Generalized fixed-T panel unit root tests

Karavias, Yiannis; Tzavalis, Elias

DOI:
10.1111/sjos.12392

Citation for published version (Harvard):
Karavias, Y & Tzavalis, E 2019, 'Generalized fixed-T panel unit root tests', *Scandinavian Journal of Statistics*, vol. 46, no. 4, pp. 1227-1251. https://doi.org/10.1111/sjos.12392

Generalized Fixed-T Panel Unit Root Tests

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

October 2018

Abstract

Panel data unit root tests which can be applied to data that do not have many time series observations are based on very restrictive error and deterministic component specification assumptions. In this paper we develop a new, doubly modified estimator, based on which we propose a panel unit root test that allows for multiple structural breaks, linear and non-linear trends, heteroscedasticity, serial correlation and error cross section heterogeneity, when the number of time series observations is finite. The test has the additional perk that it is invariant to the initial condition.

JEL classification: C22, C23

Keywords: Panel data; Unit Root; Fixed T; Nonlinear Trends; Structural Breaks; Serial Correlation

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, Joakim Westerlund, Michael Massmann, the participants of the seminar series in Lund University, Queen's University Belfast and in University of Ioannina, the participants of the 2nd IAEE Applied Econometrics Conference in Thessaloniki 2015; in the Workshop on Macroeconomic Flunctuations, Large Shocks and Breaks, Athens 2015; and in the XXXIX Simposio de la Asociación Española de Economía-Spanish Economic Association (SAEe), Palma de Mallorca, 2014, for helpful comments and suggestions.

$*$:Corresponding author

a :Department of Economics, University of Birmingham. E-mail: i.karavias@bham.ac.uk.

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

1 Introduction

Economic theory frequently predicts the existence of equilibria in which the involved processes converge, in the long run, to a constant value or maintain a constant growth rate. The existence of these predicted equilibria, and equivalently whether their originating theories hold, hinges on the presence of unit roots which can be tested by panel data unit root tests. A review of applications of panel unit root tests can be found in Choi (2015).

In this paper, we propose a general panel data unit root test for models with both general trend functions and general error processes, when the number of time series observations (T) is finite. The new test allows for the existence of multiple structural breaks, nonlinear trends, unspecified forms of short term serial correlation, heteroscedasticity and error cross section heterogeneity. We develop a novel, doubly modified estimator which is consistent under the null hypothesis of nonstationarity and as the number of cross section units N goes to infinity. Our new estimator is a corrected version of the within groups (WG) estimator which is known to be inconsistent in a finite T framework, see e.g. Nickel (1981). We first show that, when the series are nonstationary, the inconsistency of the WG estimator depends on the error variance-covariance matrix. This inconsistency is estimated by a nonparametric estimator and this is the first correction. However, the nonparametric estimator is inconsistent in the presence of complex deterministic components and must be modified as well and hence, the final corrected estimator used in the test statistic is referred to as the "doubly modified estimator", denoted with the acronym DME. The first correction is similar in spirit to the fully modified least squares estimator of Phillips and Hansen (1990), for single time series analysis, employed to correct for endogeneity bias and serial correlation. The second correction, which is that of the nonparametric estimator, is based on the idea of the variance-covariance matrix estimation method suggested by Abowd and Card (1989) and Arellano (1990).

The new test has a number of theoretical virtues which result in more robust inference. First, the problem of modelling the initial conditions, which is crucial when T is small (see e.g. Bun and Sarafidis (2015)), is avoided because the DME is invariant to the initial conditions. Second, two hypotheses are considered: a) breaks appear under the null and the alternative, as in Perron (1989) and b) breaks appear only under the alternative as in Zivot and Andrews (1992). In this latter case, when the dates of the breaks are unknown, the pdf of the limiting distribution of an infimum-type test statistic is derived analytically. This is the first paper which provides an analytic pdf of the distribution of an infimum of statistics, since Davies (1977). Third, by using information from the cross section dimension we avoid the problem of estimating the long run variance, which is a difficult econometric task even in the panel data setting (Moon and Perron (2004)).

Most panel unit root tests aim at macroeconomic data (i.e. country level data) where both N and T are large (see, e.g. Pesaran et al. (2013) and Bai and Carrion-i-Silvestre (2009)). In principle however, testing whether series are stationary or not is a way to characterize dynamic behaviour and such behaviour also appears in disaggregated data where there is a large number of units observed over only a short period of time, i.e. microeconomic panel data. The first panel data unit root test was applied on disaggregated wage data with a small time dimension (see Breitung and Meyer (1994)). Other finite T panel unit root tests have been proposed by Harris and Tzavalis (1999), Kruiniger (2008), De Blander and Dhaene (2012), Han and Phillips (2012), Karavias and Tzavalis (2014a), Choi (2014), Karavias and Tzavalis (2016) and Robertson et al. (2017) inter alia. Testing the unit root hypothesis in this type of data has attracted considerable attention, but existing tests rely on restrictive assumptions on the model specification. This happens because, since T is considered finite, cross section information must be exploited to bypass time series econometric problems such as nonstationarity, structural change and serial correlation in the errors, and it is not always clear how this can be done. Papers in the literature that consider fixed-T panel unit root testing with structural breaks are those of Hadri et al. (2012), Karavias and Tzavalis (2014b) and Karavias and Tzavalis (2017). The first two papers allow for a single structural break in the intercept and the linear trend of the model while Karavias and Tzavalis (2017) derive the local power functions of tests with one known structural break. In this paper we no longer specify the form of the trend function, i.e. intercepts and/or trends or what is the number and date of structural breaks, rather we provide a general theory of testing and sufficient conditions for the existence of the DME. Monte Carlo simulations demonstrate the excellent finite sample properties of the new test.

The paper is organized as follows. Section 2 sets up the general model and the hypotheses of interest. Section 3 introduces the DME and the test statistic. Section 4 discusses the assumptions used and Section 5 provides the asymptotic theory. Section 7 provides simulation results of Monte Carlo experiments and Section 8 concludes the paper. The proofs of the theorems are relegated to the Appendix. Matlab codes that compute the estimator and the test are available on the first author's web page.

A few words on notation. The elements of a matrix A are denoted by $[A]_{i,j}$ and $[A]_j$ denotes the *j-th* column. Define $[\cdot]_p^+ : R^{T\times T} \longrightarrow R^{T\times T}$ to be the family of linear transformations which map the main, the first p upper and first p lower diagonals of a $T \times T$ matrix to themselves while the rest of the diagonals are set to zero, where $p \in \{0, ..., T-1\}$. Additionally, $[\cdot]_p^- : R^{T\times T} \longrightarrow R^{T\times T}$ sets the main, the first p upper and first p lower diagonals equal to zero and the rest are set equal to themselves. Examples of $[\cdot]_p^+$ and $[\cdot]_p^-$ can be found in the Appendix. Let vec denote the vectorization operator and $\text{vec}_{T,T}^{-1}$ an inverse vectorization operator such that $\text{vec}_{T,T}^{-1}(\cdot) : R^{T^2} \longrightarrow R^{T\times T}$. Let vech denote the half vectorization operator. We use the Euclidean norm denoted as $\|u\| = \sqrt{u'u}$, where u is a column vector. For square matrices A, B, we write $A > B$ to indicate that $A - B$ is positive definite. For matrices A and B with the same number of rows, let $(A|B)$ denote the augmented matrix (A and B are horizontally concatenated). The column space of a matrix A is symbolized as $R\{A\}$. The notations $p\lim_N$ and $\xrightarrow{p}$ denote convergence in probability while $\xrightarrow{d}$ denotes convergence in distribution. Finally, $\lfloor \cdot \rfloor$ is the greatest integer function.

2 The model

Consider the following model:

$$\begin{aligned} y_i &= X^{(\lambda)}\pi_i^{(\lambda)} + \zeta_i, \quad \text{for } i = 1, ..., N, \\ \zeta_i &= \varphi\zeta_{i,-1} + u_i, \end{aligned} \tag{1}$$

where $y_i = (y_{i1}, ..., y_{iT})'$ is a vector of observations, $u_i = (u_{i1}, ..., u_{iT})'$, $\zeta_i = (\zeta_{i1}, ..., \zeta_{iT})'$ and $\zeta_{i,-1} = (\zeta_{i0}, ..., \zeta_{iT-1})'$ are $T \times 1$ vectors of errors, $X^{(\lambda)}$ is a $T \times k$ matrix of deterministic components and $\pi_i^{(\lambda)}$ a $k \times 1$ vector of random individual effects. The term $X^{(\lambda)}\pi_i^{(\lambda)}$ is referred to as the trend function. The scalar parameter φ is the autoregressive coefficient and $\varphi \in (-1, 1]$. If m common breaks occur, then these breaks define a partition $(T_1, ..., T_m)$ denoted by λ, where the T_j, for $j \in \{1, ..., m\}$, are the dates of the breaks. We also define $T_0 = 1$ and $T_{m+1} = T$ and denote with I the set of all admissible partitions.

The above specification nests various models which are frequently used, i.e. if there are no structural breaks $X^{(\lambda)} = e$ and $\pi_i^{(\lambda)} = \pi_i$, where e is a $T \times 1$ vector of ones and π_i is a scalar individual effect. Another popular specification is a model with individual intercepts and individual trends in which $X^{(\lambda)} = [e, \tau]$ and $\pi_i^{(\lambda)} = (\pi_{1i}, \pi_{2i})'$, where $\tau_t = t$ for $t = 1, ..., T$ and π_{1i}, π_{2i} are scalar individual effects. In models without breaks the columns of $X^{(\lambda)}$ will consist of one intercept vector and ρ trend vectors with $1 + \rho = k$, where ρ is the degree of the trend polynomial. If there are m breaks, $X^{(\lambda)}$ will contain $(m+1)$ intercept vectors and $(m+1)\rho$ trend vectors as columns, with $(m+1) + (m+1)\rho = k$. For example, in the model with one structural break in the intercept, $\rho = 0$ and $X^{(\lambda)} \equiv \left[e^{(1)}, e^{(2)}\right]$, where $e_T^{(1)}$ and $e_T^{(2)}$ are $T \times 1$ vectors whose elements are defined as follows: $e_t^{(1)} = 1$ if $t \leq T_1$ and 0 otherwise, $e_t^{(2)} = 1$ if $t > T_1$ and 0 otherwise. Also, $\pi_i^{(\lambda)} \equiv (\pi_{1i}^{(1)}, \pi_{1i}^{(2)})'$ where scalar $\pi_{1i}^{(1)}$ is the individual effect before the break and $\pi_{1i}^{(2)}$ is the individual effect after the break. A single break in a model with individual linear trends ($\rho = 1$) can be cast as $X^{(\lambda)} = \left[e^{(1)}, e^{(2)}, \tau^{(1)}, \tau^{(2)}\right]$ where $\tau^{(1)}$ and $\tau^{(2)}$ are $T \times 1$ vectors: $\tau_t^{(1)} = t$ if $t \leq T_1$ and 0 otherwise, and $\tau_t^{(2)} = t$ if $t > T_1$ and 0 otherwise. Here $\pi_i^{(\lambda)} \equiv (\pi_{1i}^{(1)}, \pi_{1i}^{(2)}, \pi_{2i}^{(1)}, \pi_{2i}^{(2)})'$ where scalars $\pi_{1i}^{(1)}$ and $\pi_{1i}^{(2)}$ are the individual effects before and after the break, while $\pi_{2i}^{(1)}$ and $\pi_{2i}^{(2)}$ are the linear trend coefficients before and after the break. Similarly (1) can nest cases of pure or partial structural change, multiple structural breaks and higher order polynomial trends. We consider $X^{(\lambda)}$ known up to the location of the breaks.

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. The common break may also be seen as the mean of possibly random breaks but we do not consider the common break date to be random, see e.g. Bai (2010). The magnitude of the break, i.e. $\pi_{1i}^{(2)} - \pi_{1i}^{(1)}$, can be different across units i, thus allowing for each individual unit to respond idiosyncratically to the effects of a structural break.

We wish to test the null hypothesis that y_{it} is a unit root process against the alternative that it is (trend) stationary. In the presence of structural change there are two approaches in doing so. The first is by allowing for structural breaks under both the null hypothesis of unit root and the alternative of stationarity as in Perron (1989); we shall denote this hypothesis as H_1. Second, by testing the null hypothesis of unit root and no breaks against the alternative of stationarity with breaks, as in Zivot and Andrews (1992); we shall denote this hypothesis as H_2. These hypotheses can be written as:

$$H_{1,0} \quad : \quad \varphi = 1,$$
$$H_{1,1} \quad : \quad \varphi < 1,$$

and

$$H_{2,0} \quad : \quad \varphi = 1 \text{ and } \pi_{ji}^{(1)} = \pi_{ji}^{(2)} = ... = \pi_{ji}^{(m+1)}, \text{ for all } j \text{ and } i.$$
$$H_{2,1} \quad : \quad \varphi < 1.$$

In applied work both hypotheses are frequently tested for various deterministic specifications. A popular example is when $X^{(\lambda)}\pi_i^{(\lambda)} = \left[e^{(1)}, e^{(2)}, \tau^{(1)}, \tau^{(2)}\right](\pi_{1i}^{(1)}, \pi_{1i}^{(2)}, \pi_{2i}^{(1)}, \pi_{2i}^{(2)})'$ for which the null hypothesis $H_{1,0}$ postulates that y_{it} is a unit root process with drift and a one-time jump in the intercept at the time of the break while, under $H_{1,1}$, y_{it} is stationary with a change taking place in both the intercept and the trend. This is the "crash and changing growth" model of Perron (1989) and is used to capture changes in both the levels and the slopes of the series. Under assumption $H_{2,0}$ the null hypothesis would be that y_{it} is a random walk process with drift, since $\pi_{1i}^{(1)} = \pi_{1i}^{(2)}$ and $\pi_{2i}^{(1)} = \pi_{2i}^{(2)}$, while under the alternative $H_{2,1}$, it would be stationary with a change in both the level and the slope, as in $H_{1,1}$.

The two hypotheses require different testing procedures when the breaks are at unknown dates. In H_1 the breaks appear under both the null and the alternative hypotheses and they can be estimated in a first step as in Bai (2010), using the first differenced series. This hypothesis is independent of structural breaks; change occurred and we are only interested whether y_{it} is nonstationary (acceptance of $H_{1,0}$) or not. In H_2 the breaks appear only under the alternative. Accepting $H_{2,0}$ means that y_{it} is nonstationary and no break occurred. In the event of rejection of $H_{2,0}$, y_{it} is stationary and breaks did happen at the estimated dates.

