Local Power of Panel Unit Root Tests Allowing for Structural Breaks

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

Abstract

The asymptotic local power of least squares based fixed-T panel unit root tests allowing for a structural break in their individual effects and/or incidental trends of the $AR(1)$ panel data model is studied. Limiting distributions of these tests are derived under a sequence of local alternatives and analytic expressions show how their means and variances are functions of the break date and the time dimension of the panel. The considered tests have non-trivial local power in a $N^{-1/2}$ neighborhood of unity when the panel data model includes individual intercepts. For panel data models with incidental trends, the power of the tests becomes trivial in this neighborhood. However, this problem does not always appear if the tests allow for serial correlation in the error term and completely vanishes in the presence of cross section correlation. These results show that fixed-T tests have very different theoretical properties than their large-T counterparts. Monte Carlo experiments demonstrate the usefulness of the asymptotic theory in small samples.

JEL classification: C22, C23

Keywords: Cross section correlation, strong factors, incidental trends, bias correction, fixed T

a :School of Economics and Granger Centre for Time Series Econometrics, University of Nottingham. University Park, Nottingham NG7 2RD, UK. Tel: +441159515480. E-mail: ioannis.karavias@nottingham.ac.uk.

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

∗ :Corresponding author

The authors would like to thank the Editor Esfandiar Maasoumi, an Associate Editor, two anonynous referees, Tassos Magdalinos and Robert Taylor for their helpful comments.

1 Introduction

There is recently growing interest in developing panel data unit root tests allowing for a break in their deterministic components, namely in their individual effects and/or individual linear trends (see, Carrion-i-Silvestre et al. (2005), Harris et al. (2005), Karavias and Tzavalis (2014a, 2014b), Chan and Pauwels (2011), Bai and Carrion-i-Silvestre (2009), Hadri et al. (2012) and Pauwels et al. (2012)). As is aptly noted by Perron (1989) in the single time-series literature, not accounting for a break point in the level and/or deterministic trend of economic series can lead to a unit root test which can hardly reject the null hypothesis of unit root from its alternative of stationarity. Panel unit root tests suffer from this problem too. But, despite the many panel unit root tests proposed in the literature, to our knowledge, there has been no attempt of studying the behaviour of these tests theoretically.

This paper constistutes the first work in this direction. It investigates the power properties of fixed-T panel unit root tests that allow for structural breaks. These tests are appropriate for panels with few time series observations and many cross-section units, often met in practice (see, e.g., Baltagi (2008)). The asymptotic theory employed considers the time dimension (T) as fixed and the cross section one (N) as going to infinity. In particular, the focus is in the asymptotic local power of two tests proposed in Karavias and Tzavalis (2014a) and in Karavias and Tzavalis (2014b). The first test generalizes the Harris and Tzavalis tests (1999) to allow for a common break, and will be henceforth denoted as HT. The second test (denoted as KT) allows, in addition to structural breaks, for serial correlation in the error term of the individual series of the panel.[1] One of the contributions of this paper is that it extends these tests to the case of cross sectionally dependent errors, so that

[1] Note that a version of the KT test for the case of no structural breaks has been suggested by Kruiniger and Tzavalis (2002), and Moon and Peron (2004) for the case that T is large.

we can study their power properties in a more general setting.

Both the above tests are based on the within groups estimator of the autoregressive coefficient of the AR(1) panel data model, which is just the least squares (LS) estimator on the transformed AR(1) model. This transformation is necessary for the removal of the individual deterministic terms and the initial conditions of the series, but it renders the LS estimator inconsistent as it induces correlation between the lagged dependent variable and the error term. The HT and KT tests correct for this inconsistency of the LS estimator (for simplicity, we will also use the term bias) in different ways. The HT test corrects for the bias of both the numerator and denominator, while the KT test corrects only for the bias of the numerator. This bias correction is fundamentally different in the fixed-T and large-T settings. As we show in the paper, it is the main source of the distinct and superior behaviour of the fixed-T tests over the large-T ones, for short panels. In the large-T setting, Moon and Perron (2004) show how the bias of the numerator is a function of a long run variance, which they estimate using kernel estimators that, as they note, have bad small sample properties. The KT test however corrects for the bias using a fixed T non-parametric estimator based on the covariance matrix estimation method of Abowd and Card (1989) and Arellano (1990, 2003). This method is consistent across the N dimension of the panel and has good small sample properties.

The paper makes a number of contributions into the literature of panel data unit root tests, which have practical implications. First, it shows that, for the standard panel data model with IID errors and individual intercepts, the HT test has higher asymptotic local power than the KT test. This can be attributed to the fact that the HT test does not require a consistent estimator of the variance of the error term, compared to the KT test. The HT test is invariant to this nuisance parameter, as it adjusts the LS estimator for its inconsistency of both its numerator and numerator. Second, as with panel unit root tests that do not allow for a break, the HT and KT tests have trivial asymptotic local power if incidental trends are included in the deterministic components of the AR(1) panel data model. The allowance for a break in the deterministic components of this panel data model does not save these tests from this problem.

Third, when short term serial correlation of arbitrary form is permitted, the KT test can increase its power and, for the panel data model with incidental trends, it has non-trivial asymptotic local power. This is important because large-T panel unit root tests have trivial power in the natural $N^{-1/2}T^{-1}$ neighbourhood of unity, when incidental trends are present (see Moon et al. (2007)). This rise of the power can be attributed to the interaction between the serial correlation parameters and the fixed T non-parametric estimator of the bias.

Finally, the paper extends the two tests for the case that the error term has a strong factor structure. It is shown that both the HT and KT tests have good power properties for the panel data model with individual intercepts. However, for the model with incidental trends, only the KT test is found to have non-trivial power. This finding is in sharp contrast with the large-T case of the KT test, which has trivial local power for panels with a large cross section dimension. It is shown that this power comes from the way the LS estimator is bias corrected which is different than in the large-T case. The above results are confirmed through a Monte Carlo experiment. This exercise also provides small sample results on the power performance of the tests and shows the usefulness of the asymptotic approximation.

The paper is organized as follows: Section 2 presents the assumptions on the data generating process required by the HT and KT tests. Section 3 derives the limiting distributions of the tests for $NIID$ errors. For the KT test allowing for serial correlation effects, this is done in Section 4. Section 5 considers the case of cross section dependence. Section 6 carries out the Monte Carlo exercise. Section 7 concludes the paper. All proofs are given in the appendix.

2 Models and Assumptions

Consider the following AR(1) dynamic panel data models allowing for a common structural break in their deterministic components (individual effects and/or individual linear trends) at time point λ, for all individual units of the panel i:

$$M1: \; y_i = a_i^{(1)} e^{(1)} + a_i^{(2)} e^{(2)} + \zeta_i, \qquad i = 1, 2, .., N,$$

$$M2: \; y_i = a_i^{(1)} e^{(1)} + a_i^{(2)} e^{(2)} + \beta_i^{(1)} \tau^{(1)} + \beta_i^{(2)} \tau^{(2)} + \zeta_i, \quad i = 1, ..., N$$

where

$$\zeta_i = \varphi \zeta_{i,-1} + u_i,$$

$\varphi \in (-1, 1]$, $y_i = (y_{i,1}, ..., y_{i,T})'$ and $y_{i,-1} = (y_{i,0}, ..., y_{i,T-1})'$ are $T \times 1$ vectors, $u_i = (u_{i,1}, ..., u_{i,T})$ is the $T \times 1$ vector of error terms $u_{i,t}$, a_i and β_i denote the individual effects and slope coefficients of the linear (incidental) trends of the panel. In particular, a_i is defined as $a_i = a_i^{(1)}$ if $t \leq T_0$ and $a_i = a_i^{(2)}$ if $t > T_0$, while $e^{(1)}$ and $e^{(2)}$ are $T \times 1$-column vectors defined as follows: $e_t^{(1)} = 1$ if $t \leq T_0$ and 0 otherwise, and $e_t^{(2)} = 1$ if $t > T_0$ and 0 otherwise. Slope coefficients β_i are defined as $\beta_i = \beta_i^{(1)}$ if $t \leq T_0$ and $\beta_i = \beta_i^{(2)}$ if $t > T_0$, while $\tau^{(1)}$ and $\tau^{(2)}$ are $T \times 1$-column vectors defined as follows: $\tau_t^{(1)} = t$ if $t \leq T_0$, and zero otherwise, and $\tau_t^{(2)} = t$ if $t > T_0$, and zero otherwise. Throughout the paper, we will denote the set of possible dates that the break can occurs with I_λ and the break fraction with $\lambda = T_0/T$, i.e. $\lambda \in I_\lambda$.

The above models nest in the same framework both the null hypothesis of unit roots in φ, i.e., $\varphi = 1$, and its alternative of stationarity, $\varphi < 1$. They can be written in a non-linear form as follows:

$$\begin{aligned} y_i &= \varphi y_{i,-1} + (1-\varphi)(a_i^{(1)} e^{(1)} + a_i^{(2)} e^{(2)}) + u_i, \qquad i = 1, 2, ..., N \quad \text{and} \\ y_i &= \varphi y_{i,-1} + \varphi \beta_i^{(1)} e^{(1)} + \varphi \beta_i^{(2)} e^{(2)} + (1-\varphi)(a_i^{(1)} e^{(1)} + a_i^{(2)} e^{(2)}) + (1-\varphi)(\beta_i^{(1)} \tau^{(1)} + \beta_i^{(2)} \tau^{(2)}) + u_i, \end{aligned}$$

respectively. The "within group" least squares (LS) (known also as least squares dummy variables (LSDV)) estimator of autoregressive coefficient φ of the models can be written as follows:

$$\hat{\varphi}^{(\lambda)} = \left(\sum_{i=1}^{N} y_{i,-1}' Q^{(\lambda)} y_{i,-1} \right)^{-1} \left(\sum_{i=1}^{N} y_{i,-1}' Q^{(\lambda)} y_i \right),$$

where $Q^{(\lambda)}$ is the $T \times T$ "within" transformation (annihilator) matrix of the individual series of the panel $y_{i,t}$. $Q^{(\lambda)}$ is defined as $Q^{(\lambda)} = I - X^{(\lambda)} \left(X^{(\lambda)\prime} X^{(\lambda)} \right)^{-1} X^{(\lambda)\prime}$, where $X^{(\lambda)} = \left(e^{(1)}, e^{(2)} \right)$ for model $M1$ and $X^{(\lambda)} = \left(e^{(1)}, e^{(2)}, \tau^{(1)}, \tau^{(2)} \right)$ for model $M2$. I denotes the $T \times T$ identity matrix. The within transformation of the data wipes off the individual effects and/or incidental trends of the panel, as well as its initial conditions $y_{i,0}$, but it results in an inconsistent estimator because it induces correlation between the transformed error and the transformed lagged dependent variable. Thus, fixed-T panel unit root tests based on it must rely on a correction of estimator $\hat{\varphi}^{(\lambda)}$ for its inconsistency (asymptotic bias) (see, e.g., Harris and Tzavalis (1999, 2004)). To study the asymptotic local power of these tests, define the autoregressive coefficient φ as $\varphi_N = 1 - c/\sqrt{N}$. Then, the hypotheses of interest become

$$H_0\colon\ c = 0 \quad \text{and} \quad H_a\colon\ c > 0,$$

where c is the local to unity parameter. The limiting distributions of the tests based on LSDV estimator $\hat{\varphi}^{(\lambda)}$ will be derived under the sequence of local alternatives φ_N, by making the following general assumptions:

Assumption A: (a1) $\{u_i\}$, $i \in \{1, 2, ..., N\}$, constitutes a sequence of $T \times 1$ independent random vectors with means $E(u_i) = 0$ and all mixed $4+\delta$ moments are finite. (a2) $E(u_i u_i') = \Gamma_i$ with $\gamma_{i,ts} = E(u_{i,t} u_{i,s}) = 0$ for $t < s$ and $s = t + p_i + 1, ..., T$, where p_i denotes the order of serial correlation for each i and $p_i \leq p_{\max}$. (a3) Define $\bar{\Gamma} = (1/N) \sum_{i=1}^{N} \Gamma_i$ for which it holds that $\lim_N (N\bar{\Gamma})^{-1}\Gamma_i = \lim_N \left(\sum_{i=1}^{N} \Gamma_i\right)^{-1} \Gamma_i = 0$ and also assume that matrix $\Gamma = \lim_N \bar{\Gamma}$ is positive definite. (a4) The error terms $u_{i,t}$ is independent of $a_i^{(1)}, a_i^{(2)}$ and $y_{i,0}$, for all i, and $Var(y_{i,0}) < +\infty$.

Assumption B: $p_{\max} = [T/2 - 2]^*$ for model $M1$ and

$$p_{\max} = \left\{ \begin{array}{c} \frac{T}{2} - 3 \text{ if } T \text{ is even and } T_0 = T/2, \\ \min\{T_0 - 2, T - T_0 - 2\} \text{ otherwise} \end{array} \right\}$$

for model $M2$, where $[.]^*$ denotes the greatest integer function.

Assumption C: (b1) $\beta_i^{(1)}$ and $\beta_i^{(2)}$ are sequences of independent random variables with finite $4+\delta$ moments. They are also independent from u_i. (b2) $\lim_N \max(E(\beta_i^{(j)2}))/(N\bar{\beta}^{(j)2}) = 0$, where $\bar{\beta}^{(j)2} = (1/N) \sum_{i=1}^{N} E(\beta_i^{(j)2})$ for $j = 1, 2$. Also, $\beta^{(j)2} = \lim_N \bar{\beta}^{(j)2}$ is finite.

Assumption D: The break fraction $\lambda \in I_\lambda = \{2/T, 3/T, ..., (T-1)/T\}$ for model $M1$ and $\lambda \in I_\lambda = \{2/T, 3/T, ..., (T-2)/T\}$ for model $M2$.

Assumption A enables us to derive the limiting distribution of the fixed-T panel data unit root tests of Karavias and Tzavalis (2014a) for $\Gamma_i = \sigma_i^2 I$. These tests (denoted as HT) extend those of Harris and Tzavalis (1999) for the case of a common break in the deterministic components of models $M1$ and $M2$. It also allows the derivation of this limiting distribution for Karavias' and Tzavalis (2014b) fixed-T panel data unit root test (denoted as KT), which allows for a structural break under heteroscedasticity and serial correlation of error terms $u_{i,t}$, for both models $M1$ and $M2$. Condition (a1) states that u_i is mean zero and that, element-wise, all possible $4+\delta$ moments of it are finite. Condition (a2) allows $u_{i,t}$ to have different types of heteroscedasticity and serial correlation across the cross section units and the time dimension of the panel. However, there is a common bound to the order of serial correlation (see also Assumption B). If $\Gamma_i = \sigma^2 I$ for all i, then (a2) is consistent with the assumption of Karavias and Tzavalis (2014a) panel data unit root tests, considering the simpler case of $u_i \sim NIID(0, \sigma^2 I)$.

Condition (a3) states that no individual variance of $u_{i,t}$ is big enough to dominate the rest. This is an assumption required by the Lindeberg-Feller CLT. Finally, condition (a4) imposes independence between the parameters of the series of the panel $y_{i,t}$ and the innovations. It also implies that $Var(y_{i,0}) < +\infty$ which is consistent with assumptions like constant, random and mean stationary initial conditions $y_{i,0}$. Covariance stationarity of $y_{i,0}$, implying $Var(y_{i,0}) = \sigma^2/(1-\varphi_N^2)$ (see Kruiniger (2008) and Madsen (2010)) is not considered. This is because, as is also aptly noted by Moon et al. (2007), this assumption implies that $Var(y_{i,0}) \rightarrow \infty$ when $\varphi_N \rightarrow 1$, which means that the variance of the initial condition increases with the number of cross-section units. This is not meaningful for cross-section data sets.

