i^*\right] = 2tr\left(F_T^{(\lambda)}\Gamma_{iT}F_T^{(\lambda)}\Gamma_{iT}\right)$. Then, under Assumption B the Lindeberg-Feller CLT implies

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} u_i^*(\Lambda_T'Q_T^{(\lambda)} - \Psi_T^{(\lambda)})u_i^* \xrightarrow{d} N\left(0, 2tr\left(F_T^{(\lambda)}\Gamma_{Tu}F_T^{(\lambda)}\Gamma_{Tu}\right)\right),$$

which can prove the theorem. By virtue of Lemma A6, we have $2tr\left(F_T^{(\lambda)}\Gamma_{Tu}F_T^{(\lambda)}\Gamma_{Tu}\right) = \frac{1}{N}2tr\left(F^{(\lambda)}\Gamma F^{(\lambda)}\Gamma\right)$ and thus either $2tr\left(F_T^{(\lambda)}\hat{\Gamma}_T F_T^{(\lambda)}\hat{\Gamma}_T\right)$ or $\frac{1}{N}2tr\left(F^{(\lambda)}\hat{\Gamma}F^{(\lambda)}\hat{\Gamma}\right)$ can be employed in estimating the variance of the above limiting distribution. Both of these estimators will be numerically equivalent.

Section D (Lemmas and Theorem Proofs for Section 4): The following lemmas are required for the proofs of Theorems 4 and 5.

Lemma A8. Under Assumption B, the inconsistency of estimator $\hat{\varphi}^{(\lambda)}$ for model (19) under H_0: $\varphi = 1$ is given by $p\lim_N(\hat{\varphi}^{(\lambda)} - 1 - \frac{tr(\Lambda' Q^{(\lambda)}\Gamma)}{tr(\Lambda' Q^{(\lambda)}\Lambda\Gamma)}) = 0$.

Proof: Following the same algebraic transformation used in Lemma A.3, relationship (19) can be written as:

$$\begin{aligned} y &= y_{-1} + e_T \otimes \beta + u \quad \text{under } H_0\text{: } \varphi = 1 \qquad (42)\\ \text{and} \quad y &= \varphi y_{-1} + \varphi(e_T^{(1)} \otimes \beta^{(1)} + e_T^{(2)} \otimes \beta^{(2)}) + (1-\varphi)(e_T^{(1)} \otimes a^{(1)} + e_T^{(2)} \otimes a^{(2)}) \\ &\quad + (1-\varphi)(\tau_T^{(1)} \otimes \beta^{(1)} + \tau_T^{(2)} \otimes \beta^{(2)}) + u \quad \text{under } H_1\text{: } \varphi < 1 \end{aligned}$$

Then, write

$$\hat{\varphi}^{(\lambda)} - 1 = \frac{y'_{-1}Q^{(\lambda)}y}{y'_{-1}Q^{(\lambda)}y_{-1}} - 1 = \frac{y'_{-1}Q^{(\lambda)}\left(y_{-1} + e_T \otimes \beta + u\right)}{y'_{-1}Q^{(\lambda)}y_{-1}} - 1 = \frac{y'_{-1}Q^{(\lambda)}u}{y'_{-1}Q^{(\lambda)}y_{-1}},$$

since $Q^{(\lambda)}\left(e_T \otimes \beta\right) = 0_T$ (see also Remark A1). By substituting backwards, y_{-1} can be written under H_0: $\varphi = 1$ as

$$y_{-1} = e_T \otimes y_0 + \Lambda\left(e_T \otimes \beta\right) + \Lambda u, \qquad (43)$$

and, thus, $y'_{-1}Q^{(\lambda)}u = \left(e_T \otimes y_0 + \Lambda\left(e_T \otimes \beta\right) + \Lambda u\right)' Q^{(\lambda)}u = u'\Lambda' Q^{(\lambda)}u$, since $Q^{(\lambda)}\left[e_T \otimes y_0 + \Lambda\left(e_T \otimes \beta\right)\right] = 0_T$ by Remark A1. Using (43), it can be shown that $y'_{-1}Q^{(\lambda)}y_{-1} = u'\Lambda' Q^{(\lambda)}\Lambda u$. Then, following similar arguments to those for the proof of Lemma A3 it can be shown $p\lim_N(\hat{\varphi}^{(\lambda)} - 1 - \frac{tr(\Lambda' Q^{(\lambda)}\Gamma)}{tr(\Lambda' Q^{(\lambda)}\Lambda\Gamma)}) = 0$.

Lemma A9. For model (19), under Assumptions B and C we have $p\lim_N\left[\hat{\Gamma}_T - \Gamma_T - \beta_T^2 e_T e_T'\right] = 0$.

Proof: Under H_0: $\varphi = 1$, model (42) implies $\Delta y = e_T \otimes \beta + u$, or $\Delta y_i^* = \beta_i e_T + u_i^*$. Then, $\hat{\Gamma}_T$ can be written as

$$\begin{aligned} \hat{\Gamma}_T &= \frac{1}{N}\sum_{i=1}^{N} \Delta y_i^* \Delta y_i^{*\prime} = \frac{1}{N}\sum_{i=1}^{N}(\beta_i e_T + u_i^*)(\beta_i e_T + u_i^*)' \\ &= \frac{1}{N}\sum_{i=1}^{N}\left(\beta_i^2 e_T e_T' + \beta_i e_T u_i^{*\prime} + \beta_i u_i^* e_T' + u_i^* u_i^{*\prime}\right). \end{aligned}$$

The last relationship shows that matrix $\hat{\Gamma}_T$ has elements of the form $\frac{1}{N}\sum_{i=1}^{N}\left(\beta_i^2 + \beta_i u_{i\mu\nu} + \beta_i u_{i\nu\mu} + u_{i\mu\nu}^2\right)$, where $E\left[\beta_i^2 + \beta_i u_{i\mu\nu} + \beta_i u_{i\nu\mu} + u_{i\mu\nu}^2\right] = E(\beta_i^2) + \gamma_{i\mu\nu T}$. Also by Assumptions B and C, it can be shown that $Var(\beta_i^2 + \beta_i u_{i\mu\nu} + \beta_i u_{i\nu m} + u_{i\mu\nu}^2)$ is finite. Then, by Chebyshev's Weak Law of Large Numbers we can obtain the following result:

$$p\lim_N\left[\frac{1}{N}\sum_{i=1}^{N}\left(\beta_i^2 + \beta_i u_{i\mu\nu} + \beta_i u_{i\nu m} + u_{i\mu\nu}^2 - E(\beta_i^2) - \gamma_{i\mu\nu T}\right)\right] = 0,$$

which implies $p\lim_N(\hat{\Gamma}_T - \Gamma_T - \beta_T^2 e_T e_T') = 0_T$, where $\beta_T^2 = \frac{1}{N}\sum_{i=1}^{N} E(\beta_i^2)$.

Lemma A10. For model (19), under Assumptions B and C we have $p\lim_N \left[\frac{tr(M_T\hat{\Gamma}_T)}{tr(M_T e_T e_T')} - \beta_T^2\right] = 0$.

Proof: First, write $tr(M_T\hat{\Gamma}_T) = \frac{1}{N}\sum_{i=1}^N \Delta y_i^{*\prime} M_T \Delta y_i^*$. Next, note that

$$\begin{aligned} E(\Delta y_i^{*\prime} M_T \Delta y_i^*) &= tr\left[M_T E(\Delta y_i^* \Delta y_i^{*\prime})\right] \quad \text{and} \\ E(\Delta y_i^* \Delta y_i^{*\prime}) &= E(\beta_i^2 e_T e_T' + \beta_i e_T u_i^{*\prime} + \beta_i u_i^* e_T' + u_i^* u_i^{*\prime}) = E(\beta_i^2) e_T e_T' + \Gamma_{iT} \end{aligned}$$

and, by Assumptions B and C. These results imply that $tr\left[M_T E(\Delta y_i^* \Delta y_i^{*\prime})\right] = E(\beta_i^2) tr(M_T e_T e_T')$, since $tr(M_T\Gamma_{iT}) = 0$. The latter holds, for all i, by the assumption that the maximum order of serial correlation, $p_{\max}$, is less than T. Also, note that the variances of $\Delta y_i^* \Delta y_i^{*\prime}$ are finite, for all i. Then, by Chebyshev's Weak Law of Large Numbers we can prove the following result: $p\lim_N \left[tr(M_T\hat{\Gamma}_T) - \beta_T^2 tr(M_T e_T e_T')\right] = 0$. The result of the lemma follows by dividing the last relationship with $tr(M_T e_T e_T')$.

Lemma A11. For model (19), under Assumptions B and C the following result holds:

$$p\lim_N \left[\frac{1}{N} tr(\Phi^{(\lambda)}\hat{\Gamma}) - \frac{1}{N} tr(\Lambda' Q^{(\lambda)} \Gamma)\right] = 0.$$

Proof: First, write $\frac{1}{N} tr(\Lambda' Q^{(\lambda)}\Gamma) = \frac{1}{N} tr(\Lambda' Q^{(\lambda)} E(uu')) = \frac{1}{N} E(u'\Lambda' Q^{(\lambda)} u)$. By Lemma A1, we have

$$\frac{1}{N} u'\Lambda' Q^{(\lambda)} u = \frac{1}{N}\sum_{i=1}^N u_i^{*\prime}\Lambda_T' Q_T^{(\lambda)} u_i^* = \frac{1}{N}\sum_{i=1}^N tr(\Lambda_T' Q_T^{(\lambda)} u_i^* u_i^{*\prime})$$

and thus, $E\left[\frac{1}{N}\sum_{i=1}^N tr(\Lambda_T' Q_T^{(\lambda)} u_i^* u_i^{*\prime})\right] = tr(\Lambda_T' Q_T^{(\lambda)} \Gamma_T)$. Also, write $\Phi^{(\lambda)} = \Phi_T^{(\lambda)} \otimes I_N$, where $\Phi_T^{(\lambda)} = \Psi_T^{(\lambda)} - tr(\Psi_T^{(\lambda)} e_T e_T') \frac{M_T}{tr(M_T e_T e_T')}$. Then, it can be easily seen that

$$\frac{1}{N} tr(\Phi^{(\lambda)}\hat{\Gamma}) = tr(\Phi_T^{(\lambda)}\hat{\Gamma}_T) = \frac{1}{N}\sum_{i=1}^N \Delta y_i^{*\prime}\Phi_T^{(\lambda)}\Delta y_i^*, \text{ with}$$

$$E(\Delta y_i^{*\prime}\Phi_T^{(\lambda)}\Delta y_i^*) = tr[\Phi_T^{(\lambda)}\left(E(\beta_i^2) e_T e_T' + \Gamma_{iT}\right)] = tr(\Phi_T^{(\lambda)}\Gamma_{iT}), \tag{44}$$

since $tr(\Phi_T^{(\lambda)} e_T e_T') = tr\left[\Psi_T^{(\lambda)} e_T e_T' - tr(\Psi_T^{(\lambda)} e_T e_T') \frac{M_T^{(\lambda)} e_T e_T'}{tr(M_T e_T e_T')}\right] = tr(\Psi_T^{(\lambda)} e_T e_T') - tr(\Psi_T^{(\lambda)} e_T e_T') \frac{tr(M_T e_T e_T')}{tr(M_T e_T e_T')} = 0$.