3 The Doubly Modified Estimator

When $\varphi = 1$, (1) can be written as:

$$y_i = y_{i,-1} + \Delta X^{(\lambda)}\pi_i^{(\lambda)} + u_i, \tag{2}$$

where $\Delta X^{(\lambda)} = X^{(\lambda)} - X_{-1}^{(\lambda)}$, $X_{-1}^{(\lambda)} = LX^{(\lambda)}$ and where L is the lag operator. Given that $X^{(\lambda)}$ contains $(m+1)$ intercept vectors and $(m+1)\rho$ trend vectors as columns, $\Delta X^{(\lambda)}$ contains $(m+1)$ differenced intercept vectors and $(m+1)\rho$ differenced trend vectors. If for example $m = \rho = 1$, $\Delta X^{(\lambda)} = \left[\Delta e^{(1)}, \Delta e^{(2)}, \Delta\tau^{(1)}, \Delta\tau^{(2)}\right]$ and there exist $(m+1)\rho = (1+1)1 = 2$ differenced trend columns: $[\Delta X^{(\lambda)}]_3 = \Delta\tau^{(1)}$ and $[\Delta X^{(\lambda)}]_4 = \Delta\tau^{(2)}$.

Define Λ a $T \times T$ matrix with $[\Lambda]_{i,j} = 1$ if $i < j$ and zeros elsewhere. Furthermore, let $P^{(\lambda)} \equiv [v_1^{(\lambda)}, ..., v_s^{(\lambda)}]$ and let $v_1^{(\lambda)}, ..., v_s^{(\lambda)}$ be the $T \times 1$ vectors which form a base of $R\{(\Delta X^{(\lambda)}|\Lambda\Delta X^{(\lambda)})\}$. Then we define the orthogonal projection matrix

$$Q^{(\lambda)} = I_T - P^{(\lambda)}(P^{(\lambda)\prime}P^{(\lambda)})^{-1}P^{(\lambda)\prime},$$

which has the property that $Q^{(\lambda)}\Delta X^{(\lambda)} = Q^{(\lambda)}\Lambda\Delta X^{(\lambda)} = 0$ by construction. We shall further assume that $Q^{(\lambda)} \neq 0$ and thus the column rank of $P^{(\lambda)}$ must be less than T.

The basic estimator that we employ is the WG estimator:

$$\hat{\varphi}^{(\lambda)} = \frac{\sum_{i=1}^{N} y_{i,-1}' Q^{(\lambda)} y_i}{\sum_{i=1}^{N} y_{i,-1}' Q^{(\lambda)} y_{i,-1}}. \tag{3}$$

This estimator has the property that, when $\varphi = 1$, it is invariant to the initial conditions of the panel y_{i0} and the individual effects $\pi_i^{(\lambda)}$. Thus, assumptions on y_{i0}, like mean and covariance stationarity made by the generalized method of moments and conditional or unconditional maximum likelihood estimation procedures (see, e.g., Bond et al. (2005) and Kruiniger (2008)) are no longer required. The WG estimator $\hat{\varphi}^{(\lambda)}$ is also attractive for its small sample properties. De Wachter et al. (2007) and Han and Phillips (2012) have noticed that the performance of the GMM estimator of φ, compared to $\hat{\varphi}^{(\lambda)}$, deteriorates in small samples due to the inaccurate estimation of its weighting matrix. Furthermore, Han and Phillips (2013) have found pathologies of the first difference maximum likelihood with a high impact on small sample performance.

The WG estimator is known to be inconsistent in dynamic panel models when the T is finite, see e.g. Nickel (1981) and Harris and Tzavalis (1999) inter alia. Let $\Gamma_N \equiv (1/N)\sum_{i=1}^{N} \Gamma_i$ where $\Gamma_i = E(u_i u_i')$. It is assumed that the maximum order of serial correlation is $p_{\max}^{(\lambda,H)}$, that is $E(u_{it}u_{is}) = 0$ if $|t-s| > p_{\max}^{(\lambda,H)}$. The properties of $p_{\max}^{(\lambda,H)}$ are given in Assumption A below. It is shown (Lemma 1) that $p\lim_{N\to\infty}\left(\hat{\varphi}^{(\lambda)} - 1 - b^{(\lambda)}/d^{(\lambda)}\right) = 0$,where $b^{(\lambda)} = \mathrm{tr}(\Lambda' Q^{(\lambda)} \Gamma_N)$ and $d^{(\lambda)} = \mathrm{tr}(\Lambda' Q^{(\lambda)} \Lambda \Gamma_N)$. The inconsistency is given by the term $b^{(\lambda)}/d^{(\lambda)}$. Ideally we would like to estimate Γ_N and plug this in $b^{(\lambda)}$ and $d^{(\lambda)}$ to bias correct the estimator.

Define $\hat{\Gamma} = (1/N)\sum_{i=1}^{N} \Delta y_i \Delta y_i'$, where $\Delta y_i = y_i - y_{i,-1}$, to be an estimator for Γ_N. Furthermore, let $E(\pi_i^{(\lambda)}\pi_i^{(\lambda)\prime}) = \Pi_i$ and $\Pi_N \equiv (1/N)\sum_{i=1}^{N} \Pi_i$. We would like to use $\hat{\Gamma}$ as an estimator for Γ_N but as it is shown in Theorem 1, $p\lim_{N\to\infty}(\hat{\Gamma} - \Gamma_N - \Delta X^{(\lambda)} \Pi_N \Delta X^{(\lambda)\prime}) = 0$ and thus $\hat{\Gamma}$ must be bias-corrected as well. This is the second correction and to do so we need an estimator for Π_N. The $k \times k$ matrix Π_N can be partitioned as

$$\Pi_N = \begin{pmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{12}' & \Pi_{22} \end{pmatrix}, \tag{4}$$

where Π_{22} is a $[(m+1)\rho] \times [(m+1)\rho]$ symmetric submatrix of nuisance parameters which are moments of the coefficients of the differenced trend vectors $[\Delta X^{(\lambda)}]_j$, for $j = m+2, ..., k$. As we will explain later, we will only need to estimate the elements of Π_{22}.

For each element of Π_{22} a deterministic $T \times T$ selection matrix $Z_{i,j}^{(\lambda)}$ is needed, and since Π_{22} is symmetric it must be that $Z_{j,i}^{(\lambda)} = Z_{i,j}^{(\lambda)}$. The $Z_{i,j}^{(\lambda)}$ will be used in creating the estimators of the elements of Π_{22}. In total we will need $[(m+1)\rho][(m+1)\rho+1]/2$ matrices $Z_{i,j}^{(\lambda)}$. For $j = m+2, ..., k$, define $[\Delta X^{*(\lambda)}]_{t,j} = [\Delta X^{(\lambda)}]_{t,j}$ for all $t = 1, ..., T$, except $[\Delta X^{*(\lambda)}]_{t,j} = 0$ at $t = T_\mu + 1$ for $\mu = 1, ..., m$. Also, by defining $\tilde{Z}$ to be a $T^2 \times [(m+1)\rho][(m+1)\rho+1]/2$ matrix with columns

$$[\tilde{Z}]_{i+\frac{(j-1)j}{2}} = \begin{cases} \mathrm{vec}\left\{\left[\Delta X_{m+1+i}^{*(\lambda)} \Delta X_{m+1+j}^{*(\lambda)\prime}\right]_p^-\right\} \text{ if } i = j, \\ \mathrm{vec}\left\{\left[\Delta X_{m+1+i}^{*(\lambda)} \Delta X_{m+1+j}^{*(\lambda)\prime} + \Delta X_{m+1+j}^{*(\lambda)} \Delta X_{m+1+i}^{*(\lambda)\prime}\right]_p^-\right\} \text{ if } i \neq j, \end{cases} \tag{5}$$

for $i = 1, ..., (m+1)\rho$ and $j = 1, ..., i$, we have that

$$\mathrm{vech}(\hat{\Pi}_{22}) = (\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'\mathrm{vec}(\hat{\Gamma}), \tag{6}$$

is an estimator for vech(Π_{22}). From this expression, we derive the $Z_{i,j}^{(\lambda)}$ to be:

$$Z_{i,j}^{(\lambda)} = \text{vec}_{T,T}^{-1}\left\{\left[\tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\right]_{i+\frac{(j-1)j}{2}}\right\}. \tag{7}$$

for $i = 1, ..., (m+1)\rho$ and $j = 1, ..., i$. The selection matrices $Z_{i,j}^{(\lambda)}$ implement the method of covariance matrix estimation of Abowd and Card (1989) and Arellano (1990).

We propose the doubly modified estimator:

$$\hat{\varphi}_{\text{DME}}^{(\lambda)} = \hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}}, \tag{8}$$

where $\hat{d}^{(\lambda)} = (1/N)\sum_{i=1}^{N} y_{i,-1}' Q^{(\lambda)} y_{i,-1}$ and $\hat{b}^{(\lambda)} = \text{tr}(\Theta^{(\lambda)}\hat{\Gamma})$. The numerator bias correction $\hat{b}^{(\lambda)}$ is based on $\Theta^{(\lambda)}$, a $T \times T$ matrix with

$$\Theta^{(\lambda)} = \Psi^{(\lambda)} - \sum_{i=m+2}^{k}\sum_{j=m+2}^{k}\left[\text{tr}(\Psi^{(\lambda)}[\Delta X^{(\lambda)}]_i[\Delta X^{(\lambda)}]_j')Z_{i,j}^{(\lambda)}\right], \tag{9}$$

where $\Psi^{(\lambda)} = [\Lambda' Q^{(\lambda)}]^{+}_{p_{\max}^{(\lambda,H)}}$.

The $T \times 1$ vectors $[\Delta X^{(\lambda)}]_j$ for $j = m+2, ..., k$ are the $(m+1)\rho$ differenced trend columns of $\Delta X^{(\lambda)}$ and these vectors are used in the estimation of Π_{22}. There is no need to estimate the elements of Π_N appearing in Π_{11} and Π_{12} because these are multiplied by 0 in the definition of $\Theta^{(\lambda)}$, since $\text{tr}(\Psi^{(\lambda)}[\Delta X^{(\lambda)}]_i[\Delta X^{(\lambda)}]_j') = 0$ if i or j are equal to $1, ..., m+1$, (see also Lemma 2). If there are no trend vectors i.e., $X^{(\lambda)} = \left[e^{(1)}, e^{(2)}\right]$, then $\rho = 0$ and we set $\Theta^{(\lambda)} = \Psi^{(\lambda)}$.

Inference on both H_1 and H_2 will be based on the t-statistic,

$$t^{(\lambda)} = \frac{\left(\hat{\varphi}_{\text{DME}}^{(\lambda)} - 1\right)}{\sqrt{\hat{V}^{(\lambda)}/(N\hat{d}^{(\lambda)2})}}, \tag{10}$$

where $\hat{V}^{(\lambda)} = F^{(\lambda)\prime}\hat{\Xi}F^{(\lambda)}$, $F^{(\lambda)} = \text{vec}(Q^{(\lambda)}\Lambda - \Theta^{(\lambda)\prime})$ and $\hat{\Xi} = (1/N)\sum_{i=1}^{N}\text{vec}(\Delta y_i \Delta y_i')\text{vec}(\Delta y_i \Delta y_i')'$ is an estimator of $\Xi_N \equiv (1/N)\sum_{i=1}^{N} V(\text{vec}(\Delta y_i \Delta y_i'))$. The estimator $\hat{\Xi}$ is a computationally attractive alternative to $\hat{\Xi}^* = (1/N)\sum_{i=1}^{N}\text{vec}(\Delta y_i\Delta y_i')\text{vec}(\Delta y_i\Delta y_i')' - (1/N^2)\sum_{i=1}^{N}\text{vec}(\Delta y_i\Delta y_i')\sum_{i=1}^{N}\text{vec}(\Delta y_i\Delta y_i')'$, which is the standard consistent estimator found in the literature, see e.g. Arellano (2003), equation 5.58. While $\hat{\Xi}$ is not a consistent estimator of Ξ_N, both $\hat{\Xi}$ and $\hat{\Xi}^*$ lead to consistent estimation of $V^{(\lambda)} = F^{(\lambda)\prime}\Xi_N F^{(\lambda)}$ as it is shown in Lemma 5, in the Appendix.

4 Assumptions

In this section we present the assumptions that we use. Regarding serial correlation and heteroscedasticity in u_{it}, we consider the following assumption:

Assumption A

(i) $\{u_i\}$ for $i = 1, ..., N$, is a sequence of independent random vectors, with $E(u_i) = 0$ and $E(u_i u_i') = \Gamma_i$.

(ii) Under $H_{1,0}$, $p_{\max}^{(\lambda,H)} = \min_j\{T_{j+1} - T_j - 2 - \rho\}$ for $j \in \{1, ..., m\}$, while under $H_{2,0}$, $p_{\max}^{(\lambda,H)} = T - 1 - \rho$.

(iii) $E(\|u_i\|^{8+\delta}) < M < +\infty$ for $i = 1, ..., N$ and $\delta > 0$.

(iv) $p\lim_{N\to\infty}\Gamma_N > 0$.

(v) $p\lim_{N\to\infty}\Xi_N > 0$.

Condition (i) of Assumption A allows for heteroscedastic and autocorrelated u_{it}. The form of heteroscedasticity and serial correlation may vary with i.

Condition (ii) determines that the maximum order of serial correlation is $p_{\max}^{(\lambda,H)}$ and therefore only $p_{\max}^{(\lambda,H)}$-dependent u_{it} are allowed. Although it is assumed that $p_{\max}^{(\lambda,H)}$ is common for all i, each cross-section unit i can exhibit a different order of serial correlation provided that this does not exceed $p_{\max}^{(\lambda,H)}$. The superscript λ denotes dependence on the trend function through the location of the structural breaks and H denotes dependence on the hypothesis being tested (H_1 or H_2). If structural breaks do not occur then $p_{\max}^{(\lambda,H)}$ no longer depends on λ and H, as in Karavias and Tzavalis (2016). This condition provides the necessary moments for the estimation of Π_N. For the covariance estimation methodology to work, it must be that for some i, j, $[p\lim_N \hat{\Gamma}]_{i,j}$ contain only nuisance parameters of Π_N and not of Γ_N. The reduction coming from ρ is a small one; $\rho = 1$ for models with linear trends and $\rho = 2$ for models with linear and quadratic trends. The value $p_{\max}^{(\lambda,H)}$ is the upper bound of serial correlation allowed. However, in a specific application if there is evidence of weaker or no dependence, it is entirely possible to select a $p \leq p_{\max}^{(\lambda,H)}$. This will result in a more powerful test because more moments are available for the estimation of Π_{22}.