Assumption B determines the maximum allowable order of serial correlation because of its interaction with the structural break. We chose to restrict the order of serial correlation and let the break date free, rather than the opposite. Finally, Assumption C is relevant only for the case of the KT test for model $M2$. Conditions (c1) and (c2) guarantee that

$\beta_i^{(1)}$ and $\beta_i^{(2)}$, which appear in the estimator of the bias correction, obey the Lindeberg-Feller CLT. Assumption D determines the possible break points. An advantage of the HT and KT tests is that the trimming of the sample depends on the deterministic specification of the panel data models $M1$ and $M2$, and it is less severe than that assumed by single time series unit root tests allowing for breaks, i.e. for $M1$, $I_\lambda = \{2/T, 3/T, ..., (T-1)/T\}$ and, therefore, only the first and last dates are trimmed out as opposed to the $\{0.15, 0.85\}$ interval, advocated in Andrews (1993).

To study the asymptotic local power of the tests, we will rely on the slope parameter, denoted as k, of local power functions of the form

$$\Phi(z_a + ck),$$

where Φ is the standard normal cumulative distribution function and z_a denotes the α-level percentile. Since Φ is strictly monotonic, a larger k means greater power for the same value of c. If k is positive, then the tests will have non-trivial power. If it is zero, they will have trivial power, which is equal to a, and, finally, if $c < 0$ they will be biased.

3 The limiting distribution of the tests if $u_i \sim NIID(0, \sigma^2 I)$

This section presents the limiting distribution of the HT and KT test statistics under the sequence of local alternatives $\varphi_N = 1 - c/\sqrt{N}$ when $u_i \sim NIID(0, \sigma^2 I)$. This is a special case of Assumption A where $p_{\max} = 0$. The assumption of normality is made only for convenience, because in this case we can calculate the analytic formula of the HT test statistic variance. For ease of exposition, it is also assumed that $\beta_i^{(j)}$ are IID. This is a special case of Assumption C. As mentioned before, the HT test corrects both the numerator and the denominator of the LS estimator $\hat{\varphi}^{(\lambda)}$ for its inconsistency, while the KT test corrects only the numerator of $\hat{\varphi}^{(\lambda)}$. This enables the KT test to be easily extended to allow for more general dependence structures of error term u_i.

3.1 Model $M1$

For model $M1$, the HT test allowing for a break is based on the following statistic:

$$Z_{HT}^{(\lambda)} = V_{HT}^{(\lambda)-1/2}\sqrt{N}(\hat{\varphi}^{(\lambda)} - 1 - B(\lambda)),$$

where $B(\lambda) = p\lim(\hat{\varphi}^{(\lambda)} - 1) = tr(\Lambda' Q^{(\lambda)})/tr(\Lambda' Q^{(\lambda)}\Lambda)$ is the inconsistency of LS estimator $\hat{\varphi}^{(\lambda)}$ under null hypothesis H_0: $c = 0$, where Λ is a $T \times T$ dimension matrix having unities at its lower than its main diagonals and zeroes elsewhere. $V_{HT}^{(\lambda)} = 2tr(A_{HT}^{(\lambda)2})/tr(\Lambda' Q^{(\lambda)}\Lambda)^2$, with $A_{HT}^{(\lambda)} = (1/2)(\Lambda' Q^{(\lambda)} + Q^{(\lambda)}\Lambda) - B(\lambda)(\Lambda' Q^{(\lambda)}\Lambda)$, is the variance of the limiting distribution of the corrected for its inconsistency LS estimator $\hat{\varphi}^{(\lambda)}$, i.e. $\sqrt{N}(\hat{\varphi}^{(\lambda)} - 1 - B(\lambda))$. The KT test is based on the following statistic:

$$Z_{KT}^{(\lambda)} = V_{KT}^{(\lambda)-1/2}\hat{\delta}^{(\lambda)}\sqrt{N}\left(\hat{\varphi}^{(\lambda)} - 1 - \frac{\hat{b}^{(\lambda)}}{\hat{\delta}^{(\lambda)}}\right),$$

where $\hat{b}^{(\lambda)}/\hat{\delta}^{(\lambda)} \equiv \hat{\sigma}^2 tr(\Lambda' Q^{(\lambda)})/\left(\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1}Q^{(\lambda)}y_{i,-1}\right)$ is a consistent estimator of the bias of $\hat{\varphi}^{(\lambda)}$ based on a consistent estimator of the bias of the numerator, namely $\hat{\sigma}^2 tr(\Lambda' Q^{(\lambda)})$, divided by $\hat{\delta}^{(\lambda)}$ which is the denominator of $\hat{\varphi}^{(\lambda)}$. $V_{KT}^{(\lambda)} = 2\sigma^4 tr(A_{KT}^{(\lambda)2})$, with $A_{KT}^{(\lambda)} = (1/2)(\Lambda' Q^{(\lambda)} + Q^{(\lambda)}\Lambda - \Psi^{(\lambda)} - \Psi^{(\lambda)\prime})$, is the variance of the limiting distribution of $\sqrt{N}\hat{\delta}^{(\lambda)}\left(\hat{\varphi}^{(\lambda)} - \hat{b}^{(\lambda)}/\hat{\delta}^{(\lambda)} - 1\right)$ and $\Psi^{(\lambda)}$ is a $T \times T$ dimension matrix having in its main diagonal the corresponding elements of matrix $\Lambda' Q^{(\lambda)}$ and zeros elsewhere.

In implementing the KT test statistic, note that a consistent estimator of σ^2 under H_0: $c = 0$ is given by $\hat{\sigma}^2 = \left[1/\left(tr(\Psi^{(\lambda)})N\right)\right] \sum_{i=1}^{N} tr(\Psi^{(\lambda)} \Delta y_i \Delta y_i')$, where $\Delta y_i = y_i - y_{i,-1}$. Matrix $\Psi^{(\lambda)}$ implies that $tr(\Psi^{(\lambda)}) = tr(\Lambda' Q^{(\lambda)})$. It is designed so as, in adjusting the numerator of the estimator $\hat{\varphi}^{(\lambda)}$ for its inconsistency, some sample information is left to test the null hypothesis. In particular, $\Psi^{(\lambda)}$ has two properties. First, it restricts the estimator $(1/N) \sum_{i=1}^{N} \Delta y_i \Delta y_i'$, which constitutes a consistent estimator of $\sigma^2 I$ based on all the available sample information, to its main diagonal.[2] This restriction uses information coming only from contemporaneous observations and, because there is no information about σ^2 in the off-diagonal elements of $\sigma^2 I$, it preserves then consistency of the estimator. Second, matrix $\Psi^{(\lambda)}$ weights the diagonal elements of $(1/N) \sum_{i=1}^{N} \Delta y_i \Delta y_i'$ in the same way that are weighted by $tr(\Lambda' Q^{(\lambda)})$, mimicking the part of the bias which is due to the within transformation matrix.

In the next theorem, we give the limiting distribution of the HT and KT test statistics for model $M1$, under the sequence of local alternatives $\varphi_N = 1 - c/\sqrt{N}$.

Theorem 1 *For model $M1$, let Assumptions A and D hold and $u_i \sim NIID(0, \sigma^2 I)$. Then, under $\varphi_N = 1 - c/\sqrt{N}$, we have*

$$V_{HT}^{(\lambda)-1/2} \sqrt{N} (\hat{\varphi}^{(\lambda)} - 1 - B(\lambda)) \xrightarrow{d} N(-ck_{HT}, 1)$$

$$\text{and} \qquad V_{KT}^{(\lambda)-1/2} \hat{\delta}^{(\lambda)} \sqrt{N} \left(\hat{\varphi}^{(\lambda)} - 1 - \frac{\hat{b}^{(\lambda)}}{\hat{\delta}^{(\lambda)}} \right) \xrightarrow{d} N(-ck_{KT}, 1),$$

[2]Notice that, under $H_0 : c = 0$, we have $\Delta y_i = u_i$ and, thus, $p\lim \hat{\sigma}^2 = p\lim[1/(tr(\Psi^{(\lambda)})N)] \sum_{i=1}^{N} tr(\Psi^{(\lambda)} \Delta y_i \Delta y_i') = \sigma^2 tr(\Lambda' Q^{(\lambda)})/tr(\Psi^{(\lambda)}) = \sigma^2$, since $tr(\Psi^{(\lambda)}) = tr(\Lambda' Q^{(\lambda)})$.

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

$$k_{HT} = \frac{T(T-2)\left[T^2(3\lambda^2-3\lambda+1)-1\right]}{4T^2(2\lambda^2-2\lambda+1)-8}\sqrt{\frac{T^4\Phi_1+T^2\Phi_2+240}{T^6R_1+T^5R_2+T^4R_3+T^2R_4+216T-136}}$$

$$\text{and} \quad k_{KT} = \frac{\sqrt{3}(T-2)}{\sqrt{T^2(2\lambda^2-2\lambda+1)+6T+10-\frac{4(-\frac{1}{T}+2(\lambda-1)\lambda T)}{(\lambda-1)\lambda}}},$$

where R_1, R_2, R_3, R_4 *and* Φ_1, Φ_2 *are polynomials of* λ *defined in the appendix (see proof of the theorem).*

The limiting distributions given by Theorem 1 imply that the asymptotic local power function of test statistics HT and KT depend on the values of slope parameters k_{HT} and k_{KT}, respectively. In Table 1, we present values of these parameters, for different values of T and λ. The results of this table indicate that the asymptotic local power behaviour of the two test statistics is different. The HT statistic has much higher power than the KT. The power of this statistic is much bigger when the break is in the beginning, or towards the end of the sample, i.e., for $\lambda = \{0.25, 0.75\}$.[3] On the other hand, the power of test statistic KT reaches its maximum point when the break is in the middle of the sample, $\lambda = \{0.50\}$. The power of statistic HT increases with T, i.e., $k_{HT} = O(T)$. The power of the KT test increases with T, but for relatively small T. As T grows large, the test has no power gains. This can be seen from $\lim_T k_{KT} = \sqrt{3}/\sqrt{2\lambda^2-2\lambda+1}$, which is independent of T. These results can be more clearly seen by the three-dimension Figures 1 and 2, presenting values of k_{HT} and k_{KT}, for different values of λ and T.

The above differences between test statistics HT and KT can be attributed to the way

[3] Analogous evidence is provided for single time series unit root tests allowing for breaks, based on a model selection Bayesian approach (see Meligkotsidou et al. (2011)).

that each of them corrects for the inconsistency of the LS estimator $\hat{\varphi}^{(\lambda)}$. As mentioned before, HT is based on a correction of LS estimator $\hat{\varphi}^{(\lambda)}$ for the inconsistency of both its numerator and denominator. On the other hand, the KT test statistic is based on an adjustment of estimator $\hat{\varphi}^{(\lambda)}$ only for the inconsistency of its numerator, which additionally requires a consistent estimator of the variance of error term $u_{i,t}$, σ^2. The later reduces the local power of the test. Finally, another result of Theorem 1 is that, under the sequence of local alternatives considered, the break function parameters do not enter the asymptotic distribution of both test statistics HT and KT. Thus, the magnitude of the break does not affect local power of the tests. Furthermore, local power is also robust, asymptotically, to the initial conditions of the panel $y_{i,0}$, which means that their magnitudes also do not affect the power of the test (see also Harvey and Leybourne (2005) and Harris et al. (2010)).

Scaling appropriately test statistics HT and KT by T and assuming that $T,N \to \infty$, with $\sqrt{N}/T \to 0$, it can be shown (see appendix) that, under $\varphi_{N,T} = 1 - c/\left(T\sqrt{N}\right)$, the limiting distributions of the large-T versions of these statistics are given as follows:

Corollary 1 *For model M1, let Assumptions A and D hold and $u_i \sim NIID(0,\sigma^2 I)$. Then, under $\varphi_{N,T} = 1 - c/\left(T\sqrt{N}\right)$, we have*

$$\begin{aligned} & V_{HT}^{*(\lambda)-1/2} T\sqrt{N}(\hat{\varphi}^{(\lambda)} - 1 - B(\lambda)) \xrightarrow{L} N\left(-ck_{HT}^{*}, 1\right), \\ \text{and} \qquad & V_{KT}^{*(\lambda)-1/2} \hat{\delta}^{(\lambda)} T\sqrt{N}\left(\hat{\varphi}^{(\lambda)} - 1 - \frac{\hat{b}^{(\lambda)}}{\hat{\delta}^{(\lambda)}}\right) \xrightarrow{d} N\left(-ck_{KT}^{*}, 1\right), \end{aligned}$$

as $T,N \to \infty$, with $\sqrt{N}/T \to 0$, where

$$k_{HT}^{*} = \frac{3\lambda^2 - 3\lambda + 1}{4(2\lambda^2 - 2\lambda + 1)}\sqrt{\frac{\Phi_1}{R_1}} \quad \text{and } k_{KT}^{*} = 0, \tag{1}$$

and

$$V_{HT}^{*(\lambda)} = \frac{36R_1}{\Phi_1(2\lambda^2 - 2\lambda + 1)^2} \text{ and } V_{KT}^{*(\lambda)} / \left(p \lim_{N,T} \hat{\delta}^{(\lambda)} \right) = \frac{36(2\lambda^4 - 4\lambda^3 + 3\lambda^2 - \lambda)}{12(\lambda - 1)\lambda(2\lambda^2 - 2\lambda + 1)^2},$$

respectively, denote the local power slope coefficients and the variances of the limiting distributions of the large-T versions of the HT and KT test statistics.

Values of power slope coefficients k_{HT}^* and k_{KT}^*, for different values of λ, are reported in Table 2. These indicate that, in contrast to the HT test, the large-T extension of test statistic KT does not have asymptotic local power.[4] Thus, the KT test can be thought of as more appropriate for short panels. The results of the table also indicate that the large-T extension of the HT test has less power than its fixed-T version. We have also found that power takes its highest values in the beginning and towards the end of the sample, i.e., for $\lambda = \{0.10, 0.90\}$, as with its fixed-T version. The smaller power of the large-T versions of test statistics HT and KT, compared to their fixed-T ones, can be attributed to the faster rate of convergence of the alternative hypotheses to the null, i.e. $\varphi_{N,T} = 1 - c/\left(T\sqrt{N}\right)$ compared to $\varphi_N = 1 - c/\sqrt{N}$ (see also Harris et al. (2010)).