To show that $tr(\Phi_T^{(\lambda)}\Gamma_{iT}) = tr(\Lambda_T' Q_T^{(\lambda)}\Gamma_{iT})$, note that $tr(\Phi_T^{(\lambda)}\Gamma_{iT})$ can be written as

$$tr(\Phi_T^{(\lambda)}\Gamma_{iT}) = tr(\Psi_T^{(\lambda)}\Gamma_{iT} - tr(\Psi_T^{(\lambda)} e_T e_T') \frac{M_T\Gamma_{iT}}{tr(M_T e_T e_T')}) = tr(\Psi_T^{(\lambda)}\Gamma_{iT}),$$

since $tr(M_T\Gamma_{iT}) = 0$. By the definition of matrix $\Psi_T^{(\lambda)}$ (see Theorem 3), we have $tr(\Psi_T^{(\lambda)}\Gamma_{iT}) = tr(\Lambda_T' Q_T^{(\lambda)}\Gamma_{iT})$, and thus $tr(\Phi_T^{(\lambda)}\Gamma_{iT}) = tr(\Psi_T^{(\lambda)}\Gamma_{iT}) = tr(\Lambda_T' Q_T^{(\lambda)}\Gamma_{iT})$. Then, it follows that $E(\Delta y_i^{*\prime}\Phi_T^{(\lambda)}\Delta y_i^*) = tr(\Lambda_T' Q_T^{(\lambda)}\Gamma_{iT})$ by (44). The result of the lemma follows immediately by applying Chebyshev's WLLN to $\frac{1}{N}\sum_{i=1}^N \Delta y_i^{*\prime}\Phi_T^{(\lambda)}\Delta y_i^*$ (see also Lemma A10).

Proof of Theorem 4: Under H_0: $\varphi = 1$ and Lemma A8, test statistic $Z^{(\lambda)}$ can be written as

$$\hat{d}^{(\lambda)}\sqrt{N}\left(\hat{\varphi}^{(\lambda)} - 1 - \frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}\right) = \hat{d}^{(\lambda)}\sqrt{N}\left(\frac{\frac{1}{N}y_{-1}'Q^{(\lambda)}y}{\hat{d}^{(\lambda)}} - 1 - \frac{\frac{1}{N}tr(\Phi^{(\lambda)}\hat{\Gamma})}{\hat{d}^{(\lambda)}}\right) = \sqrt{N}\left(\frac{1}{N}u'\Lambda'Q^{(\lambda)}u - \frac{1}{N}tr(\Phi^{(\lambda)}\hat{\Gamma})\right).$$

By Remark A4, the last relationship can be written as follows:

$$\sqrt{N}\left(\frac{1}{N}u'\Lambda'Q^{(\lambda)}u - \frac{1}{N}tr(\Phi^{(\lambda)}\hat{\Gamma})\right) = \frac{1}{\sqrt{N}}\left(u'\Lambda'Q^{(\lambda)}u - \Delta y'\Phi^{(\lambda)}\Delta y\right)$$

Substituting into this $\Delta y'\Lambda'Q^{(\lambda)}\Delta y = (u' + e_T' \otimes \beta')\Lambda'Q^{(\lambda)}(u + e_T \otimes \beta) = u'\Lambda'Q^{(\lambda)}u$, which holds under H_0: $\varphi = 1$, yields

$$\frac{1}{\sqrt{N}}\left(u'\Lambda'Q^{(\lambda)}u - \Delta y'\Phi^{(\lambda)}\Delta y\right) = \frac{1}{\sqrt{N}}\left(\Delta y'\Lambda'Q^{(\lambda)}\Delta y - \Delta y'\Phi^{(\lambda)}\Delta y\right) = \frac{1}{\sqrt{N}}(\Delta y'(\Lambda'Q^{(\lambda)} - \Phi^{(\lambda)})\Delta y).$$

By Lemma A1, the last relationship can be written as

$$\frac{1}{\sqrt{N}}(\Delta y'(\Lambda'Q^{(\lambda)} - \Phi^{(\lambda)})\Delta y) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Delta y_i^{*\prime}(\Lambda_T'Q_T^{(\lambda)} - \Phi_T^{(\lambda)})\Delta y_i^*$$

and it has

$$E\left[\Delta y_i^{*\prime}(\Lambda_T'Q_T^{(\lambda)} - \Phi_T^{(\lambda)})\Delta y_i^*\right] = tr\left[(\Lambda_T'Q_T^{(\lambda)} - \Phi_T^{(\lambda)})E(\Delta y_i^*\Delta y_i^{*\prime})\right] = 0,$$

since $E(\Delta y_i^*\Delta y_i^{*\prime}) = E[(u_i^* + \beta_i e_T)(u_i^* + \beta_i e_T)] = \Gamma_{iT} + E(\beta_i^2)e_T e_T'$ and $tr\left[(\Lambda_T'Q_T^{(\lambda)} - \Phi_T^{(\lambda)})(\Gamma_{iT} + E(\beta_i^2)e_T e_T')\right] = 0$ (see also Lemma A.11). The variance of $\Delta y_i^{*\prime}(\Lambda_T'Q_T^{(\lambda)} - \Phi_T^{(\lambda)})\Delta y_i^*$ is given as

$$\begin{aligned} Var\left[\Delta y_i^{*\prime}(\Lambda_T'Q_T^{(\lambda)} - \Phi_T^{(\lambda)})\Delta y_i^*\right] &= Var\left\{tr\left[(\Lambda_T'Q_T^{(\lambda)} - \Phi_T^{(\lambda)})E(\Delta y_i^*\Delta y_i^{*\prime})\right]\right\} \\ &= vec(Q_T^{(\lambda)}\Lambda_T - \Phi_T^{(\lambda)\prime})'Var(vec(\Delta y_i^*\Delta y_i^{*\prime}))vec(Q_T^{(\lambda)}\Lambda_T - \Phi_T^{(\lambda)\prime}) \end{aligned}$$

By Assumptions B and C, and by the Lindeberg-Feller CLT it follows that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Delta y_i^{*\prime}(\Lambda_T'Q_T^{(\lambda)} - \Phi_T^{(\lambda)})\Delta y_i^* \xrightarrow{d} N\left(0, vec(Q_T^{(\lambda)}\Lambda_T - \Phi_T^{(\lambda)\prime})'\Theta_{Tu}vec(Q_T^{(\lambda)}\Lambda_T - \Phi_T^{(\lambda)\prime})\right),$$

where $\Theta_{Tu} = p\lim_N \Theta_T = p\lim_N \frac{1}{N}\sum_{i=1}^{N} Var(vec(\Delta y_i^*\Delta y_i^{*\prime}))$. The $vec(.)$ (stacked vector) notation of the variance of the above limiting distribution comes by noticing that

$$\begin{aligned} Var\left[\Delta y'(\Lambda'Q^{(\lambda)} - \Phi^{(\lambda)})\Delta y\right] &= Var\left\{tr\left[(\Lambda'Q^{(\lambda)} - \Phi^{(\lambda)})E(\Delta y\Delta y')\right]\right\} \\ &= vec(Q^{(\lambda)}\Lambda - \Phi^{(\lambda)\prime})'Var(vec(\Delta y\Delta y'))vec(Q^{(\lambda)}\Lambda - \Phi^{(\lambda)\prime}). \end{aligned}$$

The next lemma provides a consistent estimator of Θ_T, entering the above variance function.

Lemma A12. Under Assumptions B and C, the following result holds for model (19) under H_0: $\varphi = 1$:

$$p\lim_N\left[\hat{\Theta}_T^* - \Theta_T\right] = 0, \text{ where}$$

$\hat{\Theta}_T^* = \frac{1}{N}\sum_{i=1}^N vec(\Delta y_i^* \Delta y_i^{*\prime}) vec(\Delta y_i^* \Delta y_i^{*\prime})' - \frac{1}{N}\sum_{i=1}^N vec(\Delta y_i^* \Delta y_i^{*\prime}) \frac{1}{N}\sum_{i=1}^N vec(\Delta y_i^* \Delta y_i^{*\prime})'$.

Proof: First, notice that

$$Var(vec(\Delta y_i^* \Delta y_i^{*\prime})) = E\left[vec(\Delta y_i^* \Delta y_i^{*\prime}) vec(\Delta y_i^* \Delta y_i^{*\prime})'\right] - E\left[vec(\Delta y_i^* \Delta y_i^{*\prime})\right] E\left[vec(\Delta y_i^* \Delta y_i^{*\prime})'\right].$$

Then, the proof of the lemma follows immediately, by showing element by element convergence as in Lemma A7.