This assumption is general enough to accommodate many applications, see e.g. Schwert (1987). Furthermore, while heteroscedasticity and autocorrelation consistent estimators (HAC) estimators are known to be biased and have issues with their performance (see e.g. Kiefer et al. (2000) and the remark below Theorem 2 in Moon and Perron (2004)), our method results in excellent size control as will be shown later, because of the way we use the cross section dimension. In the fixed-T literature Karavias and Tzavalis (2014b) allow for AR(2) errors with a single break. However that method cannot be further extended to allow for more breaks, trends or heteroscedasticity.

Condition (iii) of Assumption A imposes a uniform bound on the eighth moments of the errors. Uniformity until the fourth moment fulfils the Lyapunov condition which is a sufficient condition for the Lindeberg-Feller central limit theorem. Uniformity of the eighth moment is needed for so that $\hat{V}^{(\lambda)} - V^{(\lambda)} \xrightarrow{p} 0$. If the u_i are identically distributed across i then only a uniform bound on the fourth moments is needed. Conditions (iv) and (v) guarantee that the probability limits of relevant denominators will not be zero.

Assumption B

(i) $\{\pi_i^{(\lambda)}\}$ for $i = 1, ..., N$, is a sequence of independent random vectors which are independent of u_{it} for all i and t.

(ii) $E(\left\|\pi_i^{(\lambda)}\right\|^{4+\delta}) < +\infty$ for $i = 1, ..., N$, and $\delta > 0$.

Assumption B describes the probabilistic behaviour of the individual effects and is needed so that $\hat{V}^{(\lambda)} - V^{(\lambda)} \xrightarrow{p} 0$. Conditions (i)-(ii) are standard in the literature, see e.g. Bai (2013).

Assumption C

(i) If $\rho = 0$, then $\min\{T_{j+1} - T_j\} \geq 1$ for $j \in \{0, ..., m\}$.

(ii) If $\rho > 0$ then $T_1 - T_0 \geq 1 + \rho$ and $\min\{T_{j+1} - T_j\} \geq 2 + \rho$ for $j \in \{1, ..., m\}$.

The two conditions restrain the number of breaks and their position. Similar assumptions appear in the time series literature as well, see e.g. Assumption 3 of Bai and Perron (1998). Condition (i) determines the set I when (1) contains intercepts only. This condition is the weakest because for this model $\Theta^{(\lambda)} = \Psi^{(\lambda)}$ and no trend nuisance parameters need to be estimated. For a model with two breaks $I = \{(2,3), (2,4), ..., (T-2, T-1)\}$. The existence of available moments that allow the estimation of the trend nuisance parameters is guaranteed by condition (ii). In the presence of linear trends and two structural breaks, the breaks can take place in the set $I = \{(2,5), (2,6), ..., (T-6, T-3)\}$. The requirement that structural breaks in the trend model must be at least three periods apart is somewhat stronger than the single time series result by which the breaks must be at least two periods apart, see Lumsdaine and Papell (1997).

5 Asymptotic results

5.1 Asymptotic bias

The following lemma provides the inconsistency of $\hat{\varphi}^{(\lambda)}$.

Lemma 1 *Under the assumptions A and C, the null hypotheses $H_{f,0}$, $f = 1, 2$ and the dates of the breaks known, as $N \to \infty$,*

$$p \lim_{N\to\infty} \left(\hat{\varphi}^{(\lambda)} - \frac{b^{(\lambda)}}{d^{(\lambda)}} - 1 \right) = 0, \tag{11}$$

where $b^{(\lambda)} = \mathrm{tr}(\Lambda' Q^{(\lambda)} \Gamma_N)$ and $d^{(\lambda)} = \mathrm{tr}(\Lambda' Q^{(\lambda)} \Lambda \Gamma_N)$.

The above expressions show that the WG estimator is inconsistent as $N \to \infty$. The bias (inconsistency) of $\hat{\varphi}^{(\lambda)}$ is given by $b^{(\lambda)}/d^{(\lambda)}$ and depends on both the deterministic specification of the model, which is captured by $Q^{(\lambda)}$ and Λ, and the assumptions about the error terms u_{it} reflected in Γ_N.

The main idea of the paper is to propose an estimator for $b^{(\lambda)}/d^{(\lambda)}$ and with that modify the WG estimator, $\hat{\varphi}^{(\lambda)}_{\mathrm{DME}} = \hat{\varphi}^{(\lambda)} - \hat{b}^{(\lambda)}/\hat{d}^{(\lambda)}$. Note that $\hat{\varphi}^{(\lambda)}$ is only adjusted for the bias of its numerator (Phillips and Hansen (1990) and Kruiniger and Tzavalis (2002)). This method is different from that in Karavias and Tzavalis (2014b) and is much more flexible.

When estimating $b^{(\lambda)} = \mathrm{tr}(\Lambda' Q^{(\lambda)} \Gamma_N)$, since the dates of the breaks are known, $Q^{(\lambda)}$ is known and so we only need to find an estimator for Γ_N such that $\hat{\Gamma} - \Gamma_N \xrightarrow{p} 0$. Applying the $\hat{\Gamma}$ results in two problems: first, simply plugging $\hat{\Gamma}$ in $\mathrm{tr}(\Lambda' Q^{(\lambda)} \Gamma_N)$ is not sufficient because it will result in an identity, i.e. after simple algebra, it can be shown that:

$$\hat{\varphi}^{(\lambda)} - \frac{\mathrm{tr}(\Lambda' Q^{(\lambda)} \hat{\Gamma})}{\sum_{i=1}^{N} y'_{i,-1} Q^{(\lambda)} y_{i,-1}} = \frac{\sum_{i=1}^{N} \left(u'_i \Lambda' Q^{(\lambda)} u_i - u'_i \Lambda' Q^{(\lambda)} u_i \right)}{\sum_{i=1}^{N} y'_{i,-1} Q^{(\lambda)} y_{i,-1}} = 0. \tag{12}$$

This happens because we are using the full sample information twice, once to obtain the WG estimator and once to obtain $\hat{b}^{(\lambda)}$. To bypass this problem, we propose the use of a restricted form of $\Lambda' Q^{(\lambda)}$ denoted by $\Psi^{(\lambda)}$. The restriction is such that $\mathrm{tr}(\Psi^{(\lambda)} \Gamma_N) = \mathrm{tr}(\Lambda' Q^{(\lambda)} \Gamma_N)$; the matrix $\Psi^{(\lambda)}$ gives the non-zero elements of Γ_N, the same weight that $\Lambda' Q^{(\lambda)}$ does. It also gives zero weight to the zero elements of Γ and thus $\Psi^{(\lambda)} - \Lambda' Q^{(\lambda)} \neq 0$ so that the last equality in (12) is avoided.

The second problem is that, $\hat{\Gamma} - \Gamma_N \xrightarrow{p} 0$ only when $\Delta y_i = u_i$. In this particular case, one can simply set $\hat{b}^{(\lambda)} = \mathrm{tr}(\Psi^{(\lambda)} \hat{\Gamma})$ and proceed with inference. In general however, this does not hold as it is shown in the next theorem, and the more complicated $\Theta^{(\lambda)}$ matrix is needed.

Theorem 1 *Under Assumptions A and B, the null hypotheses $H_{f,0}$, $f = 1, 2$ and the dates of the breaks known, as $N \to \infty$,*

$$p \lim_{N\to\infty} \left[\hat{\Gamma} - \Gamma_N - \Delta X^{(\lambda)} \Pi_N \Delta X^{(\lambda)\prime} \right] = 0. \tag{13}$$

The adjustment of $\hat{\Gamma}$ to render it net of the break nuisance parameters is the second modification which appears in $\hat{\varphi}^{(\lambda)}_{\mathrm{DME}}$. Notice that we are not interested in estimating Γ_N per se, but in finding a

$\Theta^{(\lambda)}$ such that $\mathrm{tr}(\Theta^{(\lambda)}\hat{\Gamma}) - \mathrm{tr}(\Lambda' Q^{(\lambda)}\Gamma_N) \xrightarrow{p} 0$. This creates a dichotomy in the nuisance parameters of $\Delta X^{(\lambda)}\Pi_N \Delta X^{(\lambda)\prime}$: a) those parameters related to the "crash" vectors of $\Delta X^{(\lambda)}$ and appear in Π_{11} and Π_{12} and b) those related to the differenced trend vectors of $\Delta X^{(\lambda)}$ and appear in Π_{22}. By "crash" vectors we define those vectors that are everywhere equal to zero, except at the points $T_j + 1$, $j = 1, ..., m$ where they are either equal to 0 or equal to 1. If i.e., $\Delta X^{(\lambda)} = \left[\Delta e^{(1)}, \Delta e^{(2)}, \Delta\tau^{(1)}, \Delta\tau^{(2)}\right]$, the $\Delta e^{(1)}, \Delta e^{(2)}$ are crash vectors. Crash vectors are the outcome of differencing. The following lemma shows that there is no need to estimate the first type of nuisance parameters and that we only need estimates of the parameters in Π_{22}. The case of no structural breaks in the intercepts is trivial, since $\Delta e = 0$ and therefore there are no crash vectors.

Lemma 2 *Let $e^{(c)}$ be any "crash" vector appearing in $\Delta X^{(\lambda)}$. Then it holds that $\Psi^{(\lambda)} e^{(c)} = 0$.*

Crash vectors appear also when differencing trend vectors; in the previous example $\Delta\tau^{(1)} = -T_1 \Delta e^{(1)} + e^{(1)}$, and $\Delta\tau^{(2)} = T_1 \Delta e^{(2)} + e^{(2)}$. We avoid their interference in the Π_{22} estimation problem by working with $[\Delta X^{*(\lambda)}]_j$ in place of the original differenced trend vectors $[\Delta X^{(\lambda)}]_j$, for $j = m+2, ..., k$.

We employ the covariance matrix estimation method of Abowd and Card (1989) and Arellano (1990), which is a method of moments that leads to consistent estimation of Π_{22}. The intuition is that, because $\Gamma_N = [\Gamma_N]^{+}_{p_{\max}^{(\lambda,H)}}$ by Assumption A(ii), inside $\Gamma_N + \Delta X^{(\lambda)}\Pi_N \Delta X^{(\lambda)\prime}$ exist secondary diagonals that contain elements consisting only of Π_{22} nuisance parameters.

Theorem 2 *Under Assumptions A-C, the null hypotheses $H_{f,0}$, $f = 1, 2$ and the dates of the breaks known, as $N \to \infty$,*

$$p \lim_{N\to\infty} \left[\mathrm{vech}(\hat{\Pi}_{22}) - \mathrm{vech}(\Pi_{22})\right] = 0. \tag{14}$$

Now we are ready to use $\hat{\Gamma}$ in order to get $\hat{b}^{(\lambda)} - b^{(\lambda)} \xrightarrow{p} 0$:

Theorem 3 *Under Assumptions A-C, the null hypotheses $H_{f,0}$, $f = 1, 2$ and the dates of the breaks known, as $N \to \infty$,*

$$p \lim_{N\to\infty} \left[\mathrm{tr}(\Theta^{(\lambda)}\hat{\Gamma}) - \mathrm{tr}(\Lambda' Q^{(\lambda)}\Gamma_N)\right] = 0. \tag{15}$$

5.2 Limiting distribution

The following theorem provides the distribution of the test statistic for when the dates of the breaks are known

Theorem 4 *Under Assumptions A-C, $H_{f,0}$ with $f = 1, 2$, the dates of the breaks known, and as $N \to \infty$:*

$$t^{(\lambda)} = \frac{\left(\hat{\varphi}^{(\lambda)}_{\mathrm{DME}} - 1\right)}{\sqrt{\hat{V}^{(\lambda)}/(N\hat{d}^{(\lambda)2})}} \xrightarrow{d} N(0, 1). \tag{16}$$

If the dates of the breaks are unknown, then it becomes important whether the null hypothesis is $H_{1,0}$ or $H_{2,0}$. In the first case the breaks appear under both $H_{1,0}$ and $H_{1,1}$ and they can be estimated in a first step by Bai's (2010) method, using first differenced data. That estimator has an $o_p(\sqrt{N})$ rate of convergence and thus we can assume the break points are known and apply the results of Theorem 2.

If the null hypothesis is $H_{2,0} : \varphi = 1$ and $\pi^{(1)}_{j,i} = \pi^{(2)}_{j,i} = ... = \pi^{(m+1)}_{j,i}$ for all j, then the structural break parameters appear only under the alternative. This is not a regular hypothesis testing problem; see Davies (1977) and Andrews (1993). Following this literature, the selection of the break dates is viewed as the outcome of minimizing $t^{(\lambda)}$ over all possible combinations of break dates. In this way, the estimated dates are those that give more weight to the alternative. The null hypothesis is rejected when

$$\inf_{\lambda \in I} t^{(\lambda)} < z_a^{\inf}, \tag{17}$$

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

Theorem 5 *Under Assumptions A-C, the null hypotheses $H_{2,0}$, the dates of the breaks unknown, as $N \to \infty$*

$$\inf_{\lambda \in I} t^{(\lambda)} \xrightarrow{d} G, \tag{18}$$

where G is distributed as the infimum of a fixed number of mean-zero normal variables. These normals are correlated with asymptotic variance-covariance matrix Σ given by:

$$[\Sigma]_{\mu,\nu} = \frac{F^{(\mu)\prime}\Xi F^{(\nu)}}{\sqrt{F^{(\mu)\prime}\Xi F^{(\mu)}}\sqrt{F^{(\nu)\prime}\Xi F^{(\nu)}}}, \tag{19}$$

where μ and ν denote two different partitions that belong in I.

There are two ways of getting critical values from G. First, we could proceed by numerically integrating the analytical pdf function which has been derived by Arellano-Valle and Genton (2008). Otherwise one can proceed by using the bootstrap. The steps are the following:

1. Use the data to compute the test statistic $t^{(\lambda)}$ for each $\lambda \in I$. Compute $\inf_{\lambda \in I} t^{(\lambda)}$ as well.

2. Generate r bootstrap samples of size $T \times N$ by sampling with replacement from residuals $u_i^r = \Delta y_i$ for $i = 1, ..., N$. This resampling scheme is taken across individuals so that the time series properties of the series are maintained. Notice that $u_i^r = \Delta y_i = \Delta X^{(\lambda)} \pi_i^{(\lambda)} + u_i$ contains information on u_i as well as on the trend function and the individual effect parameters.

3. Generate bootstrap samples as

$$\begin{aligned} y_{i,-1}^r &= y_{i0} e + \Lambda u_i^r, \\ y_i^r &= y_{i,-1}^r + u_i^r, \text{ for } i = 1, ..., N. \end{aligned} \tag{20}$$

4. For each bootstrap sample, calculate the statistic $\inf_{\lambda \in I} \left(t^{r,(\lambda)} - t^{(\lambda)} \right)$ where $t^{r,(\lambda)}$ is the test statistic coming from the bootstrap sample.