The test statistics given by Theorem 1 and Corollary 1 can be readily applied in practice. To this end, for test statistic HT, first the annihilator matrix $Q^{(\lambda)}$ must be built, where $Q^{(\lambda)}$ is a deterministic matrix based on vectors $e^{(1)}$ and $e^{(2)}$. Then, given Λ which is a fixed matrix and $\Psi^{(\lambda)}$ which is a restricted form of $\Lambda' Q^{(\lambda)}$, the bias $B(\lambda) = tr(\Lambda' Q^{(\lambda)})/tr(\Lambda' Q^{(\lambda)} \Lambda)$ and matrix $A_{HT}^{(\lambda)} = (1/2)(\Lambda' Q^{(\lambda)} + Q^{(\lambda)} \Lambda) - B(\lambda)(\Lambda' Q^{(\lambda)} \Lambda)$ can be easily calculated. With these quantities at hand, the last step is to calculate the LS estimator and the $Z_{HT}^{(\lambda)}$ test

[4] Note that an analogous result has been derived by Moon and Perron (2008) for this test in the case of no break.

statistic. Similar steps to those above can be followed for the application of test statistic KT. $\hat{\sigma}^2 = \left[1/\left(tr(\Psi^{(\lambda)})N\right)\right]\sum_{i=1}^{N} tr(\Psi^{(\lambda)}\Delta y_i \Delta y_i')$ must also be calculated as part of $\hat{b}^{(\lambda)}$. $\hat{\delta}^{(\lambda)}$ is the denominator of the LS estimator. As a final note, if error terms $u_{i,t}$ are non-normal as is the most probable case, only the formula of the variance of statistic HT changes in the above theorems. This becomes $V_{HT}^{(\lambda)} = \left[k_4 \sum_{j=1}^{N} a_{HT,jj}^{(\lambda)2} + 2\sigma^4 tr(A_{HT}^{(\lambda)2})\right] / \left[\sigma^2 tr(\Lambda' Q^{(\lambda)} \Lambda)\right]^2$, where $A_{HT}^{(\lambda)} = [a_{HT,ij}^{(\lambda)}]$ and parameter k_4 can be estimated as in Harris and Tzavalis (2004).

If the date of the break is unknown, as often assumed in practice, then to test null hypothesis H_0: $c = 0$ we may either estimate it as in Bai (2010) or rely on the minimum values of the known break test statistics HT and KT over all possible break points of the sample, denoted respectively as $min_{\lambda \in I_\lambda} Z_{HT}^{(\lambda)}$ and $min_{\lambda \in I_\lambda} Z_{KT}^{(\lambda)}$. As shown in Karavias and Tzavalis (2014a), the limiting distributions of these test statistics behave like the minimum of a fixed number of correlated normal variables. The pdf of this minimum is given by Karavias and Tzavalis (2014b). Because inverting this pdf is a numerical problem and because the results are qualitatively similar to those of the known date break tests presented above, we do not pursue this issue any further.

3.2 Model $M2$

For model $M2$, which additionally considers incidental trends in the deterministic components of panel data series $y_{i,t}$, the HT and KT test statistics are defined analogously to those for model $M1$. Test statistic HT admits the same formulas, but now matrix $Q^{(\lambda)}$ is based on $X^{(\lambda)} = \left(e^{(1)}, e^{(2)}, \tau^{(1)}, \tau^{(2)}\right)$. $B(\lambda) = p\lim(\hat{\varphi}^{(\lambda)} - 1) = tr(\Lambda' Q^{(\lambda)})/tr(\Lambda' Q^{(\lambda)} \Lambda)$ denotes the inconsistency of LS estimator $\hat{\varphi}^{(\lambda)}$, for model $M2$ and $V_{HT}^{(\lambda)} = 2tr(A_{HT}^{(\lambda)2})/tr(\Lambda' Q^{(\lambda)} \Lambda)^2$, with $A_{HT}^{(\lambda)} = \frac{1}{2}(\Lambda' Q^{(\lambda)} + Q^{(\lambda)} \Lambda) - B(\lambda)(\Lambda' Q^{(\lambda)} \Lambda)$, is the variance of the limiting distribution of

$\sqrt{N}(\hat{\varphi}^{(\lambda)} - 1 - B(\lambda))$.

However, for test statistic KT, $\hat{\sigma}^2 = \left[1/\left(tr(\Psi^{(\lambda)})N\right)\right] \sum_{i=1}^{N} tr\left(\Psi^{(\lambda)} \Delta y_i \Delta y_i'\right)$ is no longer a consistent estimator of σ^2 in the case of model $M2$, due to the presence of individual coefficients (effects) β_i under null hypothesis H_0: $c = 0$. These imply that $\Delta y_i = \beta_i^{(1)} e^{(1)} + \beta_i^{(2)} e^{(2)} + u_i$ and it can be easily seen that

$$p \lim_{N} \frac{1}{N} \sum_{i=1}^{N} \Delta y_i \Delta y_i' = E(\beta_i^{(1)2}) e^{(1)} e^{(1)\prime} + E(\beta_i^{(1)2}) e^{(2)} e^{(2)\prime} + \sigma^2 I. \tag{2}$$

To render test statistic KT invariant to nuisance parameters β_i, Karavias and Tzavalis (2014b) suggested the following estimator of σ^2:

$$\hat{\sigma}^2 = \frac{1}{Ntr(\Theta^{(\lambda)})} \sum_{i=1}^{N} tr\left(\Theta^{(\lambda)} \Delta y_i \Delta y_i'\right),$$

with

$$\Theta^{(\lambda)} = \Psi^{(\lambda)} - \frac{tr(\Psi^{(\lambda)} e^{(1)} e^{(1)\prime})}{tr(M^{(1)} e^{(1)} e^{(1)\prime})} M^{(1)} - \frac{tr(\Psi^{(\lambda)} e^{(2)} e^{(2)\prime})}{tr(M^{(2)} e^{(2)} e^{(2)\prime})} M^{(2)}, \tag{3}$$

where $\Psi^{(\lambda)}$ is defined as before (i.e., it is a $T \times T$ diagonal matrix having in its main diagonal the elements of the main diagonal of the matrix $\Lambda' Q^{(\lambda)}$), $M^{(1)} = e^{(1)} e^{(1)\prime} - diag\{e^{(1)} e^{(1)\prime}, 0\}$ and $M^{(2)} = e^{(2)} e^{(2)\prime} - diag\{e^{(2)} e^{(2)\prime}, 0\}$, where $diag\{e^{(r)} e^{(r)\prime}, p\}$, $r = \{1, 2\}$, denotes two selection matrices which have zeros everywhere except from their main and p upper and p lower diagonals in which they have the elements of the matrices $e^{(r)} e^{(r)\prime}$. Matrices $M^{(r)}$, for $r = \{1, 2\}$, select the elements of $p \lim_N (1/N) \sum_{i=1}^{N} \Delta y_i \Delta y_i'$ containing individual effects $E(\beta_i^{(r)2})$. In particular, matrices $M^{(1)}$ and $M^{(2)}$ respectively select the off-diagonal elements of the right hand side of (2) where nuisance parameters $E(\beta_i^{(1)2})$ and $E(\beta_i^{(2)2})$ reside. This can be seen by noticing that $p \lim_N tr(M^{(1)} \sum_{i=1}^{N} \Delta y_i \Delta y_i')/Ntr(M^{(1)} e^{(1)} e^{(1)\prime}) = E(\beta_i^{(1)2})$ and

$p\lim_N tr(M^{(2)} \sum_{i=1}^N \Delta y_i \Delta y_i')/N tr(M^{(2)} e^{(2)} e^{(2)\prime}) = E(\beta_i^{(2)2})$. Thus, these matices are used in the adjustment of the LS estimator $\hat{\varphi}^{(\lambda)}$ for its inconsistency to render the limiting distribution of this estimator net of the individual effects β_i.

Having defined matrices $M^{(r)}$, one can see that $\Theta^{(\lambda)}$, given by (3), plays the same role that $\Psi^{(\lambda)}$ does for test statistic KT in the case of model $M1$. However, in addition to rendering the limiting distribution of $\sqrt{N}\left(\hat{\varphi}^{(\lambda)} - 1 - \hat{b}^{(\lambda)}/\hat{\delta}^{(\lambda)}\right)$ net of the diagonal elements of $\sigma^2 I$ (which is done through matrix $\Psi^{(\lambda)}$), matrix $\Theta^{(\lambda)}$ also makes this limiting distribution net of individual effects $E(\beta_i^{(r)2})$, for $r = \{1, 2\}$. The latter is done through matrices $\left[tr(\Psi^{(\lambda)} e^{(1)} e^{(1)\prime})/tr(M^{(1)} e^{(1)} e^{(1)\prime})\right] M^{(1)}$ and $\left[tr(\Psi^{(\lambda)} e^{(2)} e^{(2)\prime})/tr(M^{(2)} e^{(2)} e^{(2)\prime})\right] M^{(2)}$. Given the definition of $\Theta^{(\lambda)}$, the bias adjustment function and the variance of the limiting distribution of $\sqrt{N}\left(\hat{\varphi}^{(\lambda)} - 1 - \hat{b}^{(\lambda)}/\hat{\delta}^{(\lambda)}\right)$ will be respectively given as follows: $\hat{b}^{(\lambda)}/\hat{\delta}^{(\lambda)} = tr(\Theta^{(\lambda)}(1/N) \sum_{i=1}^N \Delta y_i \Delta y_i'))/\left(\frac{1}{N} \sum_{i=1}^N y_{i,-1}' Q^{(\lambda)} y_{i,-1}\right)$ and $V_{KT}^{(\lambda)} = vec(Q^{(\lambda)}\Lambda - \Theta^{(\lambda)\prime})' \Pi vec(Q^{(\lambda)}\Lambda - \Theta^{(\lambda)\prime})$, with $\hat{\Pi} = (1/N) \sum_{i=1}^N vec(\Delta y_i \Delta y_i') vec(\Delta y_i \Delta y_i')$.

The next theorem derives the limiting distribution of test statistics HT and KT for model $M2$ under the sequence of local alternatives $\varphi_N = 1 - c\sqrt{N}$.

Theorem 2 *For model $M2$, let Assumptions A, C and D hold with $u_i \sim NIID(0, \sigma^2 I)$ and $\beta_i^{(j)}$ be IID, for $j = 1, 2$. Then, under $\varphi_N = 1 - c\sqrt{N}$, we have*

$$V_{HT}^{(\lambda)-1/2} \sqrt{N} (\hat{\varphi}^{(\lambda)} - 1 - B(\lambda)) \xrightarrow{L} N(-ck_{HT}, 1) \quad \textit{and}$$

$$V_{KT}^{(\lambda)-1/2} \hat{\delta}^{(\lambda)} \sqrt{N} \left(\hat{\varphi}^{(\lambda)} - 1 - \frac{\hat{b}^{(\lambda)}}{\hat{\delta}^{(\lambda)}}\right) \xrightarrow{d} N(-ck_{KT}, 1),$$

as $N \to \infty$, where

$$k_{HT} = 0 \textit{ and } k_{KT} = 0.$$

The results of the theorem indicate that the well-known incidental trends problem of panel data unit root tests (see e.g. Moon et al. (2007)) also exists even if the tests allow for a break and T is fixed. Both, the HT and KT test statistics have trivial power, for model $M2$. This result also holds for the case that T grows large and is established in the next corollary.

Corollary 2 *For model $M2$, let Assumptions A, C and D hold with $u_i \sim NIID(0, \sigma^2 I)$ and $\beta_i^{(r)}$ be IID, for $r = 1, 2$. Then, under $\varphi_{N,T} = 1 - c/\left(T\sqrt{N}\right)$, we have*

$$V_{HT}^{*(\lambda)-1/2} T\sqrt{N}(\hat{\varphi}^{(\lambda)} - 1 - B(\lambda)) \xrightarrow{L} N\left(-ck_{HT}^{*}, 1\right),$$

$$\text{and} \qquad V_{KT}^{*(\lambda)-1/2} \hat{\delta}^{(\lambda)} T\sqrt{N} \left(\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{\delta}^{(\lambda)}} - 1\right) \xrightarrow{d} N\left(-ck_{KT}^{*}, 1\right),$$

as T,$N \to \infty$, with $\sqrt{N}/T \to 0$, with

$$k_{HT}^{*} = 0 \text{ and } k_{KT}^{*} = 0, \tag{4}$$

where k_{HT}^{*} and k_{KT}^{*} denote the local power slope coefficients of the large-T versions of the HT and KT test statistics.

The implementation of the test statistics given by Theorem 2 (or Corollary 2), for model $M2$, follows the same steps to those for the test statistics for model $M1$, given by Theorem 1 (or Corollary 1). More specifically, for test statistic HT matrix $Q^{(\lambda)}$, with $X^{(\lambda)} = \left(e^{(1)}, e^{(2)}, \tau^{(1)}, \tau^{(2)}\right)$, must be employed. For test statistic KT, the fixed matrices $M^{(r)}$, for $r = \{1, 2\}$, and $\Theta^{(\lambda)}$ must be built. Then, the variance function $V_{KT}^{(\lambda)-1/2}$ must be calculated, by plugging in the estimator $\hat{\Pi} = (1/N) \sum_{i=1}^{N} vec(\Delta y_i \Delta y_i')vec(\Delta y_i \Delta y_i')'$.

4 Power of the KT tests if error terms u_i are serially correlated

In this section, we consider the case that the variance-covariance matrix of the vector of error terms u_i has a more general form than $\Gamma = \sigma^2 I$, assumed in the previous section. That is, we assume that $\Gamma = [\gamma_{i,ts}]$, where $\gamma_{i,ts} = E(u_{i,t}u_{i,s}) = 0$ for $s = t + p_{\max} + 1, ..., T$ and $t < s$. This means that errors $u_{i,t}$ allow for heteroscedasticity and serial correlation of maximum lag order $p = \max p_i$, for $i = 1, ..., N$. We further allow for cross-sectional heterogeneity, by allowing the type of heteroscedasticity and serial correlation to change with i. The order of serial correlation p, considered may differ across the units of the panel, but we still impose the following bound for it: $p \leq p_{\max}$, which is smaller than T and is determined in Assumption B. This assumption provides some common nuisance parameter free moments which are exploited in the cross section dimension of the panel. We no longer impose the IID assumption on incidental trends slope coefficients $\beta_i^{(r)}$, used for simplicity in the previous section. These less restrictive assumptions enable us to investigate the combined effects of a structural break and serial correlation in $u_{i,t}$ on the asymptotic local power of panel unit root tests. As only the KT test is extended to allow for serially correlated errors $u_{i,t}$ (see, e.g., Karavias and Tzavalis (2014b)), our analysis will be focused on this test.

For both models $M1$ and $M2$, the KT test statistic under the above assumptions about u_i has analogous forms to those presented in the previous section. What changes is that, in order to take into account the p-th order serial correlation in $u_{i,t}$ which appears in the p-upper and p-lower secondary diagonals of matrix Γ, the $T \times T$ selection matrix $\Psi^{(\lambda)}$ now is defined as having in its main diagonal and its p-lower and p-upper diagonals the corresponding elements of matrix $\Lambda' Q^{(\lambda)}$, and zeroes elsewhere. Similarly, $M^{(1)}$ and $M^{(2)}$, defined before, will have zeroes in their p-lower and p-upper diagonals (as opposed to zeroes only in their main diagonal, assumed for the simple case of $\Gamma = \sigma^2 I$), i.e. $M^{(r)} = e^{(r)}e^{(r)\prime} - diag\{e^{(r)}e^{(r)\prime}, p\}$. This is needed, because in these diagonals the trend nuisance parameters appear together with the higher order serial correlation nuisance parameters.