Remark A5. Note that the estimator $\hat{\Theta}_T = \frac{1}{N}\sum_{i=1}^N vec(\Delta y_i^* \Delta y_i^{*\prime}) vec(\Delta y_i^* \Delta y_i^{*\prime})'$, suggested in the main text (see (22)), is a consistent estimator of $\frac{1}{N}\sum_{i=1}^N E\left[vec(\Delta y_i^* \Delta y_i^{*\prime}) vec(\Delta y_i^* \Delta y_i^{*\prime})'\right]$, not of Θ_T, i.e.,

$$p\lim_N [\frac{1}{N}\sum_{i=1}^N vec(\Delta y_i^* \Delta y_i^{*\prime}) vec(\Delta y_i^* \Delta y_i^{*\prime})' - \frac{1}{N}\sum_{i=1}^N E\left[vec(\Delta y_i^* \Delta y_i^{*\prime}) vec(\Delta y_i^* \Delta y_i^{*\prime})'\right]] = 0.$$

This can be proved following similar arguments to those for the proof with Lemma A12. However, when $\hat{\Theta}_T$ is plugged in the variance, it makes $\tilde{F}_T^{(\lambda)\prime}\hat{\Theta}_T \tilde{F}_T^{(\lambda)}$ a consistent estimator of variance $V^{(\lambda)}$, used by test statistic $Z^{(\lambda)}$ (see (20)). To see this, note that, under the assumption that the order of serial correlation $p \leq p_{\max}$ is the same for all i, we have:

$$\begin{aligned} & vec(Q_T^{(\lambda)}\Lambda_T - \Phi_T^{(\lambda)\prime})' Var(vec(\Delta y_i^* \Delta y_i^{*\prime})) vec(Q_T^{(\lambda)}\Lambda_T - \Phi_T^{(\lambda)\prime}) \\ = \; & vec(Q_T^{(\lambda)}\Lambda_T - \Phi_T^{(\lambda)\prime})' E\left[vec(\Delta y_i^* \Delta y_i^{*\prime}) vec(\Delta y_i^* \Delta y_i^{*\prime})'\right] vec(Q_T^{(\lambda)}\Lambda_T - \Phi_T^{(\lambda)\prime}) \end{aligned}$$

since $vec(Q_T^{(\lambda)}\Lambda_T - \Phi_T^{(\lambda)\prime})' E\left[vec(\Delta y_i^* \Delta y_i^{*\prime})\right] = tr\left[(\Lambda_T' Q_T^{(\lambda)} - \Phi_T^{(\lambda)})(\Gamma_{iT} + E(\beta_i^2) e_T e_T')\right] = 0$ by Lemma A11. The last result indicates that we only need to estimate $E\left[vec(\Delta y_i^* \Delta y_i^{*\prime}) vec(\Delta y_i^* \Delta y_i^{*\prime})'\right]$. $\hat{\Theta}_T$ and $\hat{\Theta}_T^*$ can be both used to consistently estimate $V^{(\lambda)}$, but $\hat{\Theta}_T$ may be preferred since it is computationally less demanding.

Lemma A13. For model (23), under Assumption B we have $p\lim_N (\hat{\varphi}^{(\lambda)} - 1 - \frac{tr(\Lambda' Q^{(\lambda)}\Gamma)}{tr(\Lambda' Q^{(\lambda)}\Lambda\Gamma)}) = 0$.

Proof: As in Lemma A.3, write (23) as

$$\begin{aligned} y \;=\; & \varphi y_{-1} + \varphi(e_T^{(1)} \otimes \beta^{(1)} + e_T^{(2)} \otimes \beta^{(2)}) + (1-\varphi)(e_T^{(1)} \otimes a^{(1)} + e_T^{(2)} \otimes a^{(2)}) \\ & + (1-\varphi)(\tau_T^{(1)} \otimes \beta^{(1)} + \tau_T^{(2)} \otimes \beta^{(2)}) + u \end{aligned}$$

Under H_0: $\varphi = 1$ the last relationship yields

$$y = y_{-1} + e_T^{(1)} \otimes \beta^{(1)} + e_T^{(2)} \otimes \beta^{(2)} + u. \tag{45}$$

Using this, $\hat{\varphi}^{(\lambda)} - 1$ can be written as follows:

$$\hat{\varphi}^{(\lambda)} - 1 = \frac{y_{-1}' Q^{(\lambda)} y}{y_{-1}' Q^{(\lambda)} y_{-1}} - 1 = \frac{y_{-1}' Q^{(\lambda)} \left(y_{-1} + e_T^{(1)} \otimes \beta^{(1)} + e_T^{(2)} \otimes \beta^{(2)} + u\right)}{y_{-1}' Q^{(\lambda)} y_{-1}} - 1 = \frac{y_{-1}' Q^{(\lambda)} u}{y_{-1}' Q^{(\lambda)} y_{-1}},$$

since $Q^{(\lambda)}\left(e_T^{(1)}\otimes\beta^{(1)}\right) = Q^{(\lambda)}\left(e_T^{(2)}\otimes\beta^{(2)}\right) = 0_T$ by construction (see also Remark A1). Substituting backwards y_{-1} gives

$$y_{-1} = e_T\otimes y_0 + \Lambda\left(e_T^{(1)}\otimes\beta^{(1)}\right) + \Lambda\left(e_T^{(2)}\otimes\beta^{(2)}\right) + \Lambda u. \tag{46}$$

Using this result, $y_{-1}'Q^{(\lambda)}u$ can be written as

$$y_{-1}'Q^{(\lambda)}u = \left(e_T\otimes y_0 + \Lambda\left(e_T^{(1)}\otimes\beta^{(1)}\right) + \Lambda\left(e_T^{(2)}\otimes\beta^{(2)}\right) + \Lambda u\right)' Q^{(\lambda)}u = u'\Lambda'Q^{(\lambda)}u,$$

since $Q^{(\lambda)}\left[e_T\otimes y_0 + \Lambda\left(e_T^{(1)}\otimes\beta^{(1)}\right) + \Lambda\left(e_T^{(2)}\otimes\beta^{(2)}\right)\right] = 0_{NT}$ by Remark A1. Following similar steps to the above, it can be shown that the numerator of $\hat{\varphi}^{(\lambda)}-1$ can be written as $y_{-1}'Q^{(\lambda)}y_{-1} = u'\Lambda'Q^{(\lambda)}\Lambda u$. Given the above results, it can be proved that $p\lim_N(\hat{\varphi}^{(\lambda)} - 1 - \frac{tr(\Lambda'Q^{(\lambda)}\Gamma)}{tr(\Lambda'Q^{(\lambda)}\Lambda\Gamma)}) = 0$, following analogous arguments to those for the proof of Lemma 3.

Lemma A14. For model (23), under Assumptions B and C the following result holds:

$$p\lim_N\left[\hat{\Gamma}_T - \Gamma_T - \beta_T^{(1)2}e_T^{(1)}e_T^{(1)\prime} - \beta_T^{(2)2}e_T^{(2)}e_T^{(2)\prime}\right] = 0.$$

Proof: First, note that, under H_0: $\varphi = 1$, model (45) implies that $\Delta y = e_T^{(1)}\otimes\beta^{(1)} + e_T^{(2)}\otimes\beta^{(2)} + u$, or $\Delta y_i^* = \beta_i^{(1)}e_T^{(1)} + \beta_i^{(2)}e_T^{(2)} + u_i^*$. Then, $\hat{\Gamma}_T$ can be written as

$$\hat{\Gamma}_T = \frac{1}{N}\sum_{i=1}^{N}\Delta y_i^*\Delta y_i^{*\prime} = \frac{1}{N}\sum_{i=1}^{N}(\beta_i^{(1)}e_T^{(1)} + \beta_i^{(2)}e_T^{(2)} + u_i^*)(\beta_i^{(1)}e_T^{(1)} + \beta_i^{(2)}e_T^{(2)} + u_i^*)'.$$

Based on the last relationship, the lemma can be proved following the same arguments with those for the proof of Lemma A9.

Lemma A15. For model (23), under Assumptions B and C the following results hold:

$$p\lim_N\left[\frac{tr(M_T^{(1)}\hat{\Gamma}_T)}{tr(M_T^{(1)}e_T^{(1)}e_T^{(1)\prime})} - \beta_T^{(1)2}\right] = 0 \quad\text{and}\quad p\lim_N\left[\frac{tr(M_T^{(2)}\hat{\Gamma}_T)}{tr(M_T^{(2)}e_T^{(2)}e_T^{(2)\prime})} - \beta_T^{(2)2}\right] = 0.$$

Proof: First, write $tr(M_T^{(1)}\hat{\Gamma}_T) = \frac{1}{N}\sum_{i=1}^N \Delta y_i^{*\prime}M_T^{(1)}\Delta y_i^*$, with $E(\Delta y_i^{*\prime}M_T^{(1)}\Delta y_i^*) = tr\left[M_T^{(1)}E(\Delta y_i^*\Delta y_i^{*\prime})\right]$. By Assumptions B and C, we have $E(\Delta y_i^*\Delta y_i^{*\prime}) = E(\beta_i^{(1)2})e_T^{(1)}e_T^{(1)\prime} + E(\beta_i^{(2)2})e_T^{(2)}e_T^{(2)\prime} + \Gamma_{iT}$. Thus, $tr\left[M_T^{(1)}E(\Delta y_i^*\Delta y_i^{*\prime})\right]$ can be written as

$$tr\left[M_T^{(1)}E(\Delta y_i^*\Delta y_i^{*\prime})\right] = E(\beta_i^{(1)2})tr(M_T^{(1)}e_T^{(1)}e_T^{(1)\prime}),$$

since $tr(M_T^{(1)}\Gamma_{iT}) = tr(M_T^{(1)}e_T^{(2)}e_T^{(2)\prime}) = 0$, for all i, due to condition $p_{\max} < T$ of Assumption B. Also, it can be easily seen that the variance of $\Delta y_i^*\Delta y_i^{*\prime}$ is finite, for all i. Given the above results, it can be proved that $p\lim_N\left[tr(M_T^{(1)}\hat{\Gamma}_T) - \beta_T^{(1)2}tr(M_T^{(1)}e_T^{(1)}e_T^{(1)\prime})\right] = 0$, by Chebyshev's Weak Law of Large Number. By dividing last relationship with $tr(M_T^{(1)}e_T^{(1)}e_T^{(1)\prime})$ proves the lemma.