5. Do this r times and compute the empirical distribution of $\inf_{\lambda \in I} \left(t^{r,(\lambda)} - t^{(\lambda)} \right)$. From this distribution, derive the size a left-tail critical value. If $\inf_{\lambda \in I} t^{(\lambda)}$ is less than this value, reject the null hypothesis.

In the above procedure we use *i.i.d.* resampling exploiting the independence across units. The bootstrap samples are created under the $H_{2,0}$ as it is advocated for unit root processes by Basawa et al. (1991). The proof of the consistency of the bootstrap can possibly be constructed along the lines of Horowitz (2001).

6 Simulation Results

In this section the results of a Monte Carlo study investigating the finite sample performance of the proposed test statistics are reported. Sample sizes for N and T are chosen to be $N = \{25, 50, 100, 500, 1000, 1200\}$ and $T = \{10, 15, 20, 30\}$. All experiments are conducted based on 2000 iterations. The model that we use is the "crash and changing growth" model the previous section, first with one break in the middle of the sample and then with two breaks at $\lfloor 0.35T \rfloor$ and $\lfloor 0.65T \rfloor$. For these two models we consider 5 scenarios:

1) Testing H_1 with u_{it} *i.i.d.* across i and t.
2) Testing H_1 with $u_{it} = \theta_i \varepsilon_{it} + \sigma_{it}\varepsilon_{it-1}$ where $\theta_i \sim i.i.d.\ U[0.2, 0.4]$ and $\sigma_{it} \sim i.i.d.\ U[0.5, 1.5]$.
3) Testing H_1 with $u_{it} = \theta_i \varepsilon_{it} + \sigma_{it}\varepsilon_{it-1}$ where $\theta_i \sim i.i.d.\ U[-0.4, -0.2]$ and $\sigma_{it} \sim i.i.d.\ U[0.5, 1.5]$.
4) Testing H_1 with $\varphi \sim i.i.d.\ U[0.7, 0.9]$ under the alternative and errors as in scenario 2.
5) Testing H_2 with u_{it} *i.i.d.* across i and t.

The individual effects are generated as $\pi_{1i}^{(1)} \sim i.i.d.\ U[-0.05, 0]$, $\pi_{1i}^{(2)} \sim i.i.d.\ U[0, 0.05]$, $\pi_{1i}^{(3)} \sim i.i.d.\ U[0.05, 0.1]$, $\pi_{2i}^{(1)} \sim i.i.d.\ U[0, 0.025]$, $\pi_{2i}^{(2)} \sim i.i.d.\ U[0.025, 0.05]$, $\pi_{2i}^{(3)} \sim i.i.d.\ U[0.05, 0.75]$. The errors ε_{it} are standard normal and so are the u_{it} in scenarios 1 and 5. Under the alternative $\varphi = 0.8$. The initial conditions are set equal to 0 and for scenario 5, the number of bootstrap replications is 199. The values used for this experiment are similar to those of Pesaran et al. (2013). In scenarios 1-4 we assume that the break dates are known while in scenario 5 we assume that they are unknown.

Table 1 presents the results for the model with one structural break and Table 2 presents the results for the case of two structural breaks. We find that the new tests have excellent size properties for all specifications and for all combinations of N and T considered, with the size always being close to the nominal. When it comes to the power of the tests, we have found that adding an extra break, having heterogeneous alternatives, testing H_2 when the breaks are unknown and adding heteroscedasticity and serial correlation reduces the power of the tests. Positive $\theta' s$ lead to tests with lower power than negative $\theta' s$.

7 Concluding Remarks

In this paper we propose a general methodology for testing for unit roots in panel data with a short time series dimension. This methodology, based on the novel DME estimator, allows for models with general trend functions that may contain intercepts, linear and nonlinear trends and multiple structural breaks at unknown dates. The error process is also general as the errors may have unspecified forms of short term serial correlation, heteroscedasticity and cross section heterogeneity. To examine the small sample performance of the tests, the paper conducts a Monte Carlo study. The results of this study demonstrate that the suggested tests have always size close to their nominal level and satisfactory power.

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] Andrews, D.W.K. 1993. Tests for Parameter Instability and Structural Change with Unknown Change Point. Econometrica, 59, 817-858.

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

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

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

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

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

[8] 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.

[9] 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.

[10] Breitung, J. and W. Meyer, 1994, Testing for Unit Roots in Panel Data: Are Wages on Di erent Bargaining Levels Cointegrated? Applied Economics, 26, 353-361.

[11] Bun, Maurice J. G., Sarafidis, V., 2015. Dynamic Panel Data Models. In Badi H. Baltagi, (ed) The Oxford Handbook of Panel Data. Oxford.

[12] Choi, I., 2014. Unit root tests for dependent and heterogeneous micropanels. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2475810

[13] Choi, I., 2015. Almost All About Unit Roots, Foundations, Developments, and Applications. Cambridge University Press.

[14] Davies, R. B. 1977. Hypothesis Testing When a Nuisance Parameter Is Present Only under the Alternative. Biometrika, 64, 247-254

[15] 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.

[16] 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.

[17] Feller, W., 1968. An introduction to probability theory and its applications. Vol 1, 3rd Edition, Wiley, New York.

[18] 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, Wiley Blackwell, vol. 64, pages 123-148, December.

[19] 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.

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

[21] 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.

[22] Hogg, R.V., McKean, J.W., Craig, A., 2013. Introduction to Mathematical Statistics. 7th Edition, Pearson, Harlow.

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

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

[25] 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.

[26] Karavias Y., and Tzavalis E., 2016. Local Power of Fixed-T Panel Unit Root Tests With Serially Correlated Errors and Incidental Trends. Journal of Time Series Analysis, Wiley Blackwell, vol. 37(2), pages 222-239, 03.

[27] Karavias Y., and Tzavalis E., 2017. Local power of panel unit root tests allowing for structural breaks. Econometric Reviews, 36:10, 1123-1156, DOI:10.1080/07474938.2015.1059722

[28] Kiefer, N. M., Vogelsang, T. J. and Bunzel, H., 2000. Simple Robust Testing of Regression Hypotheses. Econometrica 68, 695—714.

[29] 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.

[30] 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.

[31] 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.

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

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

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

[35] 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.

[36] 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.

[37] Robertson, D., Sarafidis, V. and Westerlund, J. 2017. Unit root inference in generally trending and cross-correlated fixed-T panels, (2017) Journal of Business and Economic Statistics, DOI: 10.1080/07350015.2016.1191501.

[38] Schwert, G.W., 1987. Effects of model specification on tests for unit roots in macroeconomic data. Journal of Monetary Economics 20, 73-103.

[39] White, H., 1984. Asymptotic theory for econometricians. New York: Academic Press.

[40] 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.

8 Appendix A

In this appendix, we provide proofs of the theorems presented in the main text of the paper. First we start with a simple example regarding the operators $[\cdot]_p^+$ and $[\cdot]_p^-$.

Example: Let $A \in R^{4\times 4}$, such that

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}.$$

Then

$$[A]_0^+ = \begin{pmatrix} a_{11} & 0 & 0 & 0 \\ 0 & a_{22} & 0 & 0 \\ 0 & 0 & a_{33} & 0 \\ 0 & 0 & 0 & a_{44} \end{pmatrix} \text{ and } [A]_1^+ = \begin{pmatrix} a_{11} & a_{12} & 0 & 0 \\ a_{21} & a_{22} & a_{23} & 0 \\ 0 & a_{32} & a_{33} & a_{34} \\ 0 & 0 & a_{43} & a_{44} \end{pmatrix}.$$

Also,

$$[A]_0^- = \begin{pmatrix} 0 & a_{12} & a_{13} & a_{14} \\ a_{21} & 0 & a_{23} & a_{24} \\ a_{31} & a_{32} & 0 & a_{34} \\ a_{41} & a_{42} & a_{43} & 0 \end{pmatrix} \text{ and } [A]_1^- = \begin{pmatrix} 0 & 0 & a_{13} & a_{14} \\ 0 & 0 & 0 & a_{24} \\ a_{31} & 0 & 0 & 0 \\ a_{41} & a_{42} & 0 & 0 \end{pmatrix}.$$

Proof of Lemma 1. Under any $H_{f,0}$, from (2) and backward substitution we have

$$y_{i,-1} = y_{i,0}e + \Lambda\Delta X^{(\lambda)}\pi_i^{(\lambda)} + \Lambda u_i, \tag{21}$$

for $i = 1, ..., N$. If we pre-multiply the above by $Q^{(\lambda)}$ ($Q^{(\lambda)} \neq 0$ by Assumption C) we get

$$Q^{(\lambda)}y_{i,-1} = Q^{(\lambda)}\Lambda u_i, \tag{22}$$

as $Q^{(\lambda)}\Lambda\Delta X^{(\lambda)} = Q^{(\lambda)}\Delta X^{(\lambda)} = 0$. This happens because the matrix $P^{(\lambda)}$ is made up by column vectors which form a base of the column space of $(\Delta X^{(\lambda)}|\Lambda\Delta X^{(\lambda)})$. Since $\Delta X^{(\lambda)}$ and $\Lambda\Delta X^{(\lambda)}$ belong to the column space spanned by $P^{(\lambda)}$, it holds that

$$P^{(\lambda)}(P^{(\lambda)\prime}P^{(\lambda)})^{-1}P^{(\lambda)\prime}\Delta X^{(\lambda)} = \Delta X^{(\lambda)},$$

and that

$$P^{(\lambda)}(P^{(\lambda)\prime}P^{(\lambda)})^{-1}P^{(\lambda)\prime}\Lambda\Delta X^{(\lambda)} = \Lambda\Delta X^{(\lambda)},$$

and thus $Q^{(\lambda)}\Lambda\Delta X^{(\lambda)} = Q^{(\lambda)}\Delta X^{(\lambda)} = 0$. When $X^{(\lambda)}$ contains intercepts, $Q^{(\lambda)}e = 0$ and (22) makes $\hat{\varphi}^{(\lambda)}$ invariant to the initial conditions.

Then,

$$\begin{aligned}
\hat{\varphi}^{(\lambda)} - \frac{\mathrm{tr}(\Lambda' Q^{(\lambda)}\Gamma_N)}{\mathrm{tr}(\Lambda' Q^{(\lambda)}\Lambda\Gamma_N)} - 1 &= \frac{\sum_{i=1}^N y_{i,-1}' Q^{(\lambda)} y_i}{\sum_{i=1}^N y_{i,-1}' Q^{(\lambda)} y_{i,-1}} - \frac{\mathrm{tr}(\Lambda' Q^{(\lambda)}\Gamma_N)}{\mathrm{tr}(\Lambda' Q^{(\lambda)}\Lambda\Gamma_N)} - 1 \\
&= \frac{\sum_{i=1}^N y_{i,-1}' Q^{(\lambda)} (y_{i,-1} + \Delta X^{(\lambda)}\pi_i^{(\lambda)} + u_i)}{\sum_{i=1}^N y_{i,-1}' Q^{(\lambda)} y_{i,-1}} - \frac{\mathrm{tr}(\Lambda' Q^{(\lambda)}\Gamma_N)}{\mathrm{tr}(\Lambda' Q^{(\lambda)}\Lambda\Gamma_N)} - 1 \\
&= 1 + \frac{\sum_{i=1}^N y_{i,-1}' Q^{(\lambda)} u_i}{\sum_{i=1}^N y_{i,-1}' Q^{(\lambda)} y_{i,-1}} - \frac{\mathrm{tr}(\Lambda' Q^{(\lambda)}\Gamma_N)}{\mathrm{tr}(\Lambda' Q^{(\lambda)}\Lambda\Gamma_N)} - 1 \\
&= \frac{\sum_{i=1}^N u_i' \Lambda' Q^{(\lambda)} u_i}{\sum_{i=1}^N u_i' \Lambda' Q^{(\lambda)} \Lambda u_i} - \frac{\mathrm{tr}(\Lambda' Q^{(\lambda)}\Gamma_N)}{\mathrm{tr}(\Lambda' Q^{(\lambda)}\Lambda\Gamma_N)}, \qquad (23)
\end{aligned}$$

where the last equality comes from (22). Notice that $E(u_i'\Lambda' Q^{(\lambda)} u_i) = \mathrm{tr}[\Lambda' Q^{(\lambda)} E(u_i u_i')] = \mathrm{tr}(\Lambda' Q^{(\lambda)}\Gamma_i)$. Thus

$$\frac{1}{N}\sum_{i=1}^N u_i'\Lambda' Q^{(\lambda)} u_i - \frac{1}{N}\sum_{i=1}^N \mathrm{tr}(\Lambda' Q^{(\lambda)}\Gamma_i) = \frac{1}{N}\sum_{i=1}^N u_i'\Lambda' Q^{(\lambda)} u_i - \mathrm{tr}(\Lambda' Q^{(\lambda)}\Gamma_N),$$

and under Assumption A, from Markov's Law of Large Numbers:

$$\frac{1}{N}\sum_{i=1}^N u_i'\Lambda' Q^{(\lambda)} u_i - \mathrm{tr}(\Lambda' Q^{(\lambda)}\Gamma_N) \xrightarrow{p} 0, \qquad (24)$$

(see e.g. White (1984), p. 33). Similarly, $E(u_i'\Lambda' Q^{(\lambda)}\Lambda u_i) = \mathrm{tr}[\Lambda' Q^{(\lambda)}\Lambda E(u_i u_i')] = \mathrm{tr}(\Lambda' Q^{(\lambda)}\Lambda\Gamma_i)$ and

$$\frac{1}{N}\sum_{i=1}^N u_i'\Lambda' Q^{(\lambda)}\Lambda u_i - \mathrm{tr}(\Lambda' Q^{(\lambda)}\Gamma_N) \xrightarrow{p} 0. \qquad (25)$$

Because we are interested in convergence in probability, the following holds by combining (24), (25) and a form of Slutsky's Theorem (see e.g. Hogg, McKean and Craig (2013), p. 297)

$$\frac{\sum_{i=1}^N u_i'\Lambda' Q^{(\lambda)} u_i}{\sum_{i=1}^N u_i'\Lambda' Q^{(\lambda)}\Lambda u_i} - \frac{\mathrm{tr}(\Lambda' Q^{(\lambda)}\Gamma_N)}{\mathrm{tr}(\Lambda' Q^{(\lambda)}\Lambda\Gamma_N)} \xrightarrow{p} 0.$$

Note that the denominator $p\lim_N \mathrm{tr}(\Lambda' Q^{(\lambda)}\Lambda\Gamma_N)$ is different than zero by Assumption A(iv) which states that $p\lim_N \Gamma_N$ is positive definite. First, notice that $\Lambda' Q^{(\lambda)}\Lambda$ is positive semidefinite as for any vector $x \in R^T$, $x'\Lambda' Q^{(\lambda)}\Lambda x = (\Lambda x)' Q^{(\lambda)} (\Lambda x) \geq 0$, because $Q^{(\lambda)}$ is positive semidefinite as a projection matrix. Second, $p\lim_N \Gamma_N$ is positive definite and can be decomposed as $L_* L_*'$ where L_* is a lower triangular matrix, by the Cholesky decomposition. Thus $\mathrm{tr}(\Lambda' Q^{(\lambda)}\Lambda p\lim_N \Gamma_N) = \mathrm{tr}(\Lambda' Q^{(\lambda)}\Lambda L_* L_*') = \mathrm{tr}(L_*'\Lambda' Q^{(\lambda)}\Lambda L_*)$. Notice that $\Lambda' Q^{(\lambda)} \neq 0$ as $Q^{(\lambda)} \neq 0$. The only other possible case for $\Lambda' Q^{(\lambda)} = 0$ would be if $Q^{(\lambda)}$ had non-zero elements in its first row and zero everywhere - but this is not possible as $Q^{(\lambda)}$ would no longer be symmetric. Furthermore, ΛL_* has the same non-zero elements with Λ. Similarly, $(\Lambda' Q^{(\lambda)})$ is not a matrix which is everywhere zero but for its first column and thus $\left(\Lambda' Q^{(\lambda)}\right)\Lambda \neq 0$. Therefore $L_*'\Lambda' Q^{(\lambda)}\Lambda L_* \neq 0$. By the same argument as before, $L_*'\Lambda' Q^{(\lambda)}\Lambda L_*$ is positive semidefinite and its eigenvalues are greater or equal than zero. Thus, $\mathrm{tr}(L_*'\Lambda' Q^{(\lambda)}\Lambda L_*) > 0$ because the trace is equal to the sum of the eigenvalues.