For the $M1$ model, which assumes that $X^{(\lambda)} = \left(e^{(1)}, e^{(2)}\right)$, the inconsistency of LS estimator $\hat{\varphi}^{(\lambda)}$ under null hypothesis H_0: $c = 0$ and Assumptions A, B and D is given by $tr(\Lambda' Q^{(\lambda)}\Gamma)/tr(\Lambda' Q^{(\lambda)}\Lambda\Gamma)$, since $p\lim_N(\hat{\varphi}^{(\lambda)} - 1 - tr(\Lambda' Q^{(\lambda)}\bar{\Gamma})/tr(\Lambda' Q^{(\lambda)}\Lambda\bar{\Gamma})) = 0$. This formula of the inconsistency of $\hat{\varphi}^{(\lambda)}$ indicates that in order to correct $\hat{\varphi}^{(\lambda)}$ we need an estimator of matrix $\bar{\Gamma}$. To this end, define the following estimator: $\hat{\Gamma} = (1/N)\sum_{i=1}^{N} \Delta y_i \Delta y_i'$. Then, by Chebyshev's Weak Law of Large Numbers, we have $p\lim_N \left[\hat{\Gamma} - \bar{\Gamma}\right] = 0$ and, thus, $p\lim_N \left[tr(\Psi^{(\lambda)}\hat{\Gamma}) - tr(\Lambda' Q^{(\lambda)}\bar{\Gamma})\right] = 0$. The last result implies that $\hat{\varphi}^{(\lambda)}$ can be adjusted for its inconsistency, by defining $\hat{b}^{(\lambda)}$ as $\hat{b}^{(\lambda)} = tr(\Psi^{(\lambda)}\hat{\Gamma})$. Then, test statistic $\sqrt{N}\left(\hat{\varphi}^{(\lambda)} - 1 - \hat{b}^{(\lambda)}/\hat{\delta}^{(\lambda)}\right)$ will be centred around zero, where $\hat{\delta}^{(\lambda)}$ is the denominator of $\hat{\varphi}^{(\lambda)}$. The variance function of this statistic is given by $V_{KT}^{(\lambda)} = 2tr\left((A_{KT}^{(\lambda)}\Gamma)^2\right)$, where $A_{KT}^{(\lambda)} = (1/2)(\Lambda' Q^{(\lambda)} + Q^{(\lambda)}\Lambda - \Psi^{(\lambda)} - \Psi^{(\lambda)\prime})$.

The above formula of the inconsistency of LS estimator $\hat{\varphi}^{(\lambda)}$, i.e., $tr(\Lambda' Q^{(\lambda)}\Gamma)/tr(\Lambda' Q^{(\lambda)}\Lambda\Gamma)$ also holds for model $M2$, which assumes that $X^{(\lambda)} = \left(e^{(1)}, e^{(2)}, \tau^{(1)}, \tau^{(2)}\right)$. Under the above assumptions and H_0: $c = 0$, it can be shown $p\lim_N(\hat{\varphi}^{(\lambda)} - 1 - tr(\Lambda' Q^{(\lambda)}\bar{\Gamma})/tr(\Lambda' Q^{(\lambda)}\Lambda\bar{\Gamma})) = 0$. However, $\hat{\Gamma}$ is an inconsistent estimator of $\bar{\Gamma}$ due to the presence of $\beta_i^{(r)}$ under the null hypothesis. It can be easily shown that $p\lim_N \left[\hat{\Gamma} - \bar{\Gamma} - \bar{\beta}^{(1)2}e^{(1)}e^{(1)\prime} - \bar{\beta}^{(2)2}e^{(2)}e^{(2)\prime}\right] = 0$. In this case, to adjust LS estimator $\hat{\varphi}^{(\lambda)}$ for its inconsistency, due to nuisance parameters $\beta_i^{(r)}$ and the presence of serial correlation in $u_{i,t}$ (both implying $p\lim_N \left[\hat{\Gamma} - \bar{\Gamma}\right] \neq 0$), we will employ matrix $\Theta^{(\lambda)}$. This matrix now is based on the modified for p-order serial correlation matrices $\Psi^{(\lambda)}$ and $M^{(r)}$, for $r = \{1, 2\}$, defined above. Then, $p\lim_N \left[tr(M^{(r)}\hat{\Gamma})/tr(M^{(r)}e^{(r)}e^{(r)\prime}) - \bar{\beta}^{(r)2}\right] =$

0, for $r = \{1,2\}$ and, hence, $p\lim_N \left[tr(\Theta^{(\lambda)}\hat{\Gamma}) - tr(\Lambda' Q^{(\lambda)}\bar{\Gamma}) \right] = 0$. The last result establishes that $\hat{b}^{(\lambda)} = tr(\Theta^{(\lambda)}\hat{\Gamma})$ constitutes a consistent estimator of the bias of the numerator of $\hat{\varphi}^{(\lambda)}$, and thus statistic $\sqrt{N}\left(\hat{\varphi}^{(\lambda)} - 1 - \hat{b}^{(\lambda)}/\hat{\delta}^{(\lambda)}\right)$ is centred around 0. Its variance has the same formula as before, i.e., $V_{KT}^{(\lambda)} = vec(Q^{(\lambda)}\Lambda - \Theta^{(\lambda)\prime})'\Pi vec(Q^{(\lambda)}\Lambda - \Theta^{(\lambda)\prime})$, with $\hat{\Pi} = (1/N)\sum_{i=1}^{N} vec(\Delta y_i \Delta y_i')vec(\Delta y_i \Delta y_i')'$.

In the next theorem, we provide the limiting distribution of test statistic KT under the sequence of local alternatives $\varphi_N = 1 - c/\sqrt{N}$, for model $M1$ allowing for serial correlation in $u_{i,t}$.

Theorem 3 *For model $M1$, let Assumptions A, B, and D hold. Then, under $\varphi_N = 1 - c/\sqrt{N}$, we have*

$$V_{KT}^{(\lambda)-1/2}\hat{\delta}^{(\lambda)}\sqrt{N}\left(\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{\delta}^{(\lambda)}} - 1\right) \xrightarrow{d} N(-ck_{KT}, 1), \ \ \textit{for model } M1,$$

as $N \to \infty$, with

$$k_{KT} = \frac{tr(F'Q^{(\lambda)}\Gamma) + tr(\Lambda' Q^{(\lambda)}\Lambda\Gamma) - tr(\Psi^{(\lambda)}\Lambda\Gamma) - tr(\Lambda'\Psi^{(\lambda)}\Gamma)}{\sqrt{2tr((A_{KT}^{(\lambda)}\Gamma)^2)}}.$$

where F is a $T \times T$ deterministic matrix independent of the order of serial correlation, defined in the Appendix.

The results of the theorem indicate that the asymptotic local power of the KT test now depends also on the values of the variance-covariance parameters $\gamma_{i,ts}$, affecting the power slope parameter k_{KT}. This can increase, or reduce, the local power of the test depending on the sign and form of $\gamma_{i,ts}$. To see this more clearly, in Panel A of Table 3 we present estimates of the power slope parameter k_{KT} assuming that error terms $u_{i,t}$ follow a MA(1) process:

$$u_{i,t} = \varepsilon_{i,t} + \theta\varepsilon_{i,t-1},$$

where $\varepsilon_{i,t} \sim NIID(0, \sigma_\varepsilon^2)$. Note that the table also considers the case that $\theta = 0$ (i.e., there is no serial correlation in $u_{i,t}$), but test statistic KT allows for serial correlation of order $p = 1$. This case can show if this test loses significant power if a higher order of serial correlation p is assumed than the correct one. The results of the table also show that the KT test has always power if $\theta \geq 0$. The finding that the test has power even if $\theta = 0$, for all cases of T_0 considered, indicates that it may be applied to test for unit roots even if higher than the correct order of serial correlation is assumed.[5] As was expected, the power of the test in this case is always less, compared to that when the correct lag order $p = 0$ is considered. This happens because in this case the test exploits less moment conditions in drawing inference about unit roots, by assuming $p = 1$ when $\theta = 0$.

Another conclusion that can be drawn from the results of the table is that, when $\theta > 0$, the power of test statistic KT becomes bigger than that of its version which does not allow for serial correlation $u_{i,t}$, presented in the previous section (see Table 2). We have found that this result can be mainly attributed to the presence of selection matrix $\Psi^{(\lambda)}$ in terms $tr(\Psi^{(\lambda)}\Lambda\Gamma)$ and $tr(\Lambda'\Psi^{(\lambda)}\Gamma)$ of the function of the slope coefficient k_{KT}, given by Theorem 3. These terms have a positive effect on k_{KT} (i.e., $tr(\Psi^{(\lambda)}\Lambda\Gamma) + tr(\Lambda'\Psi^{(\lambda)}\Gamma) < 0$) when $\theta > 0$ and a negative effect when $\theta < 0$ (i.e., $tr(\Psi^{(\lambda)}\Lambda\Gamma) + tr(\Lambda'\Psi^{(\lambda)}\Gamma) > 0$).[6] As T increases, the above sign effects of θ on the KT test are amplified. These power gains of the KT test for model $M1$, when $\theta > 0$, may be also attributed to the fact that a positive value of θ adds to

[5] We have found that this is true even for $p > 1$.

[6] The sum of traces $tr(F'Q^{(\lambda)}\Gamma) + tr(\Lambda'Q^{(\lambda)}\Lambda\Gamma)$ affects the power of the KT test, too. However, because this constitutes a parabola function which opens upwards, its effect on k_{KT} is almost symmetrical with respect to the sign of θ. Thus, the relationship between k_{KT} and θ is mainly determined by $tr(\Psi^{(\lambda)}\Lambda\Gamma)+tr(\Lambda'\Psi^{(\lambda)}\Gamma)$.

the variability of individual panel series $y_{i,t}$ driving further away the limiting distributions of the test under the null and alternative hypotheses.

For model $M2$, the limiting distributions of test statistic KT under $\varphi_N = 1 - c/\sqrt{N}$ and serially correlated error terms $u_{i,t}$ are given in the next theorem.

Theorem 4 *For model $M2$, let Assumptions A, B, C and D hold. Then, under $\varphi_N = 1 - c/\sqrt{N}$, we have*

$$V_{KT}^{(\lambda)-1/2}\hat{\delta}^{(\lambda)}\sqrt{N}\left(\hat{\varphi}^{(\lambda)} - 1 - \frac{\hat{b}^{(\lambda)}}{\hat{\delta}^{(\lambda)}}\right) \xrightarrow{d} N\left(-ck_{KT}, 1\right),$$

as $N \to \infty$, where

$$k_{KT} = \frac{tr(F'Q^{(\lambda)}\Gamma) + tr(\Lambda'Q^{(\lambda)}\Lambda\Gamma) - tr(\Theta^{(\lambda)}\Lambda\Gamma) - tr(\Lambda'\Theta^{(\lambda)}\Gamma)}{\sqrt{V_{KT}^{(\lambda)}}}.$$

The above theorem shows that, if we allow for serial correlation in $u_{i,t}$, the KT test can have non-trivial power even in the case of incidental trends. Panel B of Table 3 presents values of k_{KT} for the case that $u_{i,t} = \varepsilon_{i,t} + \theta\varepsilon_{i,t-1}$. This is done for different values of θ and T. As in Panel A, we also consider the case that $\theta = 0$.

The results of Panel B of Table 3 indicate that, for model $M2$, test statistic KT has non-trivial power only if $\theta < 0$. If $\theta = 0$, the test has trivial power, while, for $\theta > 0$ the test is biased. For $\theta < 0$, the power of the test increases with T. For a given T, it becomes bigger if the break point T_0 is located towards the end of the sample, i.e. $\lambda = 0.75$. These results are in contrast to those for model $M1$, presented in Panel A, where the KT test is found to have more power if $\theta > 0$. This can be attributed to the trace terms on the power slope parameter k_{KT} and, in particular, $tr(\Theta^{(\lambda)}\Lambda\Gamma)$ and $tr(\Lambda'\Theta^{(\lambda)}\Gamma)$. Evaluations of these terms show that negative values of θ mitigate the power reduction effects coming from the detrending of the individual panel series. In contrast to model $M1$, this now happens only when $\theta < 0$. To see this more clearly, notice the identity $tr(F'Q^{(\lambda)})+tr(\Lambda'Q^{(\lambda)}\Lambda)+tr(\Lambda'Q^{(\lambda)}) = 0$ which holds for all models. If $p = 0$ we have $tr(\Theta^{(\lambda)}\Lambda)+tr(\Lambda'\Theta^{(\lambda)}) = -tr(\Lambda'Q^{(\lambda)})$ and thus, the numerator of k_{KT} becomes 0. However, if $p > 0$, then $tr(\Theta^{(\lambda)}\Lambda\Gamma) + tr(\Lambda'\Theta^{(\lambda)}\Gamma) \neq -tr(\Lambda'Q^{(\lambda)}\Gamma)$ and thus the numerator is non-zero (see also the proof of Theorems 2 and 4).[7] The above analysis indicates that the power of the KT test can be attributed to the properties of selection matrix $\Theta^{(\lambda)}$, when $p \neq 0$.

To implement Theorems 3 and 4 one must first specify the deterministic matrices $\Psi^{(\lambda)}$ and $\Theta^{(\lambda)}$ which depend on the appropriate order of serial correlation. $\Psi^{(\lambda)}$ is a restricted version of $\Lambda'Q^{(\lambda)}$ while $\Theta^{(\lambda)}$ requires the deterministic selection matrices $M^{(r)}$, for $r = \{1, 2\}$, where all of these quantities are defined above. Then one must calculate estimators $\hat{\Gamma} = (1/N)\sum_{i=1}^{N} \Delta y_i \Delta y_i'$ and $\hat{\Pi} = (1/N)\sum_{i=1}^{N} vec(\Delta y_i \Delta y_i')vec(\Delta y_i \Delta y_i')'$. With these at hand, next we can calculate $\hat{b}^{(\lambda)} = tr(\Psi^{(\lambda)}\hat{\Gamma})$ for Theorem 3 and $\hat{b}^{(\lambda)} = tr(\Theta^{(\lambda)}\hat{\Gamma})$ for Theorem 4. The variances can be calculated accordingly, i.e. $\hat{V}_{KT}^{(\lambda)} = 2tr\left((A_{KT}^{(\lambda)}\hat{\Gamma})^2\right)$, where $A_{KT}^{(\lambda)} = (1/2)(\Lambda'Q^{(\lambda)}+Q^{(\lambda)}\Lambda-\Psi^{(\lambda)}-\Psi^{(\lambda)\prime})$ for Theorem 3 and $\hat{V}_{KT}^{(\lambda)} = vec(Q^{(\lambda)}\Lambda-\Theta^{(\lambda)\prime})'\hat{\Pi}vec(Q^{(\lambda)}\Lambda-\Theta^{(\lambda)\prime})$ for Theorem 4.

It is straightforward to show that $k_{KT} = O(1)$ in the fixed-T case and, because of the

[7] If $p = 0$,

$$\begin{aligned} tr(\Theta^{(\lambda)}\Lambda) &= tr(\Psi^{(\lambda)}\Lambda) - tr(\Psi^{(\lambda)}e^{(1)}e^{(1)\prime})\frac{tr(M^{(1)}\Lambda)}{tr(M^{(1)}e^{(1)}e^{(1)\prime})} - tr(\Psi^{(\lambda)}e^{(2)}e^{(2)\prime})\frac{tr(M^{(2)}\Lambda)}{tr(M^{(2)}e^{(2)}e^{(2)\prime})} = \\ &= -\frac{tr(\Psi^{(\lambda)}e^{(1)}e^{(1)\prime})}{2} - \frac{tr(\Psi^{(\lambda)}e^{(2)}e^{(2)\prime})}{2} \\ &= -\frac{tr(\Lambda'Q^{(\lambda)})}{2}. \end{aligned}$$

because $tr(M^{(j)}e^{(j)}e^{(j)\prime}) = 2tr(M^{(j)}\Lambda)$ and $tr(\Psi^{(\lambda)}\Lambda) = 0$.

required scaling of test statistic KT when T is asymptotic, the test always has trivial power. This is true for both models $M1$ and $M2$, and under serially correlated errors. In fact, the test has zero local power in the $N^{-1/2}T^{-1}$ neighbourhood of unity. This result corresponds to that of Moon and Perron (2004), denoted as MP, considering the case of test statistic KT without breaks and large T. Although the two tests have the same asymptotic local power for large T, they may have different properties in small samples. This can be attributed to the fact that they rely on a different type of bias correction of LS estimator $\hat{\varphi}^{(\lambda)}$. The MP test statistic assumes that error terms $u_{i,t}$ are given as $u_{i,t} = \sum_{j=0}^{\infty} d_{i,j}\varepsilon_{i,t-j}$, subject to usual restrictions. This structure of $u_{i,t}$ results in an asymptotic bias of $\varphi^{(\lambda)}$ which equals a function of the one sided long run variance of $u_{i,t}$, given as $\lambda_{e,i} = \sum_{l=}^{\infty} \sum_{j=0}^{\infty} d_{i,j}d_{i,j+l}$. Estimating this long run variance requires imposing further assumptions that ensure the consistency of the kernel estimators employed.