Lemma A16. For model (23), under Assumptions B and C the following result holds:

$$p\lim_N\left[\frac{1}{N}tr(\Omega^{(\lambda)}\hat{\Gamma}) - \frac{1}{N}tr(\Lambda'Q^{(\lambda)}\Gamma)\right] = 0.$$

Proof: Write $\Omega^{(\lambda)} = \Omega_T^{(\lambda)}\otimes I_N$, where $\Omega_T^{(\lambda)} = \Psi_T^{(\lambda)} - tr(\Psi_T^{(\lambda)}e_T^{(1)}e_T^{(1)\prime})\frac{M_T^{(1)}}{tr(M_T^{(1)}e_T^{(1)}e_T^{(1)\prime})} - tr(\Psi_T^{(\lambda)}e_T^{(2)}e_T^{(2)\prime})\frac{M_T^{(2)}}{tr(M_T^{(2)}e_T^{(2)}e_T^{(2)\prime})}$.

Then, $\frac{1}{N}tr(\Omega^{(\lambda)}\hat{\Gamma})$ can be written as follows:

$$\frac{1}{N}tr(\Omega^{(\lambda)}\hat{\Gamma}) = tr(\Omega_T^{(\lambda)}\hat{\Gamma}_T) = \frac{1}{N}\sum_{i=1}^{N}\Delta y_i^{*\prime}\Omega_T^{(\lambda)}\Delta y_i^{*}.$$

As in Lemma A15, it can be shown that

$$E(\Delta y_i^{*\prime}\Omega_T^{(\lambda)}\Delta y_i^{*}) = tr[\Omega_T^{(\lambda)}\left(E(\beta_i^{(1)2})e_T^{(1)}e_T^{(1)\prime} + E(\beta_i^{(2)2})e_T^{(2)}e_T^{(2)\prime} + \Gamma_{iT}\right)] = tr(\Omega_T^{(\lambda)}\Gamma_{iT}),$$

using the following two results:

$$\begin{aligned} tr(\Omega_T^{(\lambda)}e_T^{(1)}e_T^{(1)\prime}) &= tr\left[\Psi_T^{(\lambda)}e_T^{(1)}e_T^{(1)\prime} - tr(\Psi_T^{(\lambda)}e_T^{(1)}e_T^{(1)\prime})\tfrac{M_T^{(1)}e_T^{(1)}e_T^{(1)\prime}}{tr(M_T^{(1)}e_T^{(1)}e_T^{(1)\prime})} - tr(\Psi_T^{(\lambda)}e_T^{(2)}e_T^{(2)\prime})\tfrac{M_T^{(2)}e_T^{(1)}e_T^{(1)\prime}}{tr(M_T^{(2)}e_T^{(2)}e_T^{(2)\prime})}\right] \\ &= tr(\Psi_T^{(\lambda)}e_T^{(1)}e_T^{(1)\prime}) - tr(\Psi_T^{(\lambda)}e_T^{(1)}e_T^{(1)\prime})\tfrac{tr(M_T^{(1)}e_T^{(1)}e_T^{(1)\prime})}{tr(M_T^{(1)}e_T^{(1)}e_T^{(1)\prime})} - tr(\Psi_T^{(\lambda)}e_T^{(2)}e_T^{(2)\prime})\tfrac{tr(M_T^{(2)}e_T^{(1)}e_T^{(1)\prime})}{tr(M_T^{(2)}e_T^{(2)}e_T^{(2)\prime})} = 0, \end{aligned}$$

as $tr(M_T^{(2)}e_T^{(1)}e_T^{(1)\prime}) = 0$, and

$$\begin{aligned} tr(\Omega_T^{(\lambda)}e_T^{(2)}e_T^{(2)\prime}) &= tr\left[\Psi_T^{(\lambda)}e_T^{(2)}e_T^{(2)\prime} - tr(\Psi_T^{(\lambda)}e_T^{(1)}e_T^{(1)\prime})\tfrac{M_T^{(1)}e_T^{(2)}e_T^{(2)\prime}}{tr(M_T^{(1)}e_T^{(1)}e_T^{(1)\prime})} - tr(\Psi_T^{(\lambda)}e_T^{(2)}e_T^{(2)\prime})\tfrac{M_T^{(2)}e_T^{(2)}e_T^{(2)\prime}}{tr(M_T^{(2)}e_T^{(2)}e_T^{(2)\prime})}\right] \\ &= tr(\Psi_T^{(\lambda)}e_T^{(2)}e_T^{(2)\prime}) - tr(\Psi_T^{(\lambda)}e_T^{(1)}e_T^{(1)\prime})\tfrac{tr(M_T^{(1)}e_T^{(2)}e_T^{(2)\prime})}{tr(M_T^{(1)}e_T^{(1)}e_T^{(1)\prime})} - tr(\Psi_T^{(\lambda)}e_T^{(2)}e_T^{(2)\prime})\tfrac{tr(M_T^{(2)}e_T^{(2)}e_T^{(2)\prime})}{tr(M_T^{(2)}e_T^{(2)}e_T^{(2)\prime})} = 0, \end{aligned}$$

as $tr(M_T^{(1)}e_T^{(2)}e_T^{(2)\prime}) = 0$.

Also note that

$$\begin{aligned} tr(\Omega_T^{(\lambda)}\Gamma_{iT}) &= tr(\Psi_T^{(\lambda)}\Gamma_{iT} - tr(\Psi_T^{(\lambda)}e_T^{(1)}e_T^{(1)\prime})\frac{M_T^{(1)}\Gamma_{iT}}{tr(M_T^{(1)}e_T^{(1)}e_T^{(1)\prime})} - tr(\Psi_T^{(\lambda)}e_T^{(2)}e_T^{(2)\prime})\frac{M_T^{(2)}\Gamma_{iT}}{tr(M_T^{(2)}e_T^{(2)}e_T^{(2)\prime})}) \\ &= tr(\Psi_T^{(\lambda)}\Gamma_{iT}), \end{aligned}$$

since $tr(M_T^{(1)}\Gamma_{iT}) = tr(M_T^{(2)}\Gamma_{iT}) = 0$. By the definition of selection matrix $\Psi_T^{(\lambda)}$, we have $tr(\Psi_T^{(\lambda)}\Gamma_{iT}) = tr(\Lambda_T' Q_T^{(\lambda)}\Gamma_{iT})$. Thus, combining the above results yields $E(\Delta y_i^{*\prime}\Omega_T^{(\lambda)}\Delta y_i^{*}) = tr(\Lambda_T' Q_T^{(\lambda)}\Gamma_{iT})$. Then, the lemma can be proved by Chebyshev's WLLN (see also Lemma A15).

Proof of Theorem 5: The theorem can be proved following similar steps to those for the proof of Theorem 4.

Section E (Lemmas and Theorem Proofs for Section 5): The following lemmas are required for the proof of Theorem 6.

Lemma A17. For model (25), under Assumption B the following result hold:

$$p\lim_N(\hat{\varphi}^{(\lambda)} - 1 - \tfrac{tr(\Lambda' Q^{(\lambda)}\Gamma)}{tr(\Lambda' Q^{(\lambda)}\Lambda\Gamma)}) = 0.$$

Proof: First, write model (25) as

$$\begin{aligned} y &= \varphi y_{-1} + \varphi(e_T^{(1)}\otimes\beta^{(1)} + e_T^{(2)}\otimes\beta^{(2)}) + \varphi(\tilde{e}_T^{(1)}\otimes\delta^{(1)} + \tilde{e}_T^{(2)}\otimes\delta^{(2)}) \\ &\quad +(1-\varphi)(e_T^{(1)}\otimes a^{(1)} + e_T^{(2)}\otimes a^{(2)}) + (1-\varphi)(\tau_T^{(1)}\otimes\beta^{(1)} + \tau_T^{(2)}\otimes\beta^{(2)}) \\ &\quad +(1-\varphi)(\tau_{2T}^{(1)}\otimes\delta^{(1)} + \tau_{2T}^{(2)}\otimes\delta^{(2)}) + u \end{aligned}$$

Under H_0: $\varphi = 1$, the last relationship becomes:

$$y = y_{-1} + e_T^{(1)} \otimes \beta^{(1)} + e_T^{(2)} \otimes \beta^{(2)} + \tilde{e}_T^{(1)} \otimes \delta^{(1)} + \tilde{e}_T^{(2)} \otimes \delta^{(2)} + u. \tag{47}$$

Based on this relationship, $\hat{\varphi}^{(\lambda)} - 1$ can be written as

$$\hat{\varphi}^{(\lambda)} - 1 = \frac{y'_{-1}Q^{(\lambda)}y}{y'_{-1}Q^{(\lambda)}y_{-1}} - 1 = \frac{y'_{-1}Q^{(\lambda)}\left(y_{-1} + \sum_{j=1}^{2} e_T^{(j)} \otimes \beta^{(j)} + \sum_{j=1}^{2} \tilde{e}_T^{(j)} \otimes \delta^{(j)} + u\right)}{y'_{-1}Q^{(\lambda)}y_{-1}} - 1 = \frac{y'_{-1}Q^{(\lambda)}u}{y'_{-1}Q^{(\lambda)}y_{-1}},$$

since $Q^{(\lambda)}\left(e_T^{(j)} \otimes \beta^{(j)}\right) = Q^{(\lambda)}\left(e_T^{(j)} \otimes \delta^{(j)}\right) = 0_T$ (see Remark A1). Using

$$y_{-1} = e_T \otimes y_0 + \sum_{j=1}^{2} \Lambda\left(e_T^{(j)} \otimes \beta^{(j)}\right) + \sum_{j=1}^{2} \Lambda\left(\tilde{e}_T^{(j)} \otimes \delta^{(j)}\right) + \Lambda u,$$