Proof of Theorem 1. Rewriting (2) we have that $\Delta y_i = \Delta X^{(\lambda)}\pi_i^{(\lambda)} + u_i$ and thus,

$$E(\Delta y_i \Delta y_i') = \Delta X^{(\lambda)}\Pi_i \Delta X^{(\lambda)\prime} + \Gamma_i,$$

for $i = 1, ..., N$. The last expression holds because Assumption B requires that the individual effects and the error terms are independent. Then

$$\frac{1}{N}\sum_{i=1}^{N}\Delta y_i \Delta y_i' - \frac{1}{N}\sum_{i=1}^{N}(\Delta X^{(\lambda)}\Pi_i \Delta X^{(\lambda)\prime} + \Gamma_i) = \frac{1}{N}\sum_{i=1}^{N}\Delta y_i \Delta y_i' - \Delta X^{(\lambda)}\Pi_N \Delta X^{(\lambda)\prime} - \Gamma_N.$$

Markov's Law of Large Numbers applies elementwise under Assumptions A and B and thus,

$$\frac{1}{N}\sum_{i=1}^{N}\Delta y_i \Delta y_i' - \Delta X^{(\lambda)}\Pi_N \Delta X^{(\lambda)\prime} - \Gamma_N \xrightarrow{p} 0.$$

Proof of Lemma 2. As mentioned above, a "crash" vector is everywhere equal to zero except at the points $T_j + 1$, $i = 1, ..., m$ where it is either equal to 0 or equal to 1, but is never equal to zero everywhere. We prove the statement for $m = 1$ but the same arguments can be applied to the case of $m > 1$. Without loss of generality, assume that $e^{(c)}$ is a crash vector that is everywhere equal to zero and it is equal to 1 at $T_1 + 1$.

By definition, $\Psi^{(\lambda)} = [\Lambda' Q^{(\lambda)}]_p^+$ and $Q^{(\lambda)} = I_T - P^{(\lambda)}(P^{(\lambda)\prime}P^{(\lambda)})^{-1}P^{(\lambda)\prime}$. Because $P^{(\lambda)}$ is made by a basis of $R\{(\Delta X^{(\lambda)}|\Lambda\Delta X^{(\lambda)})\}$, $P^{(\lambda)}$ contains in one of its columns a multiple of $e^{(c)}$ because the latter appears in $\Delta X^{(\lambda)}$. Without loss of generality, assume that $P^{(\lambda)} = [e^{(c)}, B]$ where B is a matrix containing the rest of the columns of $P^{(\lambda)}$.

Define $P(e^{(c)}) = e^{(c)}(e^{(c)\prime}e^{(c)})e^{(c)\prime}$ and $M(e^c) = I_T - P(e^{(c)})$. Then, by the blockwise projection matrix formula it holds that

$$P^{(\lambda)}(P^{(\lambda)\prime}P^{(\lambda)})^{-1}P^{(\lambda)\prime} = P(e^{(c)}) + M(e^{(c)})B\left[B'M(e^{(c)})M(e^{(c)})B\right]^{-1}B'M(e^{(c)})'. \tag{26}$$

Notice that $P(e^{(c)}) = e^{(c)}(e^{(c)\prime}e^{(c)})e^{(c)\prime}$ is a matrix that is everywhere zero except at $\left[P(e^{(c)})\right]_{T_1+1,T_1+1} =$ 1. Thus $M(e^c)$ is equal to a matrix with its main diagonal elements equal to 1, except at $\left[M(e^{(c)})\right]_{T_1+1,T_1+1} =$ 0. The matrix $M(e^c)B$ is equal to B everywhere except at $[M(e^c)B]_{T_1+1,j} = 0$, for all j. This is a row full of zeroes and the ensuing algebra is about the impact of this row. For any matrix A, the matrix $[M(e^c)B]A[M(e^c)B]$ will have its $T_1 + 1$ row and its $T_1 + 1$ column full of zeroes.

The above arguments imply in successive order:

$$\begin{aligned}
[M(e^{(c)})B\left[B'M(e^{(c)})M(e^{(c)})B\right]^{-1}B'M(e^{(c)})']_{T_1+1} &= 0, \\
[P^{(\lambda)}(P^{(\lambda)\prime}P^{(\lambda)})^{-1}P^{(\lambda)\prime}]_{T_1+1} &= e^{(c)}, \\
[Q^{(\lambda)}]_{T_1+1} &= 0, \\
[\Lambda' Q^{(\lambda)}]_{T_1+1} &= 0, \\
[\Psi^{(\lambda)}]_{T_1+1} &= 0, \\
\Psi^{(\lambda)}e^{(c)} &= 0,
\end{aligned}$$

where 0 is the T-dimensional zero vector. The first equality comes from $[M(e^c)B]A[M(e^c)B]_{T+1} =$ 0, where $A = \left[B'M(e^{(c)})M(e^{(c)})B\right]^{-1}$. The second equality is derived from (26) where we plug in the result of the first equality and by $P(e^{(c)}) = e^{(c)}(e^{(c)\prime}e^{(c)})e^{(c)\prime}$ being a matrix that is everywhere zero except at $\left[P(e^{(c)})\right]_{T_1+1,T_1+1} = 1$. The third equality follows from the definition of $Q^{(\lambda)} = I_T - P^{(\lambda)}(P^{(\lambda)\prime}P^{(\lambda)})^{-1}P^{(\lambda)\prime}$ and the fourth equality from the fact that when multiplying the rows of a matrix (Λ' in this case) with a zero column in another matrix ($[Q^{(\lambda)}]_{T_1+1} = 0$), then the outcome is a zero column ($[\Lambda' Q^{(\lambda)}]_{T_1+1} = 0$). Finally, since $\Psi^{(\lambda)}$ is a restriction of $[\Lambda' Q^{(\lambda)}]_{T_1+1}$, $[\Psi^{(\lambda)}]_{T_1+1} = 0$. The last equality holds because the non-zero elements of the rows of $\Psi^{(\lambda)}$ multiply the zero elements of $e^{(c)}$ and the non-zero element of $e^{(c)}$ is multiplied by the elements in $[\Psi^{(\lambda)}]_{T_1+1} = 0$.

Proof of Theorem 2. This proof is divided in two steps. In the first step we prove the existence of $Z_{i,j}^{(\lambda)}$ in a model with no structural breaks. In the second step we show how the previous results extend to the case of m breaks.

Step 1: Consider a model in which $X^{(\lambda)}$ contains an intercept in the first column and ρ trend vectors in the rest of the columns so that $1 + \rho = k$. We emphasize this difference between the intercept and trend vectors because the nuisance parameters to be estimated arise from the existence of the trends and they are estimated using trend vector information. These parameters appear in Π_{22}.

The matrix $\Delta X^{(\lambda)}$ has $rank(\Delta X^{(\lambda)}) = \rho$ and consists of a vector of zeroes in its first column and ρ differenced trend vectors in the rest of the columns. Since $[\Delta X^{(\lambda)}]_1 = 0$, $\Delta X^{(\lambda)}\Pi_N \Delta X^{(\lambda)\prime}$ can be written as a linear combination of matrices resulting from the outer product of the differenced trend vectors, multiplied by the elements of Π_{22},

$$\Delta X^{(\lambda)}\Pi_N \Delta X^{(\lambda)\prime} = \sum_{i=2}^{k}\sum_{j=2}^{k}[\Pi_N]_{i,j}[\Delta X^{(\lambda)}]_i[\Delta X^{(\lambda)}]_j', \tag{27}$$

where $[\Delta X^{(\lambda)}]_i$ denotes the i-th column of $\Delta X^{(\lambda)}$. In the simple example mentioned in Section 2, where $X^{(\lambda)} = [e, \tau]$ and $\pi_i^{(\lambda)} = (\pi_{1i}, \pi_{2i})'$, we have that $\rho = 1$, $k = 2$ and $\Delta X^{(\lambda)} = [0, e]$. Thus $\Delta X^{(\lambda)}\Pi_N \Delta X^{(\lambda)\prime} = 0\Pi_{11}0' + 0\Pi_{12}e' + e\Pi_{21}0' + e\Pi_{22}e'$. So the parameters which appear in Π_{11} and Π_{12} are always multiplied by 0.

The total number of unique $[\Pi_N]_{i,j}$ elements in the above expression is $\rho(\rho+1)/2$, because Π_N is symmetric as a variance-covariance matrix. Therefore, we need to find $\rho(\rho+1)/2$ matrices $Z_{i,j}^{(\lambda)}$ that allow the estimation of these $\rho(\rho+1)/2$ nuisance parameters.

The maximum order of serial correlation is $p_{\max}^{(\lambda,H)}$ and therefore, Γ_N has $(T - p_{\max}^{(\lambda,H)} - 1)$ secondary upper diagonals which contain only 0-elements (it also has the same number of secondary lower diagonals but we will not consider them by symmetry), i.e. $\Gamma_N = [\Gamma_N]^{+}_{p_{\max}^{(\lambda,H)}}$. These upper secondary diagonals contain in total $(T - p_{\max}^{(\lambda,H)} - 1)(T - p_{\max}^{(\lambda,H)})/2$ zero elements. Consequently, the matrix $\Gamma_N + \Delta X^{(\lambda)}\Pi_N \Delta X^{(\lambda)\prime}$ contains $(T - p_{\max}^{(\lambda,H)} - 1)(T - p_{\max}^{(\lambda,H)})/2$ elements that are linear combinations only of the $\rho(\rho+1)/2$ nuisance parameters of Π_N; these elements do not contain serial correlation parameters. In other words, at these locations $[\Gamma_N + \Delta X^{(\lambda)}\Pi_N \Delta X^{(\lambda)\prime}]_{i,j} = [\Delta X^{(\lambda)}\Pi_N \Delta X^{(\lambda)\prime}]_{i,j}$ where $i = 1, ..., T - p_{\max}^{(\lambda,H)} - 1$ and $j = i + p_{\max}^{(\lambda,H)} + 1, ..., T$.

By Assumption A(ii) and under $H_{1,0}$ (a similar argument applies for $H_{2,0}$), $p_{\max}^{(\lambda,H)} = T - 2 - \rho$ and therefore, we have $(1+\rho)(2+\rho)/2$ elements of $\Gamma_N + \Delta X^{(\lambda)}\Pi_N \Delta X^{(\lambda)\prime}$ that contain $\rho(\rho+1)/2$ nuisance parameters coming from Π_N.

We proceed to estimate these nuisance parameters by the method of moments. By setting

$$\left[\frac{1}{N}\sum_{i=1}^{N}\Delta y_i \Delta y_i'\right]_{i,j} = [\Gamma_N + \Delta X^{(\lambda)}\Pi_N \Delta X^{(\lambda)\prime}]_{i,j} = [\Delta X^{(\lambda)}\Pi_N \Delta X^{(\lambda)\prime}]_{i,j} \tag{28}$$

for $i = 1, ...T - p_{\max}^{(\lambda,H)} - 1$ and $j = i + p_{\max}^{(\lambda,H)} + 1, ..., T$, we get a total of $(1+\rho)(2+\rho)/2$ moments to estimate $\rho(\rho+1)/2$ nuisance parameters. Because $(1+\rho)(2+\rho)/2 > \rho(\rho+1)/2$ for every ρ, we have always more moments than nuisance parameters. Furthermore, these moments are not linearly dependent because they are based on time series information from different periods. Therefore, we can estimate the $\rho(\rho+1)/2$ parameters by the method of moments, under Assumption A(ii).