For test statistic KT, the bias adjustment of $\hat{\varphi}^{(\lambda)}$ relies on selection matrices $\Psi^{(\lambda)}$, $M^{(1)}$ and $M^{(2)}$, which have zero and non-zero diagonals based on the order of serial correlation p. This test does not require an estimate of the long-run variance of $u_{i,t}$, which may be proved problematic in small samples (see, e.g., Moon and Perron (2004)). To see this more clearly, consider the case of model $M1$, where $\Psi^{(\lambda)}$ has its main, its p upper and its p lower diagonals non-zero, catching all the non-zero elements of variance-covariance matrix $\bar{\Gamma}$. Then, note that the not-yet standardized test statistic KT can be written as follows:

$$\begin{aligned}
\hat{\delta}^{(\lambda)}\left(\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{\delta}^{(\lambda)}} - 1\right) &= \hat{\delta}^{(\lambda)}\left(\frac{\sum_{i=1}^{N} y_{i,-1}' Q^{(\lambda)} y_i}{\sum_{i=1}^{N} y_{i,-1}' Q^{(\lambda)} y_{i,-1}} - \frac{\sum_{i=1}^{N} \Delta y_i' \Psi^{(\lambda)} \Delta y_i}{\sum_{i=1}^{N} y_{i,-1}' Q^{(\lambda)} y_{i,-1}} - 1\right) \\
&= \sum_{i=1}^{N} u_i'(\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)}) u_i, \qquad (5)
\end{aligned}$$

where $\left(\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)}\right)$ is a matrix whose main diagonal, its p upper and its p lower diagonals are zero. This means that the quadratic form $u_i'(\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)})u_i$ is the sum of products of the following form: $u_{i,t}u_{i,t+j}$, where $j > p_{\max}$. These products have means and variances which are free from the serial correlation nuisance parameters. Denoting $(\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)}) = C^{(\lambda)}$, with elements $C^{(\lambda)} = [c_{k,j}^{(\lambda)}]$ for $k, j = 1, ..., T$, the test statistic given by (5) can be written as

$$\sum_{i=1}^{N} u_i'(\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)})u_i = \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j \notin \hat{c}} u_{i,t}u_{i,j}c_{j,t}^{(\lambda)},$$

where $\hat{c} = [t - p_{\max}, t + p_{\max}]$ and $c_{j,t}^{(\lambda)}$ are known deterministic quantities, by construction. The limiting distribution of this statistic can be found by applying the central limit theorem to a scaled version of $\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j\notin\hat{c}} u_{i,t}u_{i,j}c_{j,t}^{(\lambda)}$. The following corollary gives the limiting distribution of the non-standardised version of KT test statistic for the cases that: i) $N \longrightarrow \infty$ and ii) N ,$T \longrightarrow \infty$, respectively, based on the above representation of $\sum_{i=1}^{N} u_i'(\Lambda' Q^{(\lambda)} - \Psi^{(\lambda)})u_i$.

Corollary 3 *For model $M1$, let Assumptions A, B and D hold. Then, under null hypothesis H_0: $c = 0$, we have:*

i) as $N \longrightarrow \infty$,

$$\sqrt{N}\hat{\delta}^{(\lambda)}\left(\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{\delta}^{(\lambda)}} - 1\right) \xrightarrow{d} N\left(0, 4\sigma^4 \sum_{t=1}^{T}\sum_{j>t} c_{j,t}^{(\lambda)2}\right)$$

and ii) as N ,$T \longrightarrow \infty$ jointly,

$$T\sqrt{N}\hat{\delta}^{(\lambda)}\left(\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{\delta}^{(\lambda)}} - 1\right) \xrightarrow{d} N\left(0, p\lim_{N,T}\left[4\sigma^4 \sum_{t=1}^{T}\sum_{j>t} c_{j,t}^{(\lambda)2}/T^2\right]\right).$$

The results of this corollary show that the unknown nuisance long run variances, which appear in the Moon and Perron (2004) version of the test statistic, are replaced by the known weights $c_{j,t}^{(\lambda)}$ in test statistic KT. These weights are functions of known quantities, which depend on the individual deterministic components of panel series $y_{i,t}$. This feature of the tests skips the problem of estimating the long run variance which is a difficult econometric task in small samples, as noted above.

5 Common factors

In this section, we extend the HT and KT test statistics to allow for cross sectional dependence in error terms $u_{i,t}$ taking the form of common factors. The assumption of cross section independence may be restrictive in panel data macroeconomic studies (see, e.g., Sarafidis and Wansbeek (2012)).

There are only a few studies examining the effect of common factors on the local power of unit root tests. Hansen (1995), in the single time series literature, considers additional exogenous covariates which lead to more powerful unit root tests. On the other hand, Moon and Perron (2004), who examine the local power of a large-T version of the KT test, find that power is unaffected by the presence of common factors in $u_{i,t}$. In our analysis, we consider the common factors to be known (observed). This is without loss of generality, as our results would be qualitatively the same even if the common factors had to be estimated in a first step as in Moon and Perron (2004). Our aim is to explore the impact of cross section dependence on the power of test statistics HT and KT.

Consider the following specifications of models $M1$ and $M2$ including a single common factor f :

$$\tilde{M}1: \quad y_i = a_i^{(1)} e^{(1)} + a_i^{(2)} e^{(2)} + \zeta_i, \qquad i = 1, 2, .., N, \quad \text{and}$$

$$\tilde{M}2: \quad y_i = a_i^{(1)} e^{(1)} + a_i^{(2)} e^{(2)} + \beta_i^{(1)} \tau^{(1)} + \beta_i^{(2)} \tau^{(2)} + \zeta_i,$$

with

$$\zeta_i = \varphi \zeta_{i,-1} + \varepsilon_i,$$

and

$$\varepsilon_i = \xi_i^{(1)} f^{(1)} + \xi_i^{(2)} f^{(2)} + u_i,$$

where $f^{(1)} = f_t^{(1)}$ if $t \leq T_0$, and zero otherwise, and $f^{(2)} = f_t^{(2)}$ if $t > T_0$, and zero otherwise.[8] That is, we assume that there is a common factor in errors $u_{i,t}$ which also undergoes a structural break at the same time. For $\xi_i^{(1)} = \xi_i^{(2)}$, both models $\tilde{M}1$ and $\tilde{M}2$ can consider the case that there is no break in the common factor process. Also note that the assumption that there is only one common factor is not restrictive, and it is made only for ease of exposition. A more general specification would be $\boldsymbol{F}^{(1)} \boldsymbol{\xi}_i^{(1)} + \boldsymbol{F}^{(2)} \boldsymbol{\xi}_i^{(2)}$, where $\boldsymbol{\xi}_i^{(j)} = (\xi_{1,i}^{(j)}, ..., \xi_{K,i}^{(j)})'$ is a $K \times 1$ vector of factor loadings and $\boldsymbol{F}^{(j)} = (f_1^{(j)}, ..., f_K^{(j)})$ is a $T \times K$ matrix of K observed factors, for $j = 1, 2$.

The reduced form of the above models can be written as follows:

$$\tilde{M}1 \quad : \quad y_i = \varphi y_{i,-1} + (1-\varphi)(a_i^{(1)} e^{(1)} + a_i^{(2)} e^{(2)}) + \xi_i^{(1)} f^{(1)} + \xi_i^{(2)} f^{(2)} + u_i, \quad \text{and}$$
$$\tilde{M}2 \quad : \quad y_i = \varphi y_{i,-1} + \sum_{j=1}^{2} \left[(1-\varphi) a_i^{(j)} e^{(j)} + \varphi \beta_i^{(j)} e^{(j)} + (1-\varphi) \beta_i^{(j)} \tau^{(j)} + \xi_i^{(j)} f^{(j)} \right] + u_i.$$

Define $Q^{(\lambda)} = I - X^{(\lambda)} \left(X^{(\lambda)\prime} X^{(\lambda)} \right)^{-1} X^{(\lambda)\prime}$, where $X^{(\lambda)} = \left(e^{(1)}, e^{(2)}, f^{(1)}, f^{(2)}, \Lambda f^{(1)}, \Lambda f^{(2)} \right)$ for model $\tilde{M}1$ and $X^{(\lambda)} = \left(e^{(1)}, e^{(2)}, \tau^{(1)}, \tau^{(2)}, f^{(1)}, f^{(2)}, \Lambda f^{(1)}, \Lambda f^{(2)} \right)$ for model $\tilde{M}2$. For our

[8] In the presentation of our results we assume that there is a break in the factor. We do so to maintain the focus of the structural break in this section as well. Equally, without loss of generality and with the same results, we could have considered two factors instead of a broken one. Breitung and Eickmeier (2011) provide the intuition that structural breaks severely inflate the number of factors.

asymptotic results, next we make the following assumption.

Assumption E: (e1) $\xi_i^{(1)}$ and $\xi_i^{(2)}$ are sequences of independent random variables with finite $4+\delta$ moments. They are also independent from u_i. (e2) $\lim_N \max(E(\xi_i^{(j)2}))/(N\bar{\xi}^{(j)2}) = 0$, where $\bar{\xi}^{(j)2} = (1/N)\sum_{i=1}^{N} E(\xi_i^{(j)2})$, for $j = 1, 2$. Also, $\xi^{(j)2} = \lim_N \bar{\xi}^{(j)2}$ is finite. (e3) $f^{(1)}$ and $f^{(2)}$ are $T \times 1$ finite, non-random vectors.

Assumption F: (f1) $T > \mathrm{col}(X^{(\lambda)})$, where $\mathrm{col}(\cdot)$ denotes the column dimension of a matrix. (f2) Product matrix $X^{(\lambda)\prime}X^{(\lambda)}$ which appears in $Q^{(\lambda)}$ is invertible. (f3) Variance function V is non-zero.

Conditions (e1) and (e2) guarantee that $\xi_i^{(j)}$ obey the Lindeberg-Feller CLT and condition (e3) states that the common factor can be seen as another type of deterministic component, which is a common approach in the large-N - fixed-T panel data literature (see Sarafidis and Wansbeek (2012)). Note that condition (e3) is weaker than that made by Moon and Perron (2004), for their large T test. This is because in the panel data factor models, the values of a common factor variable f are treated like a set of parameters which are removed from the model, just like the individual intercepts and the individual linear trends. We therefore need not assume them linear or restrict their order of integration.

Assumption F determines the maximum number of common factors and the position of the breaks. Condition (f1) puts a limit to the number of factors that can appear in models $\tilde{M}1$ and $\tilde{M}2$. This is common in the literature (see, e.g., Sarafidis and Wansbeek (2012)). Violation of this assumption may lead to an invertible $X^{(\lambda)}$ which will result in $Q^{(\lambda)} = 0$. Condition (f2) also guarantees that $Q^{(\lambda)}$ exists. If there is no serial correlation, the above conditions are sufficient for the application of the HT and KT test statistics. In the presence of serial correlation in $u_{i,t}$, a case relevant only for the KT test, condition (f3) must be satisfied as well. Since serial correlation in $u_{i,t}$ limits the available moments for estimation, condition (f3) guarantees that there are sufficient observations before and after the break for the identification of nuisance parameters $\bar{\beta}^{(j)2}$ and $\bar{\xi}^{(j)2}$, for $j = 1, 2$. We do not need to assume that the variance function of the limiting distribution of the adjusted for its inconsistency LS estimator $\hat{\varphi}^{(\lambda)}$ is known. An easy way to check this condition will be presented below. Overall, Assumption F is more general than Assumptions B and D. Assumptions E and F accommodate both models $\tilde{M}1$ and $\tilde{M}2$. The following theorem derives the limiting distribution of statistic HT under the sequence of local alternatives $\varphi_N = 1 - c/\sqrt{N}$ for model $\tilde{M}1$, as N diverges to infinity.

Theorem 5 *Let Assumptions A, E and F hold for model $\tilde{M}1$ and $u_i \sim NIID(0, \sigma^2 I)$. Then, under $\varphi_N = 1 - c/\sqrt{N}$, we have*

$$V_{HT}^{(\lambda)-1/2}\sqrt{N}(\hat{\varphi}^{(\lambda)} - 1 - B(\lambda)) \xrightarrow{d} N(-ck_{HT}, 1)$$

as $N \longrightarrow \infty$, with

$$k_{HT} = \frac{tr(F'Q^{(\lambda)}) + tr(\Lambda' Q^{(\lambda)}\Lambda) - 2B(\lambda)tr(F'Q^{(\lambda)}\Lambda)}{\sqrt{2tr(A_{HT}^{(\lambda)^2})}}$$

where $B(\lambda) = p\lim_N(\hat{\varphi}^{(\lambda)} - 1) = tr(\Lambda' Q^{(\lambda)})/tr(\Lambda' Q^{(\lambda)}\Lambda)$, $V_{HT}^{(\lambda)} = 2tr(A_{HT}^{(\lambda)2})/tr(\Lambda' Q^{(\lambda)}\Lambda)^2$, with $A_{HT}^{(\lambda)} = \frac{1}{2}(\Lambda' Q^{(\lambda)} + Q^{(\lambda)}\Lambda) - B(\lambda)(\Lambda' Q^{(\lambda)}\Lambda)$, and F is defined in the Appendix.

To calculate the value of k_{HT}, given by the above theorem, we have considered various types of processes for $f^{(1)}$ and $f^{(2)}$. Panel A of Table 4 contains the average k_{HT}, denoted $\bar{k}_{HT}$ for 5000 realizations, if f_t is generated by process $f_t = \rho f_{t,-1} + \eta_t$, where $\rho = 0.8$ and $\eta_t \sim NIID(0, 1)$. We see that the HT test statistic has reasonable power. This power is however lower than that for model $M1$, without a common factor f. An explanation of this result could be that the existence of common factor f reduces variation of the individual series of the panel, and thus the information available in the sample. Mathematically, for the HT test, less power comes from the large dimension of $X^{(\lambda)}$; every factor increases it by two and a broken factor, like in our case, increases it by four.

Summing up, our findings indicate that the existence of a factor f in the error terms $u_{i,t}$ is what determines the magnitude of power of statistic HT, for model $\tilde{M}1$. This is verified by using as common factors f various processes, even non-stationary ones.

The limiting distribution of statistic HT under $\varphi_N = 1 - c/\sqrt{N}$ for model $\tilde{M}2$, where $X^{(\lambda)} = \left(e^{(1)}, e^{(2)}, \tau^{(1)}, \tau^{(2)}, f^{(1)}, f^{(2)}, \Lambda f^{(1)}, \Lambda f^{(2)}\right)$, is given next. This theorem shows that the power of the HT test in the case of incidental trends remains trivial. The presence of common factor f does not change this result.