$y'_{-1}Q^{(\lambda)}u$ can be written as

$$y'_{-1}Q^{(\lambda)}u = \left(e_T \otimes y_0 + \sum_{j=1}^{2} \Lambda\left(e_T^{(j)} \otimes \beta^{(j)}\right) + \sum_{j=1}^{2} \Lambda\left(\tilde{e}_T^{(j)} \otimes \delta^{(j)}\right) + \Lambda u\right)' Q^{(\lambda)}u = u'\Lambda' Q^{(\lambda)}u, \tag{48}$$

as $Q^{(\lambda)}\left[e_T \otimes y_0 + \sum_{j=1}^{2} \Lambda\left(e_T^{(j)} \otimes \beta^{(j)}\right) + \sum_{j=1}^{2} \Lambda\left(\tilde{e}_T^{(j)} \otimes \delta^{(j)}\right)\right] = 0_T$ (see Remark A1). Following analogous arguments to the above, we can show that $y'_{-1}Q^{(\lambda)}y_{-1} = u'\Lambda' Q^{(\lambda)}\Lambda u$. Then, it can be easily shown $p\lim_N(\hat{\varphi}^{(\lambda)} - 1 - \frac{tr(\Lambda' Q^{(\lambda)}\Gamma)}{tr(\Lambda' Q^{(\lambda)}\Lambda\Gamma)}) = 0$ (as in Lemma A3).

Lemma A18. For model (25), under Assumptions B, C and D the following result holds:

$$p\lim_N\left[\hat{\Gamma}_T - \Gamma_T - \sum_{j=1}^{2} \beta_T^{(j)2} e_T^{(j)} e_T^{(j)\prime} - \sum_{j=1}^{2} \delta_T^{(j)2} \tilde{e}_T^{(j)} \tilde{e}_T^{(j)\prime}\right] = 0.$$

Proof: Under H_0: $\varphi = 1$, model implies $\Delta y = \sum_{j=1}^{2} e_T^{(j)} \otimes \beta^{(j)} + \sum_{j=1}^{2} \tilde{e}_T^{(j)} \otimes \delta^{(j)} + u$ and $\Delta y_i^* = \sum_{j=1}^{2} \beta_i^{(j)} e_T^{(j)} + \sum_{j=1}^{2} \delta_i^{(j)} \tilde{e}_T^{(j)} + u_i^*$. The proof of the lemma follows, immediately, by substituting the last relationship of Δy_i^* into $\hat{\Gamma}_T = \frac{1}{N}\sum_{i=1}^{N} \Delta y_i^* \Delta y_i^{*\prime}$ and following analogous arguments to those for the proof of Lemma A14.

Lemma A19. For model (23), under Assumptions B, C and D the following results hold:

$$p\lim_N\left[\frac{tr\left(J_T^{(j)}\hat{\Gamma}_T\right)}{tr(M_T^{(j)} e_T^{(j)} e_T^{(j)\prime})} - \beta_T^{(j)2}\right] = 0 \text{ and } p\lim_N\left[\frac{tr(L_T^{(j)}\hat{\Gamma}_T)}{tr(L_T^{(j)} \tilde{e}_T^{(j)} \tilde{e}_T^{(j)\prime})} - \delta_T^{(j)2}\right] = 0, \text{ for } j = 1, 2.$$

Proof: First, note that $tr(J_T^{(j)}\hat{\Gamma}_T) = \frac{1}{N}\sum_{i=1}^{N} \Delta y_i^{*\prime} J_T^{(j)} \Delta y_i^*$. Since $E(\Delta y_i^{*\prime} J_T^{(j)} \Delta y_i^*) = tr\left[J_T^{(j)} E(\Delta y_i^* \Delta y_i^{*\prime})\right]$ and $E(\Delta y_i^* \Delta y_i^{*\prime}) = \sum_{j=1}^{2} E(\beta_i^{(j)2}) e_T^{(j)} e_T^{(j)\prime} + \sum_{j=1}^{2} E(\delta_i^{(j)2}) \tilde{e}_T^{(j)} \tilde{e}_T^{(j)\prime} + \Gamma_{iT}$, $tr\left[J_T^{(j)} E(\Delta y_i^* \Delta y_i^{*\prime})\right]$ can be written as

$$
\begin{aligned}
& tr\left[J_T^{(j)}E(\Delta y_i^*\Delta y_i^{*\prime})\right] \\
= \; & tr\left[\left(M_T^{(j)} - \frac{tr(M_T^{(j)}\tilde{e}_T^{(j)}\tilde{e}_T^{(j)\prime})}{tr\left(L_T^{(j)}\tilde{e}_T^{(j)}\tilde{e}_T^{(j)\prime}\right)}L_T^{(j)}\right)E(\Delta y_i^*\Delta y_i^{*\prime})\right] \\
= \; & tr\left[\left(M_T^{(j)} - \frac{tr(M_T^{(j)}\tilde{e}_T^{(j)}\tilde{e}_T^{(j)\prime})}{tr\left(L_T^{(j)}\tilde{e}_T^{(j)}\tilde{e}_T^{(j)\prime}\right)}L_T^{(j)}\right)\left(\sum_{k=1}^{2}E(\beta_i^{(k)2})e_T^{(k)}e_T^{(k)\prime} + \sum_{k=1}^{2}E(\delta_i^{(k)2})\tilde{e}_T^{(k)}\tilde{e}_T^{(k)\prime} + \Gamma_{iT}\right)\right]
\end{aligned}
$$

Using the following results: $tr(M_T^{(j)}\Gamma_{iT}) = tr(M_T^{(j)}e_T^{(k)}e_T^{(k)\prime}) = 0$, $tr(M_T^{(j)}e_T^{(k)}e_T^{(k)\prime}) = tr(M_T^{(j)}\tilde{e}_T^{(k)}\tilde{e}_T^{(k)\prime}) = tr(L_T^{(j)}\tilde{e}_T^{(k)}\tilde{e}_T^{(k)\prime}) = 0$ and $tr(L_T^{(j)}e_T^{(k)}e_T^{(k)\prime}) = 0$, for all j and k, (see Lemma A15), the above relationship of $tr\left[J_T^{(j)}E(\Delta y_i^*\Delta y_i^{*\prime})\right]$ becomes

$$
\begin{aligned}
& tr\left[J_T^{(j)}E(\Delta y_i^*\Delta y_i^{*\prime})\right] \\
= \; & tr(M_T^{(j)}e_T^{(j)}e_T^{(j)\prime})E(\beta_i^{(j)2}) + tr(M_T^{(j)}\tilde{e}_T^{(j)}\tilde{e}_T^{(j)\prime})E(\delta_i^{(j)2}) - \frac{tr(M_T^{(j)}\tilde{e}_T^{(j)}\tilde{e}_T^{(j)\prime})}{tr\left(L_T^{(j)}\tilde{e}_T^{(j)}\tilde{e}_T^{(j)\prime}\right)}tr(L_T^{(j)}\tilde{e}_T^{(j)}\tilde{e}_T^{(j)\prime})E(\delta_i^{(j)2}) \\
= \; & tr(M_T^{(j)}e_T^{(j)}e_T^{(j)\prime})E(\beta_i^{(j)2}).
\end{aligned}
$$

Dividing the last relationship with $tr(M_T^{(j)}e_T^{(j)}e_T^{(j)\prime})$ and applying Chebyshev's WLLN proves the first result of the lemma. Similarly, we can prove the second result of the lemma. To this end, also notice that $tr\left[L_T^{(j)}E(\Delta y_i^*\Delta y_i^{*\prime})\right] = tr\left[L_T^{(j)}\tilde{e}_T^{(j)}\tilde{e}_T^{(j)\prime}\right]E(\delta_i^{(j)2})$, since $tr(L_T^{(j)}\Gamma_{iT}) = 0$.

Lemma A20. For model (23), under Assumptions B, C and D the following result holds:

$$p\lim_N\left[\tfrac{1}{N}tr(\Xi^{(\lambda)}\hat{\Gamma}) - \tfrac{1}{N}tr(\Lambda' Q^{(\lambda)}\Gamma)\right] = 0.$$

Proof: Write $\Xi^{(\lambda)} = \Xi_T^{(\lambda)} \otimes I_N$, where $\Xi_T^{(\lambda)} = \Psi_T^{(\lambda)} - \sum_{j=1}^{2}\frac{tr(\Psi_T^{(\lambda)}e_T^{(j)}e_T^{(j)\prime})J_T^{(j)}}{tr(M_T^{(j)}e_T^{(j)}e_T^{(j)\prime})} - \sum_{j=1}^{2}\frac{tr(\Psi_T^{(\lambda)}\tilde{e}_T^{(j)}\tilde{e}_T^{(j)\prime})L_T^{(j)}}{tr(L_T^{(j)}\tilde{e}_T^{(j)}\tilde{e}_T^{(j)\prime})}$. Then, $\frac{1}{N}tr(\Xi^{(\lambda)}\hat{\Gamma})$ can be written as

$$\frac{1}{N}tr(\Xi^{(\lambda)}\hat{\Gamma}) = tr(\Xi_T^{(\lambda)}\hat{\Gamma}_T) = \frac{1}{N}\sum_{i=1}^{N}\Delta y_i^{*\prime}\Xi_T^{(\lambda)}\Delta y_i^*.$$

Using Lemma A19, it can be shown that $E(\Delta y_i^{*\prime}\Xi_T^{(\lambda)}\Delta y_i^*) = tr(\Lambda_T' Q_T^{(\lambda)}\Gamma_{iT})$. Then, the proof of the lemma follows along the lines of that for Lemma A16.

Proof of Theorem 6: It follows by applying analogous arguments to those for the proof of Theorem 4.

Section F (Lemmas and Theorem Proofs for Section 6): The following lemmas are needed for the proofs of Theorems 7 and 8.