The moments in (28) can be written in matrix form as,

$$\mathrm{vec}\left([\hat{\Gamma}]^{-}_{p_{\max}^{(\lambda,H)}}\right) = \left[\begin{array}{c} \mathrm{vec}\left[[\Delta X^{(\lambda)}]_2[\Delta X^{(\lambda)}]_2'\right]^{-}_{p_{\max}^{(\lambda,H)}}, \mathrm{vec}\left[[\Delta X^{(\lambda)}]_3[\Delta X^{(\lambda)}]_2' + [\Delta X^{(\lambda)}]_2[\Delta X^{(\lambda)}]_3'\right]^{-}_{p_{\max}^{(\lambda,H)}}, \\ , ..., \mathrm{vec}\left[[\Delta X^{(\lambda)}]_k[\Delta X^{(\lambda)}]_k'\right]^{-}_{p_{\max}^{(\lambda,H)}} \end{array}\right] \mathrm{vech}(\Pi_{22}),$$

or

$$\mathrm{vec}\left([\hat{\Gamma}]^{-}_{p_{\max}^{(\lambda,H)}}\right) = \tilde{Z}\mathrm{vech}(\Pi_{22}). \tag{29}$$

The unknown parameters appear in $\mathrm{vech}(\Pi_{22})$. This is an overdetermined system and has an approximate solution given by ordinary least squares:

$$\mathrm{vech}(\hat{\Pi}_{22}) = (\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'\mathrm{vec}\left([\hat{\Gamma}]^{-}_{p_{\max}^{(\lambda,H)}}\right). \tag{30}$$

This solution is approximate with respect to the number of moments $(1+\rho)(2+\rho)/2$ available so we cannot talk about consistency as "ρ goes to infinity". By Theorem 1 (using Assumptions A and B) and the Continuous Mapping Theorem:

$$p\lim_{N\to\infty}\left[\mathrm{vec}\left([\hat{\Gamma}]^{-}_{p_{\max}^{(\lambda,H)}}\right) - \mathrm{vec}\left(\left[\Delta X^{(\lambda)}\Pi_N\Delta X^{(\lambda)\prime}\right]^{-}_{p_{\max}^{(\lambda,H)}}\right)\right] = 0. \tag{31}$$

and

$$p\lim_{N\to\infty}\left[(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'\mathrm{vec}\left([\hat{\Gamma}]^{-}_{p_{\max}^{(\lambda,H)}}\right) - (\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'\mathrm{vec}\left(\Delta X^{(\lambda)}\Pi_N\Delta X^{(\lambda)\prime}\right)\right] = 0. \tag{32}$$

therefore, asymptotically for N, (29) is no longer overdetermined and the least squares solution in (30) is the exact solution. This happens because many sample moments $\left[\frac{1}{N}\sum_{i=1}^{N}\Delta y_i\Delta y_i'\right]_{i,j}$ converge to the same limit and asymptotically they become linearly dependent moments. In other words, the rank of the augmented matrix $(\tilde{Z}|\mathrm{vec}\left([\hat{\Gamma}]^{-}_{p_{\max}^{(\lambda,H)}}\right))$ drops and becomes equal to the number of nuisance parameters.

In (31), $\mathrm{vec}\left([\hat{\Gamma}]^{-}_{p_{\max}^{(\lambda,H)}}\right)$ converges to:

$$\begin{aligned} \mathrm{vec}\left(\left[\Delta X^{(\lambda)}\Pi_N\Delta X^{(\lambda)\prime}\right]^{-}_{p_{\max}^{(\lambda,H)}}\right) &= \mathrm{vec}\left(\sum_{i=2}^{k}\sum_{j=2}^{k}[\Pi_N]_{i,j}\left[[\Delta X^{(\lambda)}]_i[\Delta X^{(\lambda)}]_j'\right]^{-}_{p_{\max}^{(\lambda,H)}}\right) \\ &= \tilde{Z}\mathrm{vech}(\Pi_{22}). \end{aligned} \tag{33}$$

Thus, by (32) and (33):

$$\begin{aligned} p\lim_{N\to\infty}\left[\mathrm{vech}(\hat{\Pi}_{22}) - (\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'\mathrm{vec}\left(\Delta X^{(\lambda)}\Pi_N\Delta X^{(\lambda)\prime}\right)\right] &= p\lim_{N\to\infty}\left[\mathrm{vech}(\hat{\Pi}_{22}) - \mathrm{vech}(\Pi_{22})\right] \\ &= 0. \end{aligned}$$

Step 2: Suppose that there are m structural breaks. Then $X^{(\lambda)}$ has $m+1$ intercept columns and $(m+1)\rho$ trend columns with $k = (m+1) + (m+1)\rho$. First differencing the intercept columns creates "crash" vectors. These vectors are everywhere equal to zero except at the points T_j+1, $j = 1, ..., m$ where they are either equal to either 0 or 1. In total $\Delta X^{(\lambda)}$ contains $m+1$ crash vector columns and $(m+1)\rho$ differenced trends columns. This makes $\Delta X^{(\lambda)}\Pi_N\Delta X^{(\lambda)\prime}$ have a specific structure:

$$\begin{array}{c@{}c}
\begin{array}{ccccccc} \phantom{R_{T_1\times T_1}} & T_1+1 & \phantom{R^*_{T_1\times(T_2-T_1-1)}} & T_2+1 & \cdots & T_m+1 & \phantom{R_{(T-T_m-1)\times(T-T_m-1)}} \end{array} & \\
\left(\begin{array}{ccccccc}
R_{T_1\times T_1} & C_{T_1\times 1} & R^*_{T_1\times(T_2-T_1-1)} & \vdots & & & \\
C_{1\times T_1} & C_{1\times 1} & \cdots & \vdots & \cdots & \cdots & \cdots \\
 & \vdots & R_{(T_2-T_1-1)\times(T_2-T_1-1)} & \vdots & & & \\
\cdots & \vdots & \cdots & \vdots & \cdots & \cdots & \cdots \\
 & \vdots & & \vdots & \ddots & & \\
\cdots & \vdots & \cdots & \vdots & \cdots & \cdots & \cdots \\
 & \vdots & & \vdots & & & R_{(T-T_m-1)\times(T-T_m-1)}
\end{array}\right) & \begin{array}{c} \\ T_1+1 \\ \\ T_2+1 \\ \\ T_m+1 \\ \\ \end{array}
\end{array}$$

There are three types of elements here, R and R^* matrices and C vectors. The R-type and R^* matrices contain the $[(m+1)\rho][(m+1)\rho+1]/2$ Π_{22}-nuisance parameters which will be estimated as in Step 1. To see this for an R-type matrix, consider the submatrix $[\Gamma_N]_{(T_j-T_{j-1}-1)\times(T_j-T_{j-1}-1)} + R_{(T_j-T_{j-1}-1)\times(T_j-T_{j-1}-1)}$ for $j=2,...,m$. In this matrix there are $(T_j-T_{j-1}-p_{\max}^{(\lambda,H)}-2)(T_j-T_{j-1}-p_{\max}^{(\lambda,H)}-1)/2$ elements that are linear combinations of the $\rho(\rho+1)/2$ nuisance parameters of Π_{22} which are relevant to the T_j-T_{j-1} period. Without loss of generality assume that T_j-T_{j-1} is the shortest within break time period, which by Assumption C(ii) is greater or equal to $2+\rho$ and thus, by Assumption A(ii), $p_{\max}^{(\lambda,H)}=T_j-T_{j-1}-2-\rho$. Therefore, we have a total of $\rho(1+\rho)/2$ available elements. Because $\rho(1+\rho)/2=\rho(1+\rho)/2$ for every ρ, we have the necessary moments. Similar arguments hold for R^* matrices. There is no need to estimate the elements appearing in the $C-$vectors because by Lemma 2, these estimates will be multiplied by zero in the $\Theta^{(\lambda)}$ matrix. The elements of Π_{11} and Π_{12} appear in the $C-$vectors.

The above arguments can be used for other forms of the trend function. Consider the model in which the breaks happen only in the trends; $X^{(\lambda)}$ has an intercept, $(m+1)\rho$ trend columns and $k=1+(m+1)\rho$. This scenario is simpler than the one above as now there are no "crash" vectors but $[\Delta X^{(\lambda)}]_1=0$. The effect of $[\Delta X^{(\lambda)}]_1$ appears only in the $C-$vectors which are multiplied by zero in the $\Theta^{(\lambda)}$ matrix. Another case is when the m breaks appear in the intercept and not in the trend. In this case $X^{(\lambda)}$ has $m+1$ intercept columns and ρ trend columns. $\Delta X^{(\lambda)}\Pi_N\Delta X^{(\lambda)\prime}$ will have the $C-$vectors as displayed above but the R and R^* matrices will contain $\rho(\rho+1)/2$ Π_{22}-nuisance parameters which are less than $[(m+1)\rho][(m+1)\rho+1]/2$ and therefore can be estimated by the available moments.

For the proof of Theorem 3 we need the following two lemmas:

Lemma 3: Let A be any $\nu\times\nu$ real matrix and B a $T^2\times\nu$ matrix such that $[B]_i=\text{vec}([J_i]_p^-)$, where J_i for $i=1,...,\nu$ are real $T\times T$ matrices. Then

$$\text{vec}_{T,T}^{-1}([BA]_i)=\left[\text{vec}_{T,T}^{-1}([BA]_i)\right]_p^-.$$

Proof of Lemma 3: The $[J_i]_p^-$ for $i=1,...,\nu$ matrices by definition have their main, the first p upper and first p lower diagonals equal to zero. Thus $[B]_i=\text{vec}([J_i]_p^-)$ is a $T^2\times 1$ vector that has zeroes in the following in the following elements:

$$\begin{aligned}
&1, ..., 1+p, \\
&\vdots \\
&(\kappa-1)T+\kappa-p, (\kappa-1)T+\kappa-p+1, ..., (\kappa-1)T+\kappa+p, \\
&\vdots \\
&(\nu-1)T+\nu-p, ..., (\nu-1)T+\nu,
\end{aligned} \tag{34}$$

where $\kappa = 2, ..., \nu-1$. Since this applies for every column $[B]_i$, these rows of B are equal to zero, i.e. $[B]_{i,j} = 0$ when i takes one of the values in (34) and for every $j = 1, ..., \nu$.

The matrix BA then has zero elements at the places where one of the zero-rows of B multiplies the columns of A. Because a zero-row of B multiplies every column of A, the outcome is a zero-row in BA :

$$[BA]_{i,j} = 0 \text{ for every } j, \text{ if } i \text{ takes one of the values in (34).}$$

Therefore BA and B have the same zero-rows and the vectors $[B]_i$ and $[BA]_i$ have zeroes at the same elements. Thus $\text{vec}^{-1}_{T,T}([BA]_i)$ has zero elements in its main, first p upper and first p lower diagonals and thus

$$\text{vec}^{-1}_{T,T}([BA]_i) = \left[\text{vec}^{-1}_{T,T}([BA]_i)\right]^-_p .$$

Lemma 4: Let A and B be any $T \times T$ matrices. Then it holds that $\text{tr}([A]^+_p[B]^-_p) = 0$.

Proof of Lemma 4: We will calculate the main diagonal elements of $[A]^+_p[B]^-_p$. By definition $\left[[A]^+_p\right]_{1,j} = 0$ if $j > p+1$ and $\left[[B]^-_p\right]_{i,1} = 0$ if $i < p+1$. Thus multiplying the first row of $[A]^+_p$ with the first column of $[B]^-_p$ the result is 0, thus $\left[[A]^+_p[B]^-_p\right]_{1,1} = 0$. Applying the same argument for the rest of the rows and columns leads to $\left[[A]^+_p[B]^-_p\right]_{j,j} = 0$ for $j = 2, ..., T$. Because the $[A]^+_p[B]^-_p$ is a matrix with a main diagonal made of zeroes, then $\text{tr}([A]^+_p[B]^-_p) = 0$.

Proof of Theorem 3. We first proceed to show that

$$\text{tr}(\Theta^{(\lambda)}\Delta X^{(\lambda)}\Pi_N\Delta X^{(\lambda)\prime}) = 0 \text{ and} \tag{35}$$

$$\text{tr}(\Theta^{(\lambda)}\Gamma_N) = \text{tr}(\Lambda' Q^{(\lambda)}\Gamma_N). \tag{36}$$

Assumption C(iii) states that $Q^{(\lambda)} \neq 0$ and thus the proof of (35) is not trivial. The expression (27) for the case of m breaks becomes:

$$\Delta X^{(\lambda)}\Pi_N\Delta X^{(\lambda)\prime} = \sum_{i=1}^{k}\sum_{j=1}^{k}[\Pi_N]_{i,j}[\Delta X^{(\lambda)}]_i[\Delta X^{(\lambda)}]'_j, \tag{37}$$

For (35), substituting $\Theta^{(\lambda)}$:

$$\begin{aligned}
\operatorname{tr}(\Theta^{(\lambda)}\Delta X^{(\lambda)}\Pi_N\Delta X^{(\lambda)\prime}) &= \operatorname{tr}\left[\begin{array}{c} \Psi^{(\lambda)}\Delta X^{(\lambda)}\Pi_N\Delta X^{(\lambda)\prime}- \\ -\sum_{i=m+2}^{k}\sum_{j=m+2}^{k}\left[\operatorname{tr}(\Psi^{(\lambda)}[\Delta X^{(\lambda)}]_i[\Delta X^{(\lambda)}]_j')Z_{i,j}^{(\lambda)}\Delta X^{(\lambda)}\Pi_N\Delta X^{(\lambda)\prime}\right] \end{array}\right] \\
&= \operatorname{tr}\left[\Psi^{(\lambda)}\Delta X^{(\lambda)}\Pi_N\Delta X^{(\lambda)\prime}\right]- \\
&\quad -\sum_{i=m+2}^{k}\sum_{j=m+2}^{k}\operatorname{tr}(\Psi^{(\lambda)}[\Delta X^{(\lambda)}]_i[\Delta X^{(\lambda)}]_j')\operatorname{tr}\left(Z_{i,j}^{(\lambda)}\Delta X^{(\lambda)}\Pi_N\Delta X^{(\lambda)\prime}\right) \\
&= \bar{I}-\overline{II}.
\end{aligned}$$

By substituting (37), $\bar{I}$ becomes

$$\begin{aligned}
\bar{I} &= \operatorname{tr}\left[\Psi^{(\lambda)}\Delta X^{(\lambda)}\Pi_N\Delta X^{(\lambda)\prime}\right] \\
&= \sum_{i=1}^{k}\sum_{j=1}^{k}[\Pi_N]_{i,j}\operatorname{tr}\left(\Psi^{(\lambda)}[\Delta X^{(\lambda)}]_i[\Delta X^{(\lambda)}]_j'\right) \\
&= \sum_{i=m+2}^{k}\sum_{j=m+2}^{k}[\Pi_N]_{i,j}\operatorname{tr}\left(\Psi^{(\lambda)}[\Delta X^{(\lambda)}]_i[\Delta X^{(\lambda)}]_j'\right). \qquad (38)
\end{aligned}$$

The last equality holds by Lemma 2 because $\Psi^{(\lambda)}$ removes the "crash" vectors and therefore we are only left with the differenced trend vectors of $\Delta X^{(\lambda)}$.

$$\begin{aligned}
\overline{II} &= \sum_{i=m+2}^{k}\sum_{j=m+2}^{k}\operatorname{tr}(\Psi^{(\lambda)}[\Delta X^{(\lambda)}]_i[\Delta X^{(\lambda)}]_j')\operatorname{tr}\left(Z_{i,j}^{(\lambda)}\Delta X^{(\lambda)}\Pi_N\Delta X^{(\lambda)\prime}\right) \\
&= \sum_{i=m+2}^{k}\sum_{j=m+2}^{k}[\Pi_N]_{i,j}\operatorname{tr}(\Psi^{(\lambda)}[\Delta X^{(\lambda)}]_i[\Delta X^{(\lambda)}]_j'). \qquad (39)
\end{aligned}$$

The last equality follows because

$$\begin{aligned}
\operatorname{tr}\left(Z_{i,j}^{(\lambda)}\Delta X^{(\lambda)}\Pi_N\Delta X^{(\lambda)\prime}\right) &= \operatorname{vec}(Z_{i,j}^{(\lambda)})'\operatorname{vec}(\Delta X^{(\lambda)}\Pi_N\Delta X^{(\lambda)\prime}) \\
&= \operatorname{vec}(Z_{i,j}^{(\lambda)})'\tilde{Z}\operatorname{vech}(\Pi_{22})
\end{aligned}$$

by (33). Since $\operatorname{vec}(Z_{i,j}^{(\lambda)}) = \left[\tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\right]_{i+\frac{(j-1)j}{2}}$, $\operatorname{vec}(Z_{i,j}^{(\lambda)})'$ is the $i+(j-1)j/2$-th row of $\tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}$. Then, given that $(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'\tilde{Z} = I_T$, it holds that $\operatorname{vec}(Z_{i,j}^{(\lambda)})'[\tilde{Z}]_{i+\frac{(j-1)j}{2}} = 1$ and $\operatorname{vec}(Z_{i,j}^{(\lambda)})'[\tilde{Z}]_\nu = 0$ for $\nu \neq i+(j-1)j/2$. The proof of (35) comes by subtracting (39) from (38).