Theorem 6 *Let Assumptions A, E and F hold for model $\tilde{M}2$ and $u_i \sim NIID(0, \sigma^2 I)$. Then, under $\varphi_N = 1 - c/\sqrt{N}$, we have*

$$V_{HT}^{(\lambda)-1/2}\sqrt{N}(\hat{\varphi}^{(\lambda)} - 1 - B(\lambda)) \xrightarrow{d} N\left(-ck_{HT}, 1\right)$$

as $N \longrightarrow \infty$, with

$$k_{HT} = 0,$$

where $B(\lambda) = p\lim(\hat{\varphi}^{(\lambda)} - 1) = tr(\Lambda' Q^{(\lambda)})/tr(\Lambda' Q^{(\lambda)}\Lambda)$, $V_{HT}^{(\lambda)} = 2tr(A_{HT}^{(\lambda)2})/tr(\Lambda' Q^{(\lambda)}\Lambda)^2$, with $A_{HT}^{(\lambda)} = \frac{1}{2}(\Lambda' Q^{(\lambda)} + Q^{(\lambda)}\Lambda) - B(\lambda)(\Lambda' Q^{(\lambda)}\Lambda)$, and F is defined in the Appendix.

To study the effects of cross-section dependence (presence of common factor f) on the power of test statistic KT, we consider the more general version of it allowing for serial correlation in error terms $u_{i,t}$. Before presenting our main results, next we make all necessary definitions to derive the liming distribution of the KT test statistic for models $\tilde{M}1$ and $\tilde{M}2$. For model $\tilde{M}1$, the inconsistency of $\hat{\varphi}^{(\lambda)}$ is given as $tr(\Lambda' Q^{(\lambda)}\Gamma)/tr(\Lambda' Q^{(\lambda)}\Lambda\Gamma)$, where $Q^{(\lambda)}$ is based on $X^{(\lambda)} = \left(e^{(1)}, e^{(2)}, f^{(1)}, f^{(2)}, \Lambda f^{(1)}, \Lambda f^{(2)}\right)$. The estimator $\hat{\Gamma}$ is inconsistent, with $p\lim_N \left[\hat{\Gamma} - \bar{\Gamma} - \bar{\xi}^{(1)2} f^{(1)} f^{(1)\prime} - \bar{\xi}^{(2)2} f^{(2)} f^{(2)\prime}\right] = 0$, since under null hypothesis H_0: $c = 0$ we have $\Delta y_i = \xi_i^{(1)} f^{(1)} + \xi_i^{(2)} f^{(2)} + \varepsilon_i$. To adjust $\hat{\Gamma}$ and, hence, $\hat{\varphi}^{(\lambda)}$ for their inconsistency, we will rely on selection matrices $M_f^{(j)} = f^{(j)} f^{(j)\prime} - diag(f^{(j)} f^{(j)\prime}, p)$, for $j = 1, 2$. These enable us to identify nuisance parameters $\xi^{(j)}$, as $p\lim_N \left[tr(M_f^{(j)}\hat{\Gamma})/tr(M_f^{(j)} f^{(j)} f^{(j)\prime}) - \bar{\xi}^{(j)2}\right] = 0$, for $j = 1, 2$. Given matrix $M_f^{(j)}$, selection matrix $\Theta^{(\lambda)}$ now becomes:

$$\Theta^{(\lambda)} = \Psi^{(\lambda)} - \frac{tr(\Psi^{(\lambda)} f^{(1)} f^{(1)\prime})}{tr(M_f^{(1)} f^{(1)} f^{(1)\prime})} M_f^{(1)} - \frac{tr(\Psi^{(\lambda)} f^{(2)} f^{(2)\prime})}{tr(M_f^{(2)} f^{(2)} f^{(2)\prime})} M_f^{(2)},$$

with $p\lim_N \left[tr(\Theta^{(\lambda)}\hat{\Gamma}) - tr(\Lambda' Q^{(\lambda)}\bar{\Gamma})\right] = 0$. Statistic $\sqrt{N}\left(\hat{\varphi}^{(\lambda)} - \hat{b}^{(\lambda)}/\hat{\delta}^{(\lambda)} - 1\right)$, where $\hat{b}^{(\lambda)} = tr(\Theta^{(\lambda)}\hat{\Gamma})$, is centred around 0 and its variance is given as $V_{KT}^{(\lambda)} = vec(Q^{(\lambda)}\Lambda - \Theta^{(\lambda)\prime})'\Pi vec(Q^{(\lambda)}\Lambda - \Theta^{(\lambda)\prime})$, with $\hat{\Pi} = (1/N)\sum_{i=1}^{N} vec(\Delta y_i \Delta y_i')vec(\Delta y_i \Delta y_i')'$.

For model $\tilde{M}2$**,** matrix $Q^{(\lambda)}$ is based on $X^{(\lambda)} = (e^{(1)}, e^{(2)}, \tau^{(1)}, \tau^{(2)}, f^{(1)}, f^{(2)}, \Lambda f^{(1)}, \Lambda f^{(2)})$. $\hat{\Gamma}$ is an inconsistent estimator of $\bar{\Gamma}$ (due to the presence of nuisance parameters $\beta_i^{(j)}$ and $\xi_i^{(j)}$, for $j = 1, 2$) with $p\lim_N \left[\hat{\Gamma} - \bar{\Gamma} - \bar{\beta}^{(1)2} e^{(1)} e^{(1)\prime} - \bar{\beta}^{(2)2} e^{(2)} e^{(2)\prime} - \bar{\xi}^{(1)2} f^{(1)} f^{(1)\prime} - \bar{\xi}^{(2)2} f^{(2)} f^{(2)\prime}\right] = 0$, since under H_0: $c = 0$ we have $\Delta y_i = \beta_i^{(1)} e^{(1)} + \beta_i^{(2)} e^{(2)} + \xi_i^{(1)} f^{(1)} + \xi_i^{(2)} f^{(2)} + \varepsilon_i$. To adjust $\hat{\Gamma}$ for its inconsistency due to nuisance parameters $\beta_i^{(j)}$ and $\xi_i^{(j)}$, we will rely on selection matrices $M^{(j)} = e^{(j)} e^{(j)\prime} - diag\{e^{(j)} e^{(j)\prime}, p\}$ and $M_f^{(j)} = f^{(j)} f^{(j)\prime} - diag(f^{(j)} f^{(j)\prime}, p)$, which imply $p\lim_N \left[tr(M^{(j)}\hat{\Gamma})/tr(M^{(j)} e^{(j)} e^{(j)\prime}) - \bar{\beta}^{(j)2}\right] = 0$ and $p\lim_N \left[tr(M_f^{(j)}\hat{\Gamma})/tr(M_f^{(j)} f^{(j)} f^{(j)\prime}) - \bar{\xi}^{(j)2}\right] = 0$. Given these definitions, selection matrix $\Theta^{(\lambda)}$ now becomes

$$\Theta^{(\lambda)} = \Psi^{(\lambda)} - \sum_{j=1}^{2} \frac{tr(\Psi^{(\lambda)} e^{(j)} e^{(j)\prime})}{tr(M^{(j)} e^{(j)} e^{(j)\prime})} M^{(j)} - \sum_{j=1}^{2} \frac{tr(\Psi^{(\lambda)} f^{(j)} f^{(j)\prime})}{tr(M_f^{(j)} f^{(j)} f^{(j)\prime})} M_f^{(j)},$$

with $p\lim_N \left[tr(\Theta^{(\lambda)}\hat{\Gamma}) - tr(\Lambda' Q^{(\lambda)}\bar{\Gamma})\right] = 0$, and then the appropriate definitions of $\hat{b}^{(\lambda)}$ and $V_{KT}^{(\lambda)}$ will apply as above.[9] The next theorem gives the limiting distributions of tests statistic KT for models $\tilde{M}1$ and $\tilde{M}2$ assuming serially correlated $u_{i,t}$.

Theorem 7 *Let Assumptions A, E, and F hold for models $\tilde{M}1$ and $\tilde{M}2$. Then, under $\varphi_N = 1 - c/\sqrt{N}$ we have*

$$V_{KT}^{(\lambda)-1/2}\hat{\delta}^{(\lambda)}\sqrt{N}\left(\hat{\varphi}^{(\lambda)} - \frac{\hat{b}^{(\lambda)}}{\hat{\delta}^{(\lambda)}} - 1\right) \xrightarrow{d} N(-ck_{KT}, 1),$$

as $N \to \infty$, with

$$k_{KT} = \frac{tr(F'Q^{(\lambda)}\Gamma) + tr(\Lambda' Q^{(\lambda)}\Lambda\Gamma) - tr(\Theta^{(\lambda)}\Lambda\Gamma) - tr(\Lambda'\Theta^{(\lambda)}\Gamma)}{\sqrt{V_{KT}^{(\lambda)}}},$$

where matrices $Q^{(\lambda)}$, $\Theta^{(\lambda)}$ are appropriately specified for each model.

Panels A and B of Table 4 present average values of k_{KT}, denoted $\bar{k}_{KT}$, for models $\tilde{M}1$ and $\tilde{M}2$, respectively, when errors terms $u_{i,t}$ are IID, over 5000 repetitions. As before, f_t is generated as $f_t = \rho f_{t,-1} + \eta_t$, where $\rho = 0.8$ and $\eta_t \sim NIID(0,1)$. The results of Panel A indicate that the presence of common factor f leads to power reduction for model $\tilde{M}1$. The values of $\bar{k}_{KT}$ are all positive, but smaller than the case without common factor f. In contrast, Panel B indicates that the inclusion of factor f in model $\tilde{M}2$ leads to non-trivial power of the KT test. This result was rather expected after the findings of Section 4, which considers the case of the KT test allowing for serial correlation in $u_{i,t}$. The power of the test

[9] To check condition (f3) it is sufficient to check that the denominators in $\Theta^{(\lambda)}$, which are based on quantities known to the researcher, are different than 0, i.e. $tr(M^{(j)}e^{(j)}e^{(j)\prime}) \neq 0$ and $tr(M_f^{(j)}f^{(j)}f^{(j)\prime}) \neq 0$. These denominators represent the number of elements the selection matrices $M^{(j)}$ and $M_f^{(j)}$ choose so that they estimate $\bar{\beta}^{(j)2}$ and $\bar{\xi}^{(j)2}$. If they are equal to zero, this means that there are zero elements available for $M^{(j)}$ and $M_f^{(j)}$ and therefore the corresponding $\bar{\beta}^{(j)2}$ and $\bar{\xi}^{(j)2}$ cannot be identified.

can be attributed to the interaction between individual trends and common factors in the bias adjustment of the LS estimator $\hat{\varphi}^{(\lambda)}$. This result is notably different than that in the large-T case, where the test is robust to the effects of common factors (see, e.g., Moon and Perron (2004)).

To implement Theorems 5, 6 and 7 one must first specify the annihilator matrix $Q^{(\lambda)}$ which contains factors $f^{(1)}$, $f^{(2)}$, i.e. for Theorem 7, $Q^{(\lambda)} = I - X^{(\lambda)} \left(X^{(\lambda)\prime} X^{(\lambda)} \right)^{-1} X^{(\lambda)\prime}$ where $X^{(\lambda)} = \left(e^{(1)}, e^{(2)}, \tau^{(1)}, \tau^{(2)}, f^{(1)}, f^{(2)}, \Lambda f^{(1)}, \Lambda f^{(2)} \right)$. Then, the HT test can be applied as before. For the KT test, the steps are also similar to those of the previous section, but care must be taken to appropriately specify $\Theta^{(\lambda)}$ because it is influenced by both the number of factors and by the order of serial correlation. It must be made sure that $\Theta^{(\lambda)}$ exists as described in footnote 9. $\hat{\Gamma}$ and $\hat{\Pi}$ can be estimated as in the previous sections.

Summing up, the results of this section indicate that cross section correlation in error terms $u_{i,t}$ affects the power performance of the HT and KT test statistics, if T is fixed. The tests are not robust to the presence of a common factor in $u_{i,t}$, as in the large-T case. For the large-T case, it can be easily seen that test statistics HT and KT have zero local power, for both models $\tilde{M}1$ and $\tilde{M}2$.

6 Monte Carlo results

In this section, we conduct a Monte Carlo study to examine if the asymptotic local power functions of the HT and KT test statistics, implied by the results of the previous section, provide good approximations of their small sample ones. This is done based on 5000 repetitions and for different values of N and T, often met in microeconomic and macroeconomic studies, i.e., $N = \{100, 300, 1000\}$ and $T = \{8, 10, 12, 15, 20, 50, 100, 200, 300\}$. For each iteration, we calculate the size of the tests at 5% level (i.e., for $c = 0$) and their power (i.e., for $c = 1$). This is done separately for the cases that $u_{i,t} \sim NIID(0, 1)$ and $u_{i,t} = \varepsilon_{i,t} + \theta\varepsilon_{i,t-1}$, with $\theta \in \{-0.8, -0.5, 0, 0.5, 0.8\}$. The nuisance parameters of the models are set to the following values: $y_{i,0} = 0$, $a_i^{(j)} = 0$ and $\beta_i^{(j)} = 0$, for all i.

Table 5 presents the results of our simulation study for the case that $u_{i,t} \sim NIID(0, 1)$. The last column of the tables gives the theoretical values (TV) of the power function and the nominal size of the tests, at $a = 5\%$. For model $M1$, the results indicate that both the HT and KT tests have size and power values which are very close to their theoretical ones. Furthermore, the results confirm that the HT test has more power towards the beginning and the end of the sample, while the KT test has more power in the middle. As was also predicted by the theory, the HT test has higher power than the KT test. The small sample power of this test is very close to that predicted by its asymptotic local power function (see column TV) even for small N (e.g., $N = 100$). However, this is not always true for the KT test, which needs very high N in order its power to converge to its theoretical value. For model $M2$, the results of Table 5 indicate that, for large N, both HT and KT tests have trivial power, as it was expected. However, in small samples (e.g., $N = 100$), both tests have some non-trivial power. This can be obviously attributed to second, or higher, order effects of the true power function, which cannot be approximated by the first-order approximation considered in our analysis. Note that, for model $M2$, the KT test has slightly higher small sample power than the HT.

Tables 6 and 7 present the results of our simulation study for the KT test statistic allowing for serial correlation in error terms $u_{i,t}$, assuming $u_{i,t} = \varepsilon_{i,t} + \theta\varepsilon_{i,t-1}$. This is done for models $M1$ and $M2$, and $T \in \{8, 10\}$. The maximum order of serial correlation allowed by the KT test is set to $p = 1$, which matches that of the MA process of $u_{i,t}$. The results of these tables are also consistent with theory. For model $M1$, the KT test has significant power when $\theta > 0$. As N increases, this power converges quite fast to its theoretical value, reported in the last column of the table. Note that both the theoretical and small sample values of the power function of the KT test statistic are higher than their corresponding values in the absence of serial correlation (see Table 5). This is also consistent with the theory. It can be attributed to the serial correlation effects of $u_{i,t}$ on the power function of the test, discussed in the previous section. For negative values of θ, the test has also significant power. This happens for $\lambda = \{0.75\}$, as was predicted by the theory.

For model $M2$, the results of Table 7 indicate that the KT test statistic has smaller power than for model $M1$. As was expected by the theory, the power of the test is non-trivial if $\theta < 0$. The KT test has also some small sample power if $\theta > 0$, which qualifies its use in practice. As was argued before, this power can be attributed to second, or higher, order effects of the true power function, which are not captured by our asymptotic approximations. Finally, another conclusion which can be drawn from the results of our simulation study reported in Tables 6 and 7 is that, when $\theta < 0$, a break towards the end of the sample increases the power of the KT test. When $\theta > 0$, the power of the test is maximized at the middle of the sample. These results apply to both models $M1$ and $M2$. They are also consistent with the theoretical results reported in Table 3.