Lemma A21. Under Assumption E, the inconsistency of the LS estimator of φ, $\hat{\varphi}^{(\lambda)}$, of model (1) under H_0: $\varphi = 1$ is given by

$$p\lim_N(\hat{\varphi}^{(\lambda)} - 1 - \tfrac{tr(\Lambda' Q^{(\lambda)} \Pi \Gamma^{\varepsilon} \Pi)}{tr(\Lambda' Q^{(\lambda)} \Lambda \Pi \Gamma^{\varepsilon} \Pi)}) = 0.$$

Proof: As in Lemma A3, write $\hat{\varphi}^{(\lambda)} - 1 = \frac{\frac{1}{N}u'\Lambda' Q^{(\lambda)} u}{\frac{1}{N}u'\Lambda' Q^{(\lambda)}\Lambda u}$. The mean of $\frac{1}{N}u'\Lambda' Q^{(\lambda)} u$ is given as

$$E(\frac{1}{N}u'\Lambda' Q^{(\lambda)} u) = \frac{1}{N}tr\left[\Lambda' Q^{(\lambda)} E(uu')\right] = \frac{1}{N}tr\left[\Lambda' Q^{(\lambda)} E(\Pi\varepsilon\varepsilon'\Pi')\right] = \frac{1}{N}tr\left[\Lambda' Q^{(\lambda)} \Pi\Gamma^{\varepsilon}\Pi'\right]$$

and its variance as $Var(\frac{1}{N}u'\Lambda' Q^{(\lambda)} u) = \frac{1}{N^2}Var(\varepsilon'\Pi'\Lambda' Q^{(\lambda)}\Pi\varepsilon)$. Quadratic formula (36) implies that $Var(\varepsilon'\Pi'\Lambda' Q^{(\lambda)}\Pi\varepsilon) = O(NT)$ and, hence, $\frac{1}{N^2}Var(\varepsilon'\Pi'\Lambda' Q^{(\lambda)}\Pi\varepsilon) = o(1)$. Given the above results, the lemma can be proved by Chebyshev's inequality, implying $P(\left|\frac{1}{N}u'\Lambda' Q^{(\lambda)} u - \frac{1}{N}tr\left[\Lambda' Q^{(\lambda)}\Pi\Gamma^{\varepsilon}\Pi'\right]\right| > \varepsilon) \leq \frac{Var(\frac{1}{N}u'\Lambda' Q^{(\lambda)} u)}{\varepsilon^2} \to 0$, and by noticing that $E(\frac{1}{N}u'\Lambda' Q^{(\lambda)}\Lambda u) = \frac{1}{N}tr\left[\Lambda' Q^{(\lambda)}\Lambda\Pi\Gamma^{\varepsilon}\Pi'\right]$. The last result follows by applying analogous arguments to the above deriving the analytic formula of $E(\frac{1}{N}u'\Lambda' Q^{(\lambda)} u)$.

Lemma A22. For model (1), under Assumption E the following result holds:

$$p\lim_N\left[\tfrac{1}{N}tr(\Psi^{(\lambda)}\hat{\Gamma}) - \tfrac{1}{N}tr(\Lambda' Q^{(\lambda)}\Pi\Gamma^{\varepsilon}\Pi)\right] = 0.$$

Proof: First, note that $\frac{1}{N}tr(\Psi^{(\lambda)}\hat{\Gamma}) = tr(\Psi_T^{(\lambda)}\hat{\Gamma}_T)$, since $\hat{\Gamma} = \hat{\Gamma}_T \otimes I_N$. Based on Remark A4, $tr(\Psi^{(\lambda)}\hat{\Gamma})$ can be written as $tr(\Psi_T^{(\lambda)}\hat{\Gamma}_T) = tr(\Psi^{(\lambda)}\frac{1}{N}\Delta y\Delta y')$. Under H_0: $\varphi = 1$, we have $\Delta y\Delta y' = uu'$ and $u = \Pi\varepsilon$. Thus, $E\left[tr(\Psi^{(\lambda)}\frac{1}{N}\Delta y\Delta y')\right]$ can be written as follows:

$$\begin{aligned} E\left[tr(\Psi^{(\lambda)}\frac{1}{N}\Delta y\Delta y')\right] &= tr\left[\Psi^{(\lambda)}\frac{1}{N}E\left(\Delta y\Delta y'\right)\right] = tr\left[\Psi^{(\lambda)}\frac{1}{N}E\left(\Delta y\Delta y'\right)\right] \\ &= \frac{1}{N}tr(\Psi^{(\lambda)}E\left(\Pi\varepsilon\varepsilon'\Pi'\right)) = \frac{1}{N}tr(\Psi^{(\lambda)}\Pi\Gamma^{\varepsilon}\Pi'). \end{aligned}$$

By the definition of matrix $\Psi^{(\lambda)}$, we have $\frac{1}{N}tr(\Psi^{(\lambda)}\Pi\Gamma^{\varepsilon}\Pi') = \frac{1}{N}tr(\Lambda' Q^{(\lambda)}\Lambda\Pi\Gamma^{\varepsilon}\Pi)$. Note that the last relationship holds, even if $\Psi^{(\lambda)}$ has more non-zero diagonals. As in Lemma A21, it can be shown that

$$Var(tr(\Psi^{(\lambda)}\frac{1}{N}\Delta y\Delta y')) = \frac{1}{N^2}Var(\Delta y'\Psi^{(\lambda)}\Delta y) = \frac{1}{N^2}Var(\varepsilon'\Pi'\Psi^{(\lambda)}\Pi\varepsilon) = o(1)$$

Given the above results on $E\left[tr(\Psi^{(\lambda)}\frac{1}{N}\Delta y\Delta y')\right]$ and $Var(tr(\Psi^{(\lambda)}\frac{1}{N}\Delta y\Delta y'))$, the lemma can be proved by applying Chebyshev's inequality.

Proof of Theorem 8: First note that test statistic $Z^{(\lambda)}$ for model (1), with u_{it} defined by (27)-(28), can be written as follows:

$$
\begin{aligned}
& \sqrt{N}\hat{d}^{(\lambda)}\left(\hat{\varphi}^{(\lambda)}-1-\frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}\right) \\
= \; & \sqrt{N}\left(\frac{1}{N}y_{-1}'Q^{(\lambda)}y_{-1}\right)\left(\frac{\frac{1}{N}y_{-1}'Q^{(\lambda)}y}{\frac{1}{N}y_{-1}'Q^{(\lambda)}y_{-1}}-1-\frac{\frac{1}{N}\Delta y'\Psi^{(\lambda)}\Delta y}{\frac{1}{N}y_{-1}'Q^{(\lambda)}y_{-1}}\right) \text{ (see Remark A4)} \\
= \; & \sqrt{N}\left(\frac{1}{N}y_{-1}'Q^{(\lambda)}y_{-1}\right)\left(\frac{\frac{1}{N}y_{-1}'Q^{(\lambda)}y}{\frac{1}{N}y_{-1}'Q^{(\lambda)}y_{-1}}-1-\frac{\frac{1}{N}\Delta y'\Psi^{(\lambda)}\Delta y}{\frac{1}{N}y_{-1}'Q^{(\lambda)}y_{-1}}\right) \\
= \; & \frac{1}{\sqrt{N}}\left(u'\Lambda'Q^{(\lambda)}u-u'\Psi^{(\lambda)}u\right)=\frac{1}{\sqrt{N}}u'\left(\Lambda'Q^{(\lambda)}-\Psi^{(\lambda)}\right)u=\frac{1}{\sqrt{N}}\varepsilon'\Pi'\left(\Lambda'Q^{(\lambda)}-\Psi^{(\lambda)}\right)\Pi\varepsilon,
\end{aligned}
$$

where matrix $\Pi'\left(\Lambda'Q^{(\lambda)}-\Psi^{(\lambda)}\right)\Pi$ has zeroes in its diagonals,

$$
\begin{aligned}
E(\varepsilon'\Pi'\left(\Lambda'Q^{(\lambda)}-\Psi^{(\lambda)}\right)\Pi\varepsilon) &= tr(\Pi'\left(\Lambda'Q^{(\lambda)}-\Psi^{(\lambda)}\right)\Pi\Gamma^{\varepsilon})=0 \\
\text{and} \quad Var(\varepsilon'\Pi'\left(\Lambda'Q^{(\lambda)}-\Psi^{(\lambda)}\right)\Pi\varepsilon) &= 2tr(F^{(\lambda)}\Gamma^{\varepsilon}F^{(\lambda)}\Gamma^{\varepsilon}),
\end{aligned}
\tag{49}
$$

with $F^{(\lambda)}=\frac{1}{2}\left(\Pi'\left(\Lambda'Q^{(\lambda)}-\Psi^{(\lambda)}\right)\Pi+\Pi'\left(Q^{(\lambda)}\Lambda-\Psi^{(\lambda)\prime}\right)\Pi\right)$ by Remark A2. By Theorem A1 of Kelejian and Prucha (2010), it can be shown that

$$
V^{(\lambda)-1/2}\left[\frac{1}{\sqrt{N}}\varepsilon'\Pi'\left(\Lambda'Q^{(\lambda)}-\Psi^{(\lambda)}\right)\Pi\varepsilon\right]\xrightarrow{d}N\left(0,1\right),
$$

where $V^{(\lambda)-1/2}=2tr(F^{(\lambda)}\Gamma^{\varepsilon}F^{(\lambda)}\Gamma^{\varepsilon})$. Note that Assumption A.2 of Kelejian and Prucha (2010) holds in the case of Theorem 8 by our Assumption E and the Remark A2 of Kapoor et al. (2007).