For (36),

$$\begin{aligned}
\operatorname{tr}(\Theta^{(\lambda)}\Gamma_N) &= \operatorname{tr}\left[\Psi^{(\lambda)}\Gamma_N-\sum_{i=m+2}^{k}\sum_{j=m+2}^{k}\left[\operatorname{tr}(\Psi^{(\lambda)}[\Delta X^{(\lambda)}]_i[\Delta X^{(\lambda)}]_j')Z_{i,j}^{(\lambda)}\Gamma_N\right]\right] \\
&= \operatorname{tr}\left(\Psi^{(\lambda)}\Gamma_N\right)-\sum_{i=m+2}^{k}\sum_{j=m+2}^{k}\operatorname{tr}(\Psi^{(\lambda)}[\Delta X^{(\lambda)}]_i[\Delta X^{(\lambda)}]_j')\operatorname{tr}\left(Z_{i,j}^{(\lambda)}\Gamma_N\right) \\
&= \operatorname{tr}\left(\Psi^{(\lambda)}\Gamma_N\right) \\
&= \operatorname{tr}(\Lambda'Q^{(\lambda)}\Gamma_N).
\end{aligned}$$

because $\operatorname{tr}\left(Z_{i,j}^{(\lambda)}\Gamma_N\right) = 0$. To see this, by definition $Z_{i,j}^{(\lambda)} = \operatorname{vec}_{T,T}^{-1}\left\{\left[\tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\right]_{i+\frac{(j-1)j}{2}}\right\}$. Also by definition, $\tilde{Z}$ has columns of the form $\operatorname{vec}\left([\cdot]^{-}_{p_{\max}^{(\lambda,H)}}\right)$. Therefore, by Lemma 3, $Z_{i,j}^{(\lambda)} = [Z_{i,j}^{(\lambda)}]^{-}_{p_{\max}^{(\lambda,H)}}$. Then since, $\Gamma_N = [\Gamma_N]^{+}_{p_{\max}^{(\lambda,H)}}$ by Lemma 4 $\operatorname{tr}(Z_{i,j}^{(\lambda)}\Gamma_N) = \operatorname{tr}(\Gamma_N Z_{i,j}^{(\lambda)}) = \operatorname{tr}([\Gamma_N]^{+}_{p_{\max}^{(\lambda,H)}}[Z_{i,j}^{(\lambda)}]^{-}_{p_{\max}^{(\lambda,H)}}) = 0$.

Overall, from Theorem 1 (using Assumptions A and B) and the Continuous Mapping Theorem,

$$\operatorname{tr}(\Theta^{(\lambda)}\hat{\Gamma}) - \operatorname{tr}(\Theta^{(\lambda)}(\Delta X^{(\lambda)}\Pi_N\Delta X^{(\lambda)\prime} + \Gamma_N)) \xrightarrow{p} 0.$$

Proof of Theorem 4. Start with

$$\begin{aligned}
\hat{d}^{(\lambda)}\left(\hat{\varphi}_{\text{DME}}^{(\lambda)} - 1\right) &= \hat{d}^{(\lambda)}\left(\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{d}^{(\lambda)}} - 1\right) \\
&= \left(\frac{1}{N}\sum_{i=1}^{N} y_{i,-1}'Q^{(\lambda)}y_{i,-1}\right)\left(\frac{\frac{1}{N}\sum_{i=1}^{N} y_{i,-1}'Q^{(\lambda)}y_i}{\frac{1}{N}\sum_{i=1}^{N} y_{i,-1}'Q^{(\lambda)}y_{i,-1}} - \frac{\operatorname{tr}(\Theta^{(\lambda)}\hat{\Gamma})}{\frac{1}{N}\sum_{i=1}^{N} y_{i,-1}'Q^{(\lambda)}y_{i,-1}} - 1\right) \\
&= \left(\frac{1}{N}\sum_{i=1}^{N} y_{i,-1}'Q^{(\lambda)}y_{i,-1}\right)\left(\frac{\frac{1}{N}\sum_{i=1}^{N} u_i'\Lambda'Q^{(\lambda)}u_i - \operatorname{tr}(\Theta^{(\lambda)}\hat{\Gamma})}{\frac{1}{N}\sum_{i=1}^{N} y_{i,-1}'Q^{(\lambda)}y_{i,-1}}\right) \\
&= \frac{1}{N}\sum_{i=1}^{N} u_i'\Lambda'Q^{(\lambda)}u_i - \operatorname{tr}(\Theta^{(\lambda)}\hat{\Gamma}) \\
&= \frac{1}{N}\sum_{i=1}^{N} \Delta y_i'\Lambda'Q^{(\lambda)}\Delta y_i - \operatorname{tr}(\Theta^{(\lambda)}\hat{\Gamma}) \\
&= \frac{1}{N}\sum_{i=1}^{N} \Delta y_i'(\Lambda'Q^{(\lambda)} - \Theta^{(\lambda)})\Delta y_i. \qquad (40)
\end{aligned}$$

where $u_i'\Lambda'Q^{(\lambda)}u_i = \Delta y_i'\Lambda'Q^{(\lambda)}\Delta y_i$ because $Q^{(\lambda)}\Delta X^{(\lambda)} = 0$ and $\Delta X^{(\lambda)\prime}\Lambda'Q^{(\lambda)} = 0$ because of the way $Q^{(\lambda)}$ is constructed. Then, since $E(\pi_i^{(\lambda)}u_i) = 0$, by Assumption B(i),

$$\begin{aligned}
E\left[\Delta y_i'(\Lambda'Q^{(\lambda)} - \Theta^{(\lambda)})\Delta y_i\right] &= E\left[u_i'(\Lambda'Q^{(\lambda)} - \Theta^{(\lambda)})u_i + \pi_i^{(\lambda)\prime}\Delta X^{(\lambda)\prime}(\Lambda'Q^{(\lambda)} - \Theta^{(\lambda)})\Delta X^{(\lambda)}\pi_i^{(\lambda)}\right] \\
&= \operatorname{tr}\left[(\Lambda'Q^{(\lambda)} - \Theta^{(\lambda)})\Gamma_i\right] + \operatorname{tr}\left[(\Lambda'Q^{(\lambda)} - \Theta^{(\lambda)})\Delta X^{(\lambda)}\Pi_N\Delta X^{(\lambda)\prime}\right].
\end{aligned}$$

But

$$\operatorname{tr}\left[(\Lambda'Q^{(\lambda)} - \Theta^{(\lambda)})\Gamma_i\right] = \operatorname{tr}\left\{\left[\Lambda'Q^{(\lambda)} - \Theta^{(\lambda)}\right]^{-}_{p_{\max}^{(\lambda,H)}}[\Gamma_i]^{+}_{p_{\max}^{(\lambda,H)}}\right\} = 0, \qquad (41)$$

by Lemma 4. Notice that $\left(\Lambda'Q^{(\lambda)} - \Theta^{(\lambda)}\right) = \left[\Lambda'Q^{(\lambda)} - \Theta^{(\lambda)}\right]^{-}_{p_{\max}^{(\lambda,H)}}$ as,

$$\begin{aligned}
\Lambda'Q^{(\lambda)} - \Theta^{(\lambda)} &= \Lambda'Q^{(\lambda)} - \Psi^{(\lambda)} + \sum_{i=m+2}^{k}\sum_{j=m+2}^{k}\left[\operatorname{tr}(\Psi^{(\lambda)}[\Delta X^{(\lambda)}]_i[\Delta X^{(\lambda)}]_j')Z_{i,j}^{(\lambda)}\right] \qquad (42) \\
&= \Lambda'Q^{(\lambda)} - \left[\Lambda'Q^{(\lambda)}\right]^{+}_{p_{\max}^{(\lambda,H)}} + \sum_{i=m+2}^{k}\sum_{j=m+2}^{k}\left[\operatorname{tr}(\Psi^{(\lambda)}[\Delta X^{(\lambda)}]_i[\Delta X^{(\lambda)}]_j')Z_{i,j}^{(\lambda)}\right] \\
&= \left[\Lambda'Q^{(\lambda)}\right]^{-}_{p_{\max}^{(\lambda,H)}} + \left[\sum_{i=m+2}^{k}\sum_{j=m+2}^{k}\operatorname{tr}(\Psi^{(\lambda)}[\Delta X^{(\lambda)}]_i[\Delta X^{(\lambda)}]_j')Z_{i,j}^{(\lambda)}\right]^{-}_{p_{\max}^{(\lambda,H)}} \\
&= \left[\Lambda'Q^{(\lambda)} - \Theta^{(\lambda)}\right]^{-}_{p_{\max}^{(\lambda,H)}},
\end{aligned}$$

because the $Z_{i,j}^{(\lambda)} = \left[Z_{i,j}^{(\lambda)}\right]^{-}_{p_{\max}^{(\lambda,H)}}$ by Lemma 3. Furthermore,

$$\begin{aligned}\mathrm{tr}\left[(\Lambda' Q^{(\lambda)} - \Theta^{(\lambda)})\Delta X^{(\lambda)}\Pi_N \Delta X^{(\lambda)\prime}\right] &= \mathrm{tr}\left[\Lambda' Q^{(\lambda)}\Delta X^{(\lambda)}\Pi_N \Delta X^{(\lambda)\prime}\right] - \mathrm{tr}\left[\Theta^{(\lambda)}\Delta X^{(\lambda)}\Pi_N \Delta X^{(\lambda)\prime}\right] \\ &= 0,\end{aligned}$$

by the construction of $Q^{(\lambda)}$ and by (35). Therefore,

$$E\left[\Delta y_i'(\Lambda' Q^{(\lambda)} - \Theta^{(\lambda)})\Delta y_i\right] = 0.$$

Its variance is given by,

$$\begin{aligned}V\left[\Delta y_i'(\Lambda' Q^{(\lambda)} - \Theta^{(\lambda)})\Delta y_i\right] &= V\left\{\mathrm{tr}\left[(\Lambda' Q^{(\lambda)} - \Theta^{(\lambda)})\Delta y_i \Delta y_i'\right]\right\} \\ &= V\left[\mathrm{vec}(Q^{(\lambda)}\Lambda - \Theta^{(\lambda)\prime})'\mathrm{vec}(\Delta y_i \Delta y_i')\right] \\ &= \mathrm{vec}(Q^{(\lambda)}\Lambda - \Theta^{(\lambda)\prime})' V\left[\mathrm{vec}(\Delta y_i \Delta y_i')\right]\mathrm{vec}(Q^{(\lambda)}\Lambda - \Theta^{(\lambda)\prime}).\end{aligned}$$

Denote $V_i^{(\lambda)} = \mathrm{vec}(Q^{(\lambda)}\Lambda - \Theta^{(\lambda)\prime})' V\left[\mathrm{vec}(\Delta y_i \Delta y_i')\right]\mathrm{vec}(Q^{(\lambda)}\Lambda - \Theta^{(\lambda)\prime})$. Then by the Lindeberg-Feller CLT (see e.g. Feller (1968), p.254), under Assumptions A and B,

$$\frac{\sum_{i=1}^{N}\Delta y_i'(\Lambda' Q^{(\lambda)} - \Theta^{(\lambda)})\Delta y_i}{\sqrt{\sum_{i=1}^{N} V_i^{(\lambda)}}} \xrightarrow{d} N(0,1). \tag{43}$$

By similar arguments it is straightforward to see that $\hat{V}^{(\lambda)} - (1/N)\sum_{i=1}^{N} V_i^{(\lambda)} \xrightarrow{p} 0$. This is an estimator of fourth moments of the u_i and requires a uniform bound on the 8th moments. Note that the fourth order individual effects are removed by the $\mathrm{vec}(Q^{(\lambda)}\Lambda - \Theta^{(\lambda)\prime})'$s and thus there is no need to make 8th order moment assumptions on them. To see this, in

$$\hat{V}^{(\lambda)} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{vec}(Q^{(\lambda)}\Lambda - \Theta^{(\lambda)\prime})'\left[\mathrm{vec}(\Delta y_i \Delta y_i')\mathrm{vec}(\Delta y_i \Delta y_i')'\right]\mathrm{vec}(Q^{(\lambda)}\Lambda - \Theta^{(\lambda)\prime}),$$

the vectorization in the middle is equal to

$$\mathrm{vec}(\Delta y_i \Delta y_i') = \mathrm{vec}(u_i u_i') + \mathrm{vec}(u_i \pi_i^{(\lambda)\prime}\Delta X^{(\lambda)\prime}) + \mathrm{vec}(\Delta X^{(\lambda)}\pi_i^{(\lambda)} u_i') + \mathrm{vec}(\Delta X^{(\lambda)}\pi_i^{(\lambda)}\pi_i^{(\lambda)\prime}\Delta X^{(\lambda)\prime}).$$