Table 8 presents the results of the tests for models $\tilde{M}1$ and $\tilde{M}2$ including a common factor in error terms. This factor f_t is generated as before (see Section 5), i.e., $f_t = \rho f_{t,-1} + \eta_t$, with $\rho = 0.8$ and $\eta_t \sim NIID(0,1)$. The slope coefficients of this factor $\xi_i^{(j)}$ are set to zero, i.e., $\xi_i^{(j)} = 0$, for all i and j. The results of the table clearly indicate that both the size and power of the HT and KT test statistics are close to their theoretical values, for all cases of N, T and λ considered. For model $\tilde{M}1$ both tests have non-trivial power, while for model

$\tilde{M}2$ only the KT test has non-trivial power, which is in accordance to our theoretical rsults of Section 5.

Finally, Table 9 shows how good the large-T approximations the HT and KT tests statistics, given by Corollaries 1 and 2, are for a variety of different values of N and T. In particular, we consider the following cases of N and T: $N = \{10, 20, 50\}$ and $T = \{50, 100, 200, 300\}$, often used in macroeconomic studies. The results of the table indicate that, for model $M1$, the HT test statistic is a bit oversized when T is much larger than N. For instance, the size of the test is equal to 0.07, for $N = 10$, $T = 300$ and $\lambda = 0.5$. As N increases, the size of the test tends to its nominal value of 0.05. The power of the test is also close to its theoretical values, with the approximation becoming better as N increases. The KT test statistic displays similar behaviour to that of HT, but it is a bit undersized for large T and very small N. Consistently with the theory, the KT test does not have power anywhere. For model $M2$, this result holds for both HT and KT test statistics. When T becomes large, both the HT and KT tests have trivial power. The quality of the approximations seems to be unaffected by the relative position of the break in the sample.

7 Conclusions

This paper analyses the asymptotic local power properties of least-squares based fixed-T panel unit root tests allowing for a structural break in the deterministic components of the AR(1) panel data model, namely its individual effects and/or slope coefficients of its individual linear (incidental) trends. This is done by assuming that the cross-section dimension of the panel data models (N) grows large.

The paper derives the limiting distributions of the panel unit root test proposed in Karavias and Tzavalis (2014a) (denoted HT) under local alternatives. This is studied together with the test of Karavias and Tzavalis (2014b), denoted as KT, which allows for a structural break and serial correlation in the error terms of the AR(1) panel data model. In this paper, we have extended the above tests to allow for cross section dependence. Both of these tests are based on the least squares dummy variables estimator of the autoregressive coefficient of the AR(1) panel data model, which is corrected for its inconsistency due to the deterministic components of the panel, the cross section dependence and/or serial correlation effects of the error term.

The results of the paper lead to a number of conclusions. First, they show that, for the standard AR(1) panel data model with white noise error terms and individual effects, both the HT and KT tests have significant asymptotic local power. The HT test has much higher power than the KT test. This can be attributed to the fact that, in order to adjust for the inconsistency of the least squares estimator, the KT test requires consistent estimation of the variance of the error term. The HT test does not depend on this nuisance parameter, as it adjusts the least squares estimator for both the inconsistency of its numerator and denominator, and thus the variance of the error terms is cancelled out. The HT test is found to have more power when the break is towards the beginning or the end of the sample, while the KT test has more power when the break is towards the middle of the sample.

Second, both the HT and KT tests have asymptotically trivial power in the case that the AR(1) allows also for incidental trends. The allowance for a common break in the slope coefficients of the incidental trends does not change the behaviour of the tests. This problem does not always exist for the KT test extended for serial correlation of the error terms. In this case, the paper presents circumstances that the KT test has non-trivial power. In particular, this happens when the error terms follow a MA(1) procedure with negative serial correlation. The power of the KT in this case can be attributed to the effects of the serial correlation of error term on the adjustment of the least squares estimator of the autoregressive coefficient for its inconsistency, upon which the KT test is based on. In contrast to large-T panel data unit root tests, the power function of fixed-T tests depends on the values of nuisance parameters capturing serial correlation effects which can affect the asymptotic (over N) power of the tests.

Third, we find that the existence of a common factor in the error terms changes the behaviour of the tests. For the model with intercepts, the presence of the common factor reduces the power of both the HT and KT tests. For the model with trends, the HT has trivial power, while the KT has positive. Fourth, we compare our results to those of the large-T literature and we show that the desirable properties of the KT test presented in this paper, i.e. non-trivial power for the case with incidental trends, are derived from the fixed-T estimator of the bias of the within groups estimator.

The above results are confirmed through a Monte Carlo simulation exercise. This exercise has shown that the empirical probabilities of rejection are very close to their theoretical values, which means that the asymptotic theory provides a good approximation of small sample results of fixed-T panel data unit roots. The above findings suggest that there are several theoretical arguments in favour of the use of fixed T tests in practice, especially in the case where incidental trends are in the model.

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, Econometric Society, Econometric Society, vol. 61(4), p 821-56.

[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, M., 2003. Panel data econometrics. Oxford University Press.

[5] Bai, J., 2010. Common breaks in means and variances for panel data. Journal of Econometrics, Elsevier, vol. 157(1), pages 78-92.

[6] Bai J., Carrion-I-Silvestre, J.L., 2009. Structural Changes, Common Stochastic Trends, and Unit Roots in Panel Data. Review of Economic Studies, vol. 76(2), 471-501.

[7] Baltagi, B., 2008, Econometric analysis of panel data. John Wiley & Sons.

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

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

[10] Chan, F., Pauwels, L.L., 2011. Model specification in panel data unit root tests with an unknown break. Mathematics and Computers in Simulation. 81, 1299–1309.

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

[12] Hadri K., Larsson, R., Rao, Y., 2012. Testing for stationarity with a break in panels where the time dimension is finite. Bulletin of Economic Research, 64, s123-s148.

[13] Hansen, B., E., 1995. Rethinking the Univariate Approach to Unit Root Testing: Using Covariates to Increase Power. Econometric Theory, Cambridge University Press, vol. 11(05), pages 1148-1171.

[14] Harris D., Harvey D., Leybourne S., and Sakkas N., 2010. Local asymptotic power of the Im-Pesaran-Shin panel unit root test and the impact of initial observations. Econometric Theory 26, 311-324.

[15] Harris, D., Leybourne, S., and McCabe, B., 2005. Panel Stationarity Tests for Purchasing Power Parity with Cross-Sectional Dependence. Journal of Business & Economic Statistics, vol. 23, 395-409.

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

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

[18] Harvey, D.I., Leybourne S.J., 2005. On testing for unit roots and the initial observation. Econometrics Journal 8, 97–111.

[19] Karavias, Y., Tzavalis, E., 2014a. Testing for unit roots in short panels allowing for a structural break. Computational Statistics & Data Analysis, Elsevier, vol. 76(C), pages 391-407.

[20] Karavias, Y., and Tzavalis, E., 2014b. Testing for unit roots in panels with structural changes, spatial and temporal dependence when the time dimension is finite. Granger Centre Discussion Paper Series, No 14/03.

[21] Karavias, Y., Tzavalis, E., 2014c. A fixed-T version of Breitung's panel data unit root test. Economics Letters, Elsevier, vol. 124(1), pages 83-87.

[22] 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.

[23] 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.

[24] Madsen E., 2010. Unit root inference in panel data models where the time-series dimension is fixed: a comparison of different tests. Econometrics Journal 13, 63-94.

[25] Meligkotsidou, L., Tzavalis, E., Vrontos I.D., 2011. A Bayesian analysis of unit roots and structural breaks in the level, the trend and the error variance of autoregressive models of economic series. Econometric Reviews, 30 (2), 208-249.

[26] Moon, H.R., Perron, B., 2008. Asymptotic local power of pooled t-ratio tests for unit roots in panels with fixed effects. Econometrics Journal 11, 80-104.

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

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

[29] Sarafidis, V., Wansbeek, T., 2012. Cross-Sectional Dependence in Panel Data Analysis. Econometric Reviews, Taylor & Francis Journals, Taylor & Francis Journals, vol. 31(5), pages 483-531.

7.1 Appendix

In this appendix, we provide proofs of the theorems and the corollary presented in the main text of the paper.

Proof of Theorem 1: First, we derive the limiting distribution of the HT test statistic, under the sequence of local alternatives $\varphi_N = 1 - c/\sqrt{N}$. Define vector $w = (1, \varphi_N, \varphi_N^2, ..., \varphi_N^{T-1})'$ and matrix

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

Under null hypothesis H_0: $c = 0$, we have $\Omega = \Lambda$. The first order Taylor expansions of Ω and w yields

$$\Omega = \Lambda + F(\varphi_N - 1) + o_p(1) \text{ and} \tag{6}$$

$$w = e + f(\varphi_N - 1) + o_P(1), \tag{7}$$

respectively, where $F = (d\Omega/d\varphi_N) \mid_{c=0}$ and $f = (dw/d\varphi_N) \mid_{c=0}$. Based on the above definitions of w and Ω, vector $y_{i,-1}$ can be written as

$$y_{i,-1} = wy_{i,0} + \Omega X^{(\lambda)}\gamma_i^{(\lambda)} + \Omega u_i, \tag{8}$$

where $\gamma_i^{(\lambda)} = (a_i^{(1)}(1-\varphi_N), a_i^{(2)}(1-\varphi_N))' = (1-\varphi_N)(a_i^{(1)}, a_i^{(2)})'$. Using last relationship of $y_{i,-1}$, the HT test statistic for model $M1$ can be written under $\varphi_N = 1 - c/\sqrt{N}$ as follows:

$$\begin{aligned}
&\sqrt{N}(\hat{\varphi}^{(\lambda)} - \varphi_N - B(\lambda)) && (9)\\
&= \sqrt{N}\left(\frac{\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1}Q^{(\lambda)}(\varphi_N y_{i,-1} + X^{(\lambda)}\gamma_i^{(\lambda)} + u_i)}{\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1}Q^{(\lambda)}y_{i,-1}} - \varphi_N - B(\lambda)\right),\\
&= \sqrt{N}\left(\frac{\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1}Q^{(\lambda)}u_i}{\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1}Q^{(\lambda)}y_{i,-1}} - B(\lambda)\frac{\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1}Q^{(\lambda)}y_{i,-1}}{\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1}Q^{(\lambda)}y_{i,-1}}\right),\\
&= \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N} y'_{i,-1}Q^{(\lambda)}u_i - \frac{1}{\sqrt{N}}B(\lambda)\sum_{i=1}^{N} y'_{i,-1}Q^{(\lambda)}y_{i,-1}}{\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1}Q^{(\lambda)}y_{i,-1}} = \frac{(A)-(B)}{(C)}. && (10)
\end{aligned}$$

Next, we derive asymptotic results of each of quantities (A), (B) and (C), defined by (10).

Substituting (8) in (A), we have

$$
\begin{aligned}
(A) \quad \equiv \quad & \frac{1}{\sqrt{N}} \sum_{i=1}^{N} y_{i,-1}' Q^{(\lambda)} u_i = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(y_{i,0} w' + \gamma_i^{(\lambda)\prime} X^{(\lambda)\prime} \Omega' + u_i' \Omega' \right) Q^{(\lambda)} u_i \\
= \quad & \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(y_{i,0} w' Q^{(\lambda)} u_i + \gamma_i^{(\lambda)\prime} X^{(\lambda)\prime} \Omega' Q^{(\lambda)} u_i + u_i' \Omega' Q^{(\lambda)} u_i \right)
\end{aligned}
$$

Using relationships (6)-(7), we can find the following limits of the summands entering into the last relationship of (A). First, it can be shown that

$$
\begin{aligned}
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} y_{i,0} w' Q^{(\lambda)} u_i \quad &= \quad \frac{1}{\sqrt{N}} \sum_{i=1}^{N} y_{i,0} (e' + f'(\varphi_N - 1)) Q^{(\lambda)} u_i + o_P(1), \\
&= \quad \frac{1}{\sqrt{N}} \sum_{i=1}^{N} y_{i,0} e' Q^{(\lambda)} u_i + \frac{c}{N} \sum_{i=1}^{N} y_{i,0} f' Q^{(\lambda)} u_i + o_P(1), \\
&= \quad o_P(1), \qquad (11)
\end{aligned}
$$

since $e' Q^{(\lambda)} = 0$ and $E(y_{i,0} u_i) = 0$ by assumption (a4), and

$$
\begin{aligned}
& \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \gamma_i^{(\lambda)\prime} X^{(\lambda)\prime} \Omega' Q^{(\lambda)} u_i \\
= \quad & \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \gamma_i^{(\lambda)\prime} X^{(\lambda)\prime} (\Lambda' + F'(\varphi_N - 1) + o_p(1)) Q^{(\lambda)} u_i, \\
= \quad & \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \gamma_i^{(\lambda)\prime} X^{(\lambda)\prime} \Lambda' Q^{(\lambda)} u_i + \frac{c}{N} \sum_{i=1}^{N} \gamma_i^{(\lambda)\prime} X^{(\lambda)\prime} F' Q^{(\lambda)} u_i + o_p(1), \\
= \quad & \frac{c}{N} \sum_{i=1}^{N} (a_i^{(1)}, a_i^{(2)})' X^{(\lambda)\prime} \Lambda' Q^{(\lambda)} u_i + \frac{c^2}{N^{3/2}} \sum_{i=1}^{N} (a_i^{(1)}, a_i^{(2)})' X^{(\lambda)\prime} F' Q^{(\lambda)} u_i + o_p(1), \\
= \quad & o_p(1), \qquad (12)
\end{aligned}
$$

since $E(a_i^{(\lambda)} u_i) = 0$ by assumption (a4). Finally, we have

$$\begin{aligned} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} u_i' \Omega' Q^{(\lambda)} u_i &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} u_i' \left(\Lambda' + F'(\varphi_N - 1) + o_p(1) \right) Q^{(\lambda)} u_i, \\ &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} u_i' \Lambda' Q^{(\lambda)} u_i - \frac{c}{N} \sum_{i=1}^{N} u_i' F' Q^{(\lambda)} u_i + o_p(1), \end{aligned}$$

where

$$\frac{c}{N} \sum_{i=1}^{N} u_i' F' Q^{(\lambda)} u_i \xrightarrow{p} c\sigma^2 tr(F'Q^{(\lambda)}) \quad \text{and} \tag{13}$$

$$\sqrt{N} \left(\frac{1}{N} \sum_{i=1}^{N} u_i' \Lambda' Q^{(\lambda)} u_i - \sigma^2 tr(\Lambda' Q^{(\lambda)}) \right) \xrightarrow{d} N(0, V_{HT,A}), \tag{14}$$

where $V_{HT,A}$ is the variance of the last limiting distribution. Based on the asymptotic results given by equations (11)-(14), we can show that

$$(A) \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{N} y_{i,-1}' Q^{(\lambda)} u_i \xrightarrow{d} N\left(-c\sigma^2 tr(F'Q^{(\lambda)}), V_{HT,A}\right). \tag{15}$$

To derive asymptotic results for summand (B), write it as follows:

$$\begin{aligned} (B) &\equiv \frac{1}{\sqrt{N}} B(\lambda) \sum_{i=1}^{N} y_{i,-1}' Q^{(\lambda)} y_{i,-1} \\ &= \frac{1}{\sqrt{N}} B(\lambda) \sum_{i=1}^{N} \left(y_{i,0} w' + \gamma_i^{(\lambda)\prime} X^{(\lambda)\prime} \Omega' + u_i' \Omega' \right) Q^{(\lambda)} \left(y_{i,0} w + \Omega X^{(\lambda)} \gamma_i^{(\lambda)} + \Omega u_i \right). \end{aligned}$$