Proof of Theorem 9: The proof of this theorem is based on the theorem of Beran and Ducharme (1991), as presented in Horowitz (2001). In particular, the case of our test is similar to that of the Example 2.1 of Horowitz (2001). To prove the theorem, consider the following (unscaled and nonstandardized) version of our test statistic: $\hat{d}^{(\lambda)}\left(\hat{\varphi}^{(\lambda)}-1-\frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}\right)$ under Assumption E. Let $G_N(\pi,F_0)$ be $G_N(\pi,F_0)=P_N\left[\hat{d}^{(\lambda)}\left(\hat{\varphi}^{(\lambda)}-1-\frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}\right)\right]$ and F_N be the empirical distribution function of the data. Then, the bootstrap for this distribution is given as

$$
\sqrt{N}\left[\hat{d}^{B(\lambda)}\left(\hat{\varphi}^{B(\lambda)}-1-\frac{\hat{b}^{B(\lambda)}}{\hat{d}^{B(\lambda)}}\right)-\hat{d}^{(\lambda)}\left(\hat{\varphi}^{(\lambda)}-1-\frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}\right)\right]
$$

This bootstrap is centred around $\hat{d}^{(\lambda)}\left(\hat{\varphi}^{(\lambda)}-1-\frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}\right)$, since the latter is the mean of the distribution from which the bootstrap sample is drawn. Also notice that $E\left[\hat{d}^{(\lambda)}\left(\hat{\varphi}^{(\lambda)}-1-\frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}\right)\right]=0$ by (49) and that statistic $\sqrt{N}\left[\hat{d}^{(\lambda)}\left(\hat{\varphi}^{(\lambda)}-1-\frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}\right)-E\left[\hat{d}^{(\lambda)}\left(\hat{\varphi}^{(\lambda)}-1-\frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}\right)\right]\right]=Z^{(\lambda)}$. The bootstrap estimator is $G_N(\pi,F_N)=P_N^B\left[\hat{d}^{B(\lambda)}\left(\hat{\varphi}^{B(\lambda)}-1-\frac{\hat{b}^{B(\lambda)}}{\hat{d}^{B(\lambda)}}\right)-\hat{d}^{(\lambda)}\left(\hat{\varphi}^{(\lambda)}-1-\frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}\right)\leq\pi\right]$, where P_N^B is the probability distribution induced by the sampling process. Based on the arguments of Horowitz (2001), it can be seen that the conditions of the theorem of Beran and Ducharme (1991) hold in our case, and thus the bootstrap is consistent. For the case that the date of the break is unknown, the bootstrap of minimum statistic is also consistent. This can be proved by similar arguments.

Table 1: Size and power of $\min_{\lambda \in I} Z^{(\lambda)}$ in the case of individual intercepts

		N	50	50	50	100	100	100	200	200	200
θ	λ/T	φ\T	8	10	15	8	10	15	8	10	15
-0.5	0.25	1	0.034	0.036	0.0465	0.047	0.045	0.0485	0.0365	0.0555	0.0555
		0.95	0.044	0.049	0.042	0.0635	0.068	0.054	0.071	0.073	0.064
		0.9	0.05	0.048	0.046	0.079	0.065	0.057	0.097	0.010	0.071
	0.5	1	0.041	0.039	0.040	0.033	0.051	0.044	0.043	0.042	0.057
		0.95	0.0435	0.0415	0.041	0.0625	0.07	0.0415	0.081	0.0775	0.0635
		0.9	0.056	0.047	0.0475	0.075	0.0695	0.0415	0.1065	0.072	0.0635
	0.75	1	0.033	0.036	0.038	0.0505	0.039	0.0535	0.055	0.0595	0.053
		0.95	0.048	0.053	0.0485	0.056	0.0645	0.0415	0.0755	0.0665	0.05
		0.9	0.0575	0.0485	0.0455	0.0815	0.0705	0.063	0.097	0.084	0.0645
0	0.25	1	0.0665	0.058	0.049	0.061	0.067	0.063	0.067	0.0575	0.0555
		0.95	0.166	0.1935	0.15	0.264	0.2725	0.209	0.424	0.428	0.343
		0.9	0.33	0.3	0.234	0.5645	0.4935	0.3565	0.8315	0.7435	0.607
	0.5	1	0.0575	0.0585	0.056	0.058	0.055	0.0635	0.0595	0.0525	0.054
		0.95	0.182	0.183	0.1335	0.277	0.2775	0.225	0.418	0.4335	0.3325
		0.9	0.337	0.302	0.2155	0.5435	0.5235	0.356	0.819	0.779	0.57
	0.75	1	0.0585	0.057	0.058	0.057	0.0655	0.059	0.06	0.0515	0.047
		0.95	0.1545	0.1875	0.148	0.279	0.282	0.2265	0.416	0.4385	0.3585
		0.9	0.335	0.327	0.2415	0.558	0.4975	0.364	0.8075	0.748	0.5865
0.5	0.25	1	0.0285	0.0295	0.0345	0.0375	0.039	0.0395	0.042	0.0355	0.0495
		0.95	0.19	0.174	0.1415	0.3225	0.3005	0.259	0.601	0.555	0.423
		0.9	0.4205	0.385	0.25	0.743	0.6725	0.454	0.974	0.9365	0.737
	0.5	1	0.0315	0.029	0.0435	0.044	0.0385	0.042	0.042	0.0445	0.0465
		0.95	0.1675	0.1775	0.151	0.345	0.326	0.2485	0.6035	0.5765	0.414
		0.9	0.411	0.3605	0.2445	0.7555	0.67	0.438	0.972	0.939	0.7285
	0.75	1	0..031	0.0355	0.0385	0.0375	0.039	0.0475	0.048	0.044	0.0505
		0.95	0.182	0.1745	0.137	0.324	0.328	0.242	0.586	0.571	0.413
		0.9	0.404	0.373	0.243	0.742	0.6815	0.446	0.9785	0.942	0.7375

Table 2: Size and power of $\min_{\lambda \in I} Z^{(\lambda)}$ in the case of incidental trends

		N	50	50	50	100	100	100	200	200	200
θ	λ/T	φ\T	8	10	15	8	10	15	8	10	15
-0.5	0.25	1	0.0295	0.033	0.041	0.0375	0.0445	0.0495	0.0475	0.0525	0.056
		0.95	0.044	0.0475	0.0665	0.0465	0.0525	0.0895	0.0615	0.0685	0.1275
		0.9	0.0385	0.0585	0.0985	0.059	0.0935	0.132	0.0735	0.1235	0.2315
	0.5	1	0.034	0.0345	0.051	0.044	0.038	0.049	0.039	0.0445	0.05
		0.95	0.0345	0.048	0.0575	0.047	0.064	0.0885	0.058	0.0765	0.1025
		0.9	0.04	0.0565	0.083	0.0485	0.0775	0.125	0.0695	0.118	0.1815
	0.75	1	0.0315	0.0385	0.041	0.048	0.046	0.054	0.042	0.045	0.056
		0.95	0.039	0.0495	0.0675	0.0385	0.0535	0.0745	0.0545	0.0685	0.119
		0.9	0.04	0.049	0.0725	0.052	0.0795	0.109	0.078	0.1175	0.1645
0	0.25	1	0.058	0.064	0.0785	0.061	0.0595	0.077	0.0455	0.0615	0.0715
		0.95	0.0635	0.0715	0.107	0.065	0.0805	0.098	0.074	0.0705	0.123
		0.9	0.0785	0.11	0.192	0.11	0.144	0.25	0.116	0.1725	0.361
	0.5	1	0.048	0.0625	0.078	0.052	0.0645	0.072	0.059	0.059	0.074
		0.95	0.066	0.0855	0.0895	0.072	0.069	0.1115	0.0655	0.081	0.109
		0.9	0.086	0.109	0.18	0.0865	0.1215	0.2365	0.1325	0.163	0.3225
	0.75	1	0.059	0.059	0.078	0.054	0.0585	0.0705	0.069	0.06	0.0675
		0.95	0.055	0.061	0.103	0.06	0.0755	0.111	0.0665	0.0825	0.107
		0.9	0.085	0.1145	0.1675	0.0855	0.1305	0.216	0.126	0.171	0.3145
0.5	0.25	1	0.0405	0.0495	0.0645	0.0545	0.0485	0.073	0.046	0.0455	0.054
		0.95	0.043	0.047	0.073	0.052	0.0565	0.0905	0.047	0.0705	0.095
		0.9	0.06	0.072	0.144	0.0675	0.0985	0.204	0.0815	0.1185	0.3205
	0.5	1	0.0485	0.0495	0.05	0.056	0.048	0.06	0.047	0.0505	0.053
		0.95	0.038	0.045	0.067	0.0425	0.0465	0.0745	0.0575	0.072	0.0995
		0.9	0.05	0.0565	0.1515	0.0685	0.0915	0.2005	0.0765	0.1345	0.3075
	0.75	1	0.048	0.05	0.0485	0.0595	0.048	0.0595	0.0505	0.0435	0.06
		0.95	0.048	0.0565	0.071	0.054	0.0685	0.0865	0.0565	0.0685	0.112
		0.9	0.0575	0.077	0.139	0.0685	0.097	0.1945	0.089	0.1415	0.306

Table 3: Size and power of $\min_{\lambda \in I} Z^{(\lambda)}$ in the case of individual intercepts and SMA errors