Thus $\mathrm{vec}(\Delta y_i \Delta y_i')\mathrm{vec}(\Delta y_i \Delta y_i')'$ is a function of (to save space the transposed matrices are omitted):

$$\begin{aligned}(i) &: \mathrm{vec}(u_i u_i')\mathrm{vec}(u_i u_i')' \\ (ii) &: \mathrm{vec}(\Delta X^{(\lambda)}\pi_i^{(\lambda)}\pi_i^{(\lambda)\prime}\Delta X^{(\lambda)\prime})\mathrm{vec}(\Delta X^{(\lambda)}\pi_i^{(\lambda)}\pi_i^{(\lambda)\prime}\Delta X^{(\lambda)\prime})' \\ (iii) &: \mathrm{vec}(\Delta X^{(\lambda)}\pi_i^{(\lambda)} u_i')\mathrm{vec}(\Delta X^{(\lambda)}\pi_i^{(\lambda)} u_i')' \\ (iv) &: \mathrm{vec}(\Delta X^{(\lambda)}\pi_i^{(\lambda)} u_i')\mathrm{vec}(\Delta X^{(\lambda)}\pi_i^{(\lambda)}\pi_i^{(\lambda)\prime}\Delta X^{(\lambda)\prime})' \\ (v) &: \mathrm{vec}(u_i u_i')\mathrm{vec}(u_i \pi_i^{(\lambda)\prime}\Delta X^{(\lambda)\prime})'.\end{aligned}$$

The matrix $(i) : \mathrm{vec}(u_i u_i')\mathrm{vec}(u_i u_i')'$ contains as elements cross products of the errors of fourth order, and the squared moments of these elements are bounded by Assumption A(iii) (this can be seen by applying the Cauchy-Schwarz inequality). The matrices (iii) and (iv) contain cross products in which the elements of $\pi_i^{(\lambda)}$ appear at most in the power of two and therefore their squared moments are bounded by Assumption B(ii). The matrices (ii) and (iv) contain elements of $\pi_i^{(\lambda)}$ in higher powers but are equal to zero when multiplied by $\text{vec}(Q^{(\lambda)}\Lambda - \Theta^{(\lambda)\prime})'$ and thus there is no need to make higher order assumptions on the moments of $\pi_i^{(\lambda)}$:

$$\begin{aligned}
\text{vec}(Q^{(\lambda)}\Lambda - \Theta^{(\lambda)\prime})'\text{vec}(\Delta X^{(\lambda)}\pi_i^{(\lambda)}\pi_i^{(\lambda)\prime}\Delta X^{(\lambda)\prime}) &= \text{tr}\left[\left(\Lambda' Q^{(\lambda)} - \Theta^{(\lambda)}\right)\left(\Delta X^{(\lambda)}\pi_i^{(\lambda)}\pi_i^{(\lambda)\prime}\Delta X^{(\lambda)\prime}\right)\right] \\
&= 0
\end{aligned}$$

by (35). Notice that (35) applies for any matrix Π_N as we have not made any assumptions regarding its structure, and thus for $\pi_i^{(\lambda)}\pi_i^{(\lambda)\prime}$.

Proof of Theorem 5. The distribution G is derived by combining the result of Theorem 4 and the Continuous Mapping Theorem. We now proceed to derive the formula for Σ. First, notice that under $H_{2,0} : \varphi = 1$ and $\pi_{j,i}^{(1)} = \pi_{j,i}^{(2)} = ... = \pi_{j,i}^{(m+1)}$, for all j. Therefore, there are no structural breaks and thus $\Delta X^{(\lambda)} \equiv \Delta X$ and $Q^{(\mu)}\Delta X = Q^{(\mu)}\Lambda\Delta X = 0$ for every partition μ. This holds because the column space of $R\{(\Delta X|\Lambda\Delta X)\}$ is a subset of the column space of $R\{(\Delta X^{(\lambda)}|\Lambda\Delta X^{(\lambda)})\}$, i.e., if $\Delta X^{(\lambda)} = \left[\Delta e^{(1)}, \Delta e^{(2)}, \Delta\tau^{(1)}, \Delta\tau^{(2)}\right]$, $\Delta X = \left[\Delta e^{(1)} + \Delta e^{(2)}, \Delta\tau^{(1)} + \Delta\tau^{(2)}\right]$. Then, using the same algebra as in the previous theorem

$$\begin{aligned}
t^{(\mu)}t^{(\nu)} &= \frac{\left(\hat{\varphi}_{\text{DME}}^{(\mu)} - 1\right)}{\sqrt{\hat{V}^{(\mu)}/(N\hat{d}^{(\mu)2})}}\frac{\left(\hat{\varphi}_{\text{DME}}^{(\nu)} - 1\right)}{\sqrt{\hat{V}^{(\nu)}/(N\hat{d}^{(\nu)2})}} \\
&= \frac{1}{\frac{1}{N}\sqrt{\hat{V}^{(\mu)}}\sqrt{\hat{V}^{(\nu)}}}\hat{d}^{(\mu)}\left(\hat{\varphi}_{\text{DME}}^{(\mu)} - 1\right)\hat{d}^{(\nu)}\left(\hat{\varphi}_{\text{DME}}^{(\nu)} - 1\right) \\
&= \frac{1}{\frac{1}{N}\sqrt{\hat{V}^{(\mu)}}\sqrt{\hat{V}^{(\nu)}}}\left(\frac{1}{N}\sum_{i=1}^{N}\Delta y_i'(\Lambda' Q^{(\mu)} - \Theta^{(\mu)})\Delta y_i\right)\left(\frac{1}{N}\sum_{i=1}^{N}\Delta y_i'(\Lambda' Q^{(\nu)} - \Theta^{(\nu)})\Delta y_i\right) \\
&= \frac{\frac{1}{N}\left(\sum_{i=1}^{N}\Delta y_i'(\Lambda' Q^{(\mu)} - \Theta^{(\mu)})\Delta y_i\right)\left(\sum_{i=1}^{N}\Delta y_i'(\Lambda' Q^{(\nu)} - \Theta^{(\nu)})\Delta y_i\right)}{\sqrt{\hat{V}^{(\mu)}}\sqrt{\hat{V}^{(\nu)}}} \\
&= \frac{\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\Delta y_i'(\Lambda' Q^{(\mu)} - \Theta^{(\mu)})\Delta y_i\Delta y_j'(\Lambda' Q^{(\nu)} - \Theta^{(\nu)})\Delta y_j}{\sqrt{\hat{V}^{(\mu)}}\sqrt{\hat{V}^{(\nu)}}}.
\end{aligned}$$

By Assumption A(i), (41), (42) and Lemma A.1. in Kelejian and Prucha (2010),

$$\begin{aligned}
E\left[\Delta y_i'(\Lambda' Q^{(\mu)} - \Theta^{(\mu)})\Delta y_i\Delta y_j'(\Lambda' Q^{(\nu)} - \Theta^{(\nu)})\Delta y_j\right] &= F^{(\mu)}\Xi F^{(\nu)} \text{ if } i = j, \\
E\left[\Delta y_i'(\Lambda' Q^{(\mu)} - \Theta^{(\mu)})\Delta y_i\Delta y_j'(\Lambda' Q^{(\nu)} - \Theta^{(\nu)})\Delta y_j\right] &= 0 \text{ if } i \neq j.
\end{aligned}$$

Therefore,

$$p\lim_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\Delta y_i'(\Lambda' Q^{(\mu)} - \Theta^{(\mu)})\Delta y_i\Delta y_j'(\Lambda' Q^{(\nu)} - \Theta^{(\nu)})\Delta y_j - F^{(\mu)}\Xi F^{(\nu)}\right] = 0.$$

and by Slutcky's Theorem:

$$p\lim_{N\to\infty}\left[\frac{\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\Delta y_i'(\Lambda' Q^{(\mu)} - \Theta^{(\mu)})\Delta y_i\Delta y_j'(\Lambda' Q^{(\nu)} - \Theta^{(\nu)})\Delta y_j}{\sqrt{\hat{V}^{(\mu)}}\sqrt{\hat{V}^{(\nu)}}} - \frac{F^{(\mu)}\Xi F^{(\nu)}}{\sqrt{F^{(\mu)}\Xi F^{(\mu)}}\sqrt{F^{(\nu)}\Xi F^{(\nu)}}}\right] = 0.$$

Lemma 5: $p\lim_{N\to\infty}\left[F^{(\lambda)\prime}\hat{\Xi}F^{(\lambda)} - F^{(\lambda)\prime}\hat{\Xi}^{*}F^{(\lambda)}\right] = 0.$

Table 1: Crash and Changing Growth Model with one break

		Scenario 1		Scenario 2		Scenario 3		Scenario 4		Scenario 5	
T	N	$H_{1,0}$	$H_{1,1}$	$H_{1,0}$	$H_{1,1}$	$H_{1,0}$	$H_{1,1}$	$H_{1,0}$	$H_{1,1}$	$H_{2,0}$	$H_{2,1}$
10	25	0.053	0.109	0.052	0.068	0.057	0.078	0.050	0.060	0.044	0.087
	50	0.053	0.144	0.047	0.075	0.055	0.084	0.053	0.095	0.048	0.092
	100	0.060	0.185	0.051	0.093	0.048	0.107	0.058	0.115	0.057	0.146
	500	0.044	0.466	0.050	0.167	0.058	0.222	0.049	0.276	0.051	0.387
	1000	0.059	0.702	0.045	0.226	0.051	0.360	0.051	0.376	0.057	0.608
	1200	0.054	0.777	0.046	0.276	0.042	0.396	0.058	0.442	0.052	0.651
20	25	0.067	0.207	0.055	0.134	0.069	0.194	0.062	0.203	0.047	0.171
	50	0.063	0.275	0.054	0.159	0.065	0.269	0.060	0.285	0.044	0.243
	100	0.061	0.392	0.059	0.241	0.060	0.410	0.053	0.417	0.046	0.429
	500	0.053	0.927	0.052	0.642	0.059	0.927	0.060	0.819	0.061	0.927
	1000	0.051	0.997	0.051	0.884	0.048	0.997	0.059	0.912	0.058	0.997
	1200	0.052	0.999	0.049	0.935	0.053	0.999	0.059	0.936	0.056	0.999
30	25	0.064	0.218	0.064	0.132	0.060	0.238	0.061	0.234	0.054	0.173
	50	0.072	0.304	0.041	0.169	0.056	0.355	0.059	0.374	0.047	0.270
	100	0.061	0.468	0.057	0.258	0.051	0.558	0.060	0.528	0.049	0.495
	500	0.064	0.961	0.041	0.697	0.048	0.990	0.055	0.910	0.058	0.881
	1000	0.061	0.999	0.047	0.929	0.062	1.000	0.056	0.979	0.063	0.956
	1200	0.057	1.000	0.047	0.961	0.057	1.000	0.047	0.989	0.057	1.000

Proof of Lemma 5: Substituting $\hat{\Xi}$ and $\hat{\Xi}^*$:

$$F^{(\lambda)\prime}\hat{\Xi}F^{(\lambda)} - F^{(\lambda)\prime}\hat{\Xi}^*F^{(\lambda)} = \left[F^{(\lambda)\prime}\frac{1}{N}\sum_{i=1}^{N}\mathrm{vec}(\Delta y_i \Delta y_i')\right]\left[\frac{1}{N}\sum_{i=1}^{N}\mathrm{vec}(\Delta y_i \Delta y_i')'F^{(\lambda)}\right].$$

Now, consider

$$\begin{aligned}\left[F^{(\lambda)\prime}\frac{1}{N}\sum_{i=1}^{N}\mathrm{vec}(\Delta y_i \Delta y_i')\right] &= \frac{1}{N}\sum_{i=1}^{N}\mathrm{vec}(Q^{(\lambda)}\Lambda - \Theta^{(\lambda)\prime})'\mathrm{vec}(\Delta y_i \Delta y_i') \\ &= \frac{1}{N}\sum_{i=1}^{N}\mathrm{tr}\left[(\Lambda' Q^{(\lambda)} - \Theta^{(\lambda)})\Delta y_i \Delta y_i'\right] \\ &= \mathrm{tr}\left[(\Lambda' Q^{(\lambda)} - \Theta^{(\lambda)})\frac{1}{N}\sum_{i=1}^{N}\Delta y_i \Delta y_i'\right].\end{aligned}$$

From Theorem 1:

$$\frac{1}{N}\sum_{i=1}^{N}\Delta y_i \Delta y_i' - \Delta X^{(\lambda)}\Pi_N \Delta X^{(\lambda)\prime} - \Gamma_N \xrightarrow{p} 0.$$

and thus

$$p\lim_{N\to\infty}\mathrm{tr}\left[(\Lambda' Q^{(\lambda)} - \Theta^{(\lambda)})\frac{1}{N}\sum_{i=1}^{N}\Delta y_i \Delta y_i'\right] = 0$$

by $Q^{(\lambda)}\Delta X^{(\lambda)} = 0$, (35) and (36).

Table 2: Crash and Changing Growth Model with two breaks

		Scenario 1		Scenario 2		Scenario 3		Scenario 4		Scenario 5	
T	N	$H_{1,0}$	$H_{1,1}$	$H_{1,0}$	$H_{1,1}$	$H_{1,0}$	$H_{1,1}$	$H_{1,0}$	$H_{1,1}$	$H_{2,0}$	$H_{2,1}$
10	25	0.067	0.110	0.040	0.062	0.050	0.082	0.061	0.075	0.065	0.088
	50	0.052	0.143	0.058	0.078	0.071	0.100	0.051	0.098	0.052	0.116
	100	0.062	0.175	0.044	0.080	0.064	0.123	0.058	0.124	0.051	0.143
	500	0.052	0.465	0.057	0.176	0.058	0.238	0.050	0.286	0.055	0.355
	1000	0.049	0.758	0.059	0.236	0.058	0.404	0.050	0.412	0.062	0.572
	1200	0.057	0.812	0.058	0.291	0.051	0.488	0.052	0.426	0.049	0.671
20	25	0.058	0.173	0.048	0.085	0.061	0.117	0.060	0.117	0.053	0.153
	50	0.063	0.200	0.050	0.111	0.056	0.164	0.049	0.173	0.048	0.199
	100	0.060	0.298	0.046	0.132	0.050	0.214	0.049	0.242	0.052	0.337
	500	0.049	0.792	0.052	0.344	0.058	0.616	0.063	0.597	0.062	0.775
	1000	0.056	0.961	0.045	0.497	0.052	0.868	0.050	0.722	0.063	0.944
	1200	0.061	0.983	0.046	0.590	0.052	0.906	0.052	0.770	0.051	0.962
30	25	0.058	0.212	0.054	0.126	0.056	0.202	0.054	0.231	0.055	0.179
	50	0.054	0.316	0.056	0.153	0.048	0.310	0.056	0.294	0.052	0.280
	100	0.052	0.477	0.046	0.229	0.051	0.443	0.055	0.462	0.059	0.478
	500	0.051	0.965	0.049	0.642	0.061	0.955	0.048	0.819	0.050	0.861
	1000	0.060	0.999	0.055	0.900	0.059	0.999	0.051	0.920	0.059	0.969
	1200	0.052	1.000	0.052	0.931	0.048	1.000	0.054	0.949	0.063	0.987