		N	50	50	50	100	100	100	200	200	200
μ	λ/T	$\varphi\backslash$T	8	10	15	8	10	15	8	10	15
0.4	0.25	1	0.086	0.063	0.065	0.065	0.069	0.079	0.065	0.063	0.062
		0.95	0.137	0.152	0.148	0.212	0.203	0.144	0.287	0.292	0.209
		0.9	0.246	0.208	0.172	0.343	0.308	0.228	0.557	0.493	0.35
	0.5	1	0.075	0.065	0.083	0.064	0.048	0.061	0.059	0.064	0.053
		0.95	0.139	0.156	0.123	0.202	0.185	0.178	0.281	0.267	0.221
		0.9	0.229	0.24	0.175	0.344	0.311	0.222	0.557	0.531	0.349
	0.75	1	0.07	0.065	0.085	0.073	0.075	0.061	0.051	0.059	0.071
		0.95	0.18	0.149	0.145	0.208	0.196	0.165	0.261	0.268	0.221
		0.9	0.222	0.25	0.164	0.361	0.309	0.245	0.549	0.501	0.364
0.8	0.25	1	0.069	0.071	0.088	0.063	0.091	0.081	0.064	0.07	0.072
		0.95	0.178	0.15	0.136	0.19	0.208	0.152	0.243	0.269	0.216
		0.9	0.233	0.208	0.159	0.28	0.281	0.221	0.446	0.408	0.356
	0.5	1	0.058	0.083	0.064	0.068	0.064	0.079	0.062	0.061	0.085
		0.95	0.158	0.159	0.147	0.18	0.163	0.181	0.252	0.263	0.226
		0.9	0.227	0.205	0.184	0.296	0.276	0.239	0.467	0.438	0.323
	0.75	1	0.073	0.075	0.064	0.068	0.07	0.081	0.065	0.073	0.067
		0.95	0.159	0.154	0.146	0.188	0.203	0.179	0.269	0.279	0.214
		0.9	0.234	0.201	0.196	0.309	0.297	0.236	0.482	0.445	0.314

Table 4: Size and power of $\min Z^{(\lambda)}$ in the case of individual intercepts and SAR errors

		N	50	50	50	100	100	100	200	200	200
μ	λ/T	$\varphi\backslash$T	8	10	15	8	10	15	8	10	15
0.4	0.25	1	0.082	0.068	0.073	0.063	0.072	0.059	0.072	0.054	0.072
		0.95	0.163	0.143	0.138	0.181	0.187	0.177	0.29	0.255	0.208
		0.9	0.221	0.229	0.16	0.331	0.314	0.216	0.516	0.485	0.361
	0.5	1	0.054	0.059	0.073	0.058	0.046	0.073	0.064	0.054	0.071
		0.95	0.161	0.172	0.115	0.195	0.197	0.153	0.304	0.3	0.215
		0.9	0.221	0.229	0.176	0.333	0.291	0.232	0.56	0.498	0.35
	0.75	1	0.07	0.075	0.072	0.058	0.056	0.056	0.064	0.071	0.065
		0.95	0.154	0.143	0.146	0.18	0.217	0.153	0.267	0.263	0.212
		0.9	0.227	0.224	0.176	0.33	0.314	0.252	0.549	0.497	0.358
0.8	0.25	1	0.129	0.124	0.13	0.106	0.133	0.124	0.121	0.129	0.137
		0.95	0.192	0.185	0.216	0.204	0.215	0.231	0.261	0.281	0.27
		0.9	0.214	0.222	0.216	0.258	0.28	0.25	0.383	0.373	0.352
	0.5	1	0.12	0.123	0.146	0.1	0.14	0.138	0.125	0.146	0.146
		0.95	0.165	0.197	0.201	0.208	0.222	0.2	0.254	0.297	0.273
		0.9	0.229	0.217	0.219	0.274	0.278	0.273	0.407	0.409	0.369
	0.75	1	0.113	0.133	0.148	0.115	0.12	0.133	0.112	0.145	0.137
		0.95	0.196	0.176	0.191	0.207	0.221	0.217	0.267	0.294	0.265
		0.9	0.234	0.255	0.227	0.288	0.299	0.246	0.377	0.382	0.314

Table 5: Size and power of $\min_{\lambda \in I} Z^{(\lambda)}$ in the case of incidental trends and SMA errors

		N	50	50	50	100	100	100	200	200	200
μ	λ/T	$\varphi\backslash$T	8	10	15	8	10	15	8	10	15
0.4	0.25	1	0.074	0.067	0.087	0.05	0.069	0.064	0.058	0.058	0.065
		0.95	0.067	0.074	0.113	0.058	0.078	0.105	0.074	0.079	0.116
		0.9	0.077	0.101	0.161	0.078	0.103	0.19	0.09	0.149	0.274
	0.5	1	0.07	0.063	0.075	0.076	0.074	0.07	0.061	0.062	0.065
		0.95	0.083	0.076	0.102	0.053	0.072	0.105	0.064	0.075	0.119
		0.9	0.088	0.093	0.148	0.069	0.111	0.176	0.082	0.127	0.233
	0.75	1	0.06	0.064	0.07	0.062	0.067	0.057	0.056	0.069	0.06
		0.95	0.069	0.09	0.088	0.078	0.073	0.114	0.062	0.066	0.11
		0.9	0.074	0.092	0.134	0.089	0.098	0.185	0.115	0.132	0.268
0.8	0.25	1	0.09	0.076	0.103	0.068	0.071	0.070	0.065	0.075	0.072
		0.95	0.094	0.089	0.109	0.073	0.09	0.125	0.083	0.094	0.113
		0.9	0.088	0.102	0.153	0.094	0.124	0.192	0.118	0.121	0.276
	0.5	1	0.076	0.095	0.094	0.078	0.07	0.086	0.072	0.077	0.064
		0.95	0.078	0.091	0.114	0.071	0.077	0.106	0.073	0.086	0.136
		0.9	0.108	0.097	0.151	0.102	0.118	0.169	0.11	0.143	0.272
	0.75	1	0.082	0.092	0.087	0.074	0.082	0.092	0.064	0.08	0.084
		0.95	0.053	0.076	0.096	0.087	0.072	0.123	0.071	0.072	0.125
		0.9	0.086	0.099	0.152	0.086	0.127	0.193	0.106	0.142	0.261

Table 6: Size and power of $\min_{\lambda \in I} Z^{(\lambda)}$ in the case of incidental trends and SAR errors

		N	50	50	50	100	100	100	200	200	200
μ	λ/T	$\varphi\backslash$T	8	10	15	8	10	15	8	10	15
0.4	0.25	1	0.071	0.071	0.085	0.076	0.076	0.08	0.058	0.067	0.073
		0.95	0.059	0.086	0.099	0.087	0.093	0.107	0.073	0.093	0.109
		0.9	0.082	0.104	0.156	0.079	0.106	0.188	0.097	0.142	0.272
	0.5	1	0.053	0.063	0.06	0.068	0.074	0.076	0.059	0.064	0.061
		0.95	0.082	0.083	0.085	0.066	0.091	0.098	0.067	0.101	0.133
		0.9	0.093	0.119	0.136	0.091	0.121	0.192	0.095	0.112	0.293
	0.75	1	0.073	0.082	0.075	0.066	0.06	0.071	0.066	0.063	0.051
		0.95	0.086	0.063	0.093	0.081	0.064	0.107	0.062	0.084	0.12
		0.9	0.086	0.081	0.129	0.069	0.091	0.184	0.081	0.152	0.268
0.8	0.25	1	0.129	0.133	0.153	0.116	0.124	0.145	0.127	0.135	0.174
		0.95	0.134	0.143	0.175	0.128	0.139	0.183	0.126	0.148	0.177
		0.9	0.136	0.167	0.235	0.137	0.157	0.259	0.145	0.2	0.334
	0.5	1	0.132	0.134	0.152	0.131	0.144	0.16	0.118	0.153	0.162
		0.95	0.12	0.129	0.195	0.127	0.14	0.193	0.135	0.148	0.212
		0.9	0.14	0.155	0.211	0.112	0.168	0.263	0.155	0.183	0.321
	0.75	1	0.128	0.14	0.141	0.121	0.133	0.17	0.125	0.125	0.162
		0.95	0.112	0.136	0.17	0.129	0.139	0.198	0.13	0.158	0.218
		0.9	0.141	0.14	0.226	0.133	0.142	0.243	0.149	0.179	0.319

Table 7: Size and power of $\min_{\lambda \in I} Z^{(\lambda_1,\lambda_2)}$ with two structural breaks.

		N	50	50	50	100	100	100	200	200	200
Model	$\lambda_1/T, \lambda_2/T$	φ\T	8	10	15	8	10	15	8	10	15
Intercepts	$0.25, 0.5$	1	0.065	0.074	0.066	0.064	0.065	0.054	0.066	0.056	0.054
		0.95	0.2105	0.212	0.1955	0.271	0.287	0.244	0.444	0.4795	0.411
		0.9	0.399	0.3685	0.318	0.613	0.596	0.5125	0.856	0.8425	0.7945
	$0.5, 0.75$	1	0.0665	0.0595	0.0605	0.063	0.0625	0.0615	0.059	0.052	0.059
		0.95	0.2025	0.1935	0.1855	0.2915	0.286	0.265	0.45	0.465	0.409
		0.9	0.424	0.3745	0.3135	0.6025	0.593	0.5125	0.872	0.869	0.799
	$0.25, 0.75$	1	0.0625	0.058	0.0635	0.058	0.0555	0.0525	0.056	0.055	0.055
		0.95	0.205	0.212	0.1635	0.273	0.28	0.2595	0.449	0.45	0.415
		0.9	0.399	0.391	0.3385	0.592	0.583	0.5125	0.856	0.8455	0.8145
Trends	$0.25, 0.5$	1	0.057	0.0615	0.0605	0.0495	0.0505	0.05	0.0485	0.0645	0.0465
		0.95	0.071	0.0655	0.074	0.057	0.0685	0.0795	0.06	0.064	0.089
		0.9	0.076	0.1135	0.158	0.0705	0.1115	0.2135	0.0915	0.147	0.343
	$0.5, 0.75$	1	0.0605	0.0615	0.0565	0.062	0.049	0.056	0.056	0.0615	0.048
		0.95	0.0585	0.071	0.083	0.0575	0.056	0.08	0.0565	0.0645	0.0865
		0.9	0.0835	0.099	0.1705	0.084	0.1265	0.198	0.0815	0.1295	0.336
	$0.25, 0.75$	1	0.0645	0.067	0.0585	0.0525	0.059	0.061	0.0445	0.0495	0.0485
		0.95	0.0625	0.073	0.0905	0.05	0.061	0.0845	0.057	0.067	0.0995
		0.9	0.081	0.094	0.162	0.0715	0.129	0.2105	0.0985	0.146	0.333