By similar arguments to those applied to derive results (11)-(14), we can prove the following asymptotic results:

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(y_{i,0}^2 w'Q^{(\lambda)}w + y_{i,0}w'Q^{(\lambda)}\Omega X^{(\lambda)}\gamma_i^{(\lambda)} + y_{i,0}w'Q^{(\lambda)}\Omega u_i\right) = o_p(1), \quad (16)$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\gamma_i^{(\lambda)\prime}X^{(\lambda)\prime}\Omega'Q^{(\lambda)}wy_{i,0}+\gamma_i^{(\lambda)\prime}X^{(\lambda)\prime}\Omega'Q^{(\lambda)}\Omega X^{(\lambda)}\gamma_i^{(\lambda)}+\Omega X^{(\lambda)}\gamma_i^{(\lambda)}\Omega u_i) = o_p(1), \quad (17)$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(u_i'\Omega'Q^{(\lambda)}wy_{i,0} + u_i'\Omega'Q^{(\lambda)}\Omega X^{(\lambda)}\gamma_i^{(\lambda)}\right) = o_p(1), \quad (18)$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}u_i'\Omega'Q^{(\lambda)}\Omega u_i = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}u_i'(\Lambda' + F'(\varphi_N - 1))Q^{(\lambda)}(\Lambda + F(\varphi_N - 1))u_i + o_p(1), \quad (19)$$

where

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}u_i'\Lambda'Q^{(\lambda)}\Lambda u_i - \sigma^2 tr(\Lambda'Q^{(\lambda)}\Lambda)\right) \xrightarrow{d} N(0, V_{HT,B}), \quad (20)$$

$$-\frac{c}{N}\sum_{i=1}^{N}u_i'F'Q^{(\lambda)}\Lambda u_i \xrightarrow{p} \sigma^2 tr(F'Q^{(\lambda)}\Lambda), \quad (21)$$

$$-\frac{c}{N}\sum_{i=1}^{N}u_i'\Lambda'Q^{(\lambda)}Fu_i \xrightarrow{p} \sigma^2 tr(\Lambda'Q^{(\lambda)}F) \quad \text{and} \quad (22)$$

$$\frac{c^2}{N^{3/2}}\sum_{i=1}^{N}u_i'F'Q^{(\lambda)}Fu_i = o_p(1). \quad (23)$$

Based on the above results, given by equations (16)-(23), it can be shown that

$$(B) \equiv \frac{1}{\sqrt{N}}B(\lambda)\sum_{i=1}^{N}y_{i,-1}'Q^{(\lambda)}y_{i,-1} \xrightarrow{d} \quad (24)$$
$$N\left(-c\sigma^2 B(\lambda)[tr(F'Q^{(\lambda)}\Lambda) + tr(\Lambda'Q^{(\lambda)}F)], B^2(\lambda)V_{HT,B}\right).$$

Finally, following similar arguments to the above, we can easily show that, for quantity (C), the following asymptotic result holds:

$$(C) \equiv \frac{1}{N} \sum_{i=1}^{N} y'_{i,-1} Q^{(\lambda)} y_{i,-1} \xrightarrow{p} \sigma^2 tr(\Lambda' Q^{(\lambda)} \Lambda). \tag{25}$$

Using asymptotic results (15), (24) and (25), equation (10) implies that

$$\sqrt{N}(\hat{\varphi}^{(\lambda)} - \varphi_N - B(\lambda)) \xrightarrow{d} N\left(-c\frac{tr(F'Q^{(\lambda)}) - 2B(\lambda)tr(F'Q^{(\lambda)}\Lambda)}{tr(\Lambda'Q^{(\lambda)}\Lambda)}, V_{HT}^{(\lambda)}\right), \tag{26}$$

$$\text{or} \sqrt{N}(\hat{\varphi}^{(\lambda)} - 1 - B(\lambda)) \xrightarrow{d} N\left(-c\frac{tr(F'Q^{(\lambda)}) + tr(\Lambda'Q^{(\lambda)}\Lambda) - 2B(\lambda)tr(F'Q^{(\lambda)}\Lambda)}{tr(\Lambda'Q^{(\lambda)}\Lambda)}, V_{HT}^{(\lambda)}\right),$$

since $tr(\Lambda'Q^{(\lambda)}) - B(\lambda)tr(\Lambda'Q^{(\lambda)}\Lambda) = 0$. Note that the analytic formula of variance $V_{HT}^{(\lambda)}$ of the last limiting distribution is the same with that of the HT test under null hypothesis H_0: $c = 0$, given by $V_{HT}^{(\lambda)} = 2tr(A_{HT}^{(\lambda)2})/tr(\Lambda'Q^{(\lambda)}\Lambda)^2$. This does not depend on local parameter c. It remains the same under the null and sequence of local alternative hypotheses (see, e.g., Madsen (2010) and Karavias and Tzavalis (2014c)), given as $V_{HT}^{(\lambda)} = 2tr(A_{HT}^{(\lambda)2})/tr(\Lambda'Q^{(\lambda)}\Lambda)^2$. Scaling by $V_{HT}^{(\lambda)-1/2}$ the above limiting distribution yields

$$V_{HT}^{(\lambda)-1/2}\sqrt{N}(\hat{\varphi}^{(\lambda)} - 1 - B(\lambda)) \xrightarrow{d} N\left(-ck_{HT}, 1\right), \text{ with} \tag{27}$$

$$k_{HT} = \frac{tr(F'Q^{(\lambda)}) + tr(\Lambda'Q^{(\lambda)}\Lambda) - 2B(\lambda)tr(F'Q^{(\lambda)}\Lambda)}{\sqrt{2tr(A_{HT}^{(\lambda)^2})}}.$$

Substituting into the above formula of k_{HT} the following identities:

$$\begin{aligned}
tr(F'Q^{(\lambda)}\Lambda) &= tr(\Lambda'Q^{(\lambda)}F) = \qquad (28)\\
&= \frac{6}{144}(3\lambda^2-3\lambda+1)T^3 - \frac{1}{12}(2\lambda^2-2\lambda+1)T^2 - \frac{1}{24}T + \frac{1}{6}\\
tr(\Lambda'Q^{(\lambda)}\Lambda) + tr(F'Q^{(\lambda)}) + tr(\Lambda'Q^{(\lambda)}) &= 0, \qquad (29)\\
tr(F'Q^{(\lambda)}) &= -\frac{T^2}{6}(2\lambda^2-2\lambda+1) + \frac{T}{2} - \frac{4}{6}, \qquad (30)\\
tr(\Lambda'Q^{(\lambda)}) &= -\frac{T-2}{2}, \qquad (31)\\
tr(\Lambda'Q^{(\lambda)}\Lambda) &= \frac{T^2}{6}(2\lambda^2-2\lambda+1) - \frac{2}{6}, \qquad (32)
\end{aligned}$$

$$\begin{aligned}
tr(A_{HT}^{(\lambda)^2}) &= tr\left[\left(\frac{1}{2}(\Lambda'Q^{(\lambda)} + Q^{(\lambda)}\Lambda) - B(\lambda)(\Lambda'Q^{(\lambda)}\Lambda)\right)^2\right], \qquad (33)\\
tr\left(\left(\Lambda'Q^{(\lambda)} + Q^{(\lambda)}\Lambda\right)^2\right) &= \frac{T^2}{6}(2\lambda^2-2\lambda+1) + T - \frac{7}{3}, \qquad (34)\\
tr\left(\left(\Lambda'Q^{(\lambda)}\Lambda\right)^2\right) &= \frac{1}{90}(2\lambda^4-4\lambda^3+6\lambda^2-4\lambda+1)T^4 \qquad (35)\\
&\quad + \frac{1}{36}(2\lambda^2-2\lambda+1)T^2 - \frac{7}{90},\\
tr\left(\left(\Lambda'Q^{(\lambda)} + Q^{(\lambda)}\Lambda\right)\left(\Lambda'Q^{(\lambda)}\Lambda\right)\right) &= \frac{T-2}{2}, \qquad (36)
\end{aligned}$$

yields the results of Theorem 1, for the HT test statistic. Note that $2tr(A_{HT}^{(\lambda)^2})$ can be analytically written as

$$2tr(A_{HT}^{(\lambda)^2}) = \frac{D}{S}, \text{ where}$$

$$
\begin{aligned}
D &= T^6R_1 + T^5R_2 + T^4R_3 + T^2R_4 + 216T - 136,\\
S &= T^4\Phi_1 + T^2\Phi_2 + 240,\\
R_1 &= 40\lambda^6 - 120\lambda^5 + 204\lambda^4 - 208\lambda^3 + 162\lambda^2 - 78\lambda + 17,\\
R_2 &= -216\lambda^4 + 432\lambda^3 - 528\lambda^2 + 312\lambda - 78,\\
R_3 &= 216\lambda^4 - 432\lambda^3 + 588\lambda^2 - 372\lambda + 108,\\
R_4 &= -120\lambda^2 + 120\lambda - 144,\\
\Phi_1 &= 240\lambda^4 - 480\lambda^3 + 480\lambda^2 - 240\lambda + 60 \text{ and}\\
\Phi_2 &= -480\lambda^2 + 480\lambda - 240.
\end{aligned}
$$

To derive the limiting distribution of the KT test under the sequence of local alternatives $\varphi_N = 1 - c/\sqrt{N}$, write

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

$$
= \frac{1}{\sqrt{N}}\sum_{i=1}^{N} y'_{i,-1}Q^{(\lambda)}u_i - \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \Delta y'_i \Psi^{(\lambda)}\Delta y_i, \qquad (37)
$$

where Δy_i can be written as

$$
\Delta y_i = u_i + (\varphi_N - 1)y_{i,-1} + X^{(\lambda)}\gamma_i^{(\lambda)}. \qquad (38)
$$

The limiting distribution of the KT test under $\varphi_N = 1 - c/\sqrt{N}$ can be proved by obtaining asymptotic results for the two summands entering into equation (37), i.e., $\left(1/\sqrt{N}\right)\sum_{i=1}^{N} y_{i,-1}' Q^{(\lambda)} u_i$ and $\left(1/\sqrt{N}\right)\sum_{i=1}^{N} \Delta y_i' \Psi^{(\lambda)} \Delta y_i$, following analogous to the proof of (27) steps. The formula of slope power parameter k_{KT} is given as

$$k_{KT} = \frac{tr(F'Q^{(\lambda)}) + tr(\Lambda' Q^{(\lambda)} \Lambda)}{\sqrt{2tr(A_{KT}^{(\lambda)^2})}}. \tag{39}$$

Substituting the following identities into the above formula of k_{KT}:

$$tr(A_{KT}^{(\lambda)^2}) = tr\left(\left(\frac{1}{2}(\Lambda' Q^{(\lambda)} + Q^{(\lambda)}\Lambda - \Psi^{(\lambda)} - \Psi^{(\lambda)\prime})\right)^2\right), \tag{40}$$

$$\begin{aligned}
tr(\Psi^{(\lambda)}\Lambda) &= tr(\Lambda'\Psi^{(\lambda)}) = 0, && (41)\\
2tr(A_{KT}^{(\lambda)^2}) &= 2tr(P^{(\lambda)}) - 2tr(Z^{(\lambda)2}),\ \text{with } Z^{(\lambda)} = \frac{1}{2}(\Psi^{(\lambda)\prime} + \Psi^{(\lambda)}) && (42)\\
\text{and }\ P^{(\lambda)} &= \frac{1}{2}(\Lambda' Q^{(\lambda)})^2 + \frac{1}{2}\Lambda' Q^{(\lambda)}\Lambda, && (43)\\
tr\left((\Lambda' Q^{(\lambda)})^2\right) &= -\frac{T^2}{12}(2\lambda^2 - 2\lambda - 1) + \frac{T}{2} - \frac{5}{6}\ \text{and} && (44)\\
tr(Z^{(\lambda)2}) &= \frac{-\frac{1}{T} + 2(\lambda-1)\lambda T}{6(\lambda-1)\lambda} - 1 && (45)
\end{aligned}$$

yields the results of Theorem 1, for the KT test statistic.

Proof of Corollary 1: The results of the corollary and, in particular, those of equation (1) can be derived based on analogous arguments to those applied for the proof of Theorem 1.

To obtain the analytic formula of k_{HT}^*, given by equation (1), scale (9) by T, replace φ_N

with $\varphi_{N,T}$, and apply asymptotic theory for $N \to \infty$, as in Theorem 1. Then, we will have

$$T\sqrt{N}(\hat{\varphi}^{(\lambda)} - \varphi_{N,T} - B(\lambda)) \xrightarrow{d} N\left(-c\frac{tr(F'Q^{(\lambda)}) - 2B(\lambda)tr(F'Q^{(\lambda)}\Lambda)}{tr(\Lambda'Q^{(\lambda)}\Lambda)}, T^2V_{HT}^{(\lambda)}\right).$$

Multiplying with $\left(T^2V_{HT}^{(\lambda)}\right)^{-1/2}$ and using $\varphi_{N,T} = 1 - c/\left(T\sqrt{N}\right)$, the last limiting distribution can be written as

$$T\left(T^2V_{HT}^{(\lambda)}\right)^{-1/2}\sqrt{N}(\hat{\varphi}^{(\lambda)} - 1 - B(\lambda)) \xrightarrow{d} N\left(-c\frac{1}{T}k_{HT}, 1\right) \tag{46}$$

where $k_{HT} = \left[tr(F'Q^{(\lambda)}) + tr(\Lambda'Q^{(\lambda)}\Lambda) - 2B(\lambda)tr(F'Q^{(\lambda)}\Lambda)\right]/\sqrt{2tr(A_{HT}^{(\lambda)2})}$ (see proof of Theorem 1). By taking the limit for $T \to \infty$ of k_{HT} and $T^2V_{HT}^{(\lambda)}$, (46) can be written as

$$T\left(V_{HT}^{*(\lambda)}\right)^{-1/2}\sqrt{N}(\hat{\varphi}^{(\lambda)} - 1 - B(\lambda)) \xrightarrow{d} N\left(-ck_{HT}^{*}, 1\right), \text{ where}$$

$$\begin{aligned} k_{HT}^{*} &\equiv \lim_T \frac{1}{T}k_{HT} = \frac{3\lambda^2 - 3\lambda + 1}{4(2\lambda^2 - 2\lambda + 1)}\sqrt{\frac{\Phi_1}{R_1}} \text{ and} \\ V_{HT}^{*(\lambda)} &\equiv \lim_T T^2V_{HT}^{(\lambda)} = \frac{36R_1}{\Phi_1(2\lambda^2 - 2\lambda - 1)^2}. \end{aligned}$$

The analytic formulas of the last two limits are derived based on the results of identities (28)-(36). The above results have been derived by taking limits sequentially, first for $N \to \infty$ and then for $T \to \infty$. Joint convergence in N, T requires the extra assumption that $\sqrt{N}/T \to 0$, see also Moon and Perron (2008). However, for $c = 0$ there is no need to specify the relative rate of convergence between N and T (see Hahn and Kuersteiner (2002) and Karavias and Tzavalis (2014a)).

Considering now the KT test, by the definition of $V_{KT}^{(\lambda)}$, we have that $\lim_T V_{KT}^{(\lambda)} = +\infty$.

To clarify how this is not a problem for the implementation of the statistic, notice that the definition of the variance $V_{KT}^{(\lambda)} = 2tr(A_{KT}^{(\lambda)^2})$ is made for convenience (our notation gives weight to $\hat{\delta}^{(\lambda)}$ and thus to the bias correction aspect of the test) and does not correspond exactly to the variance of $\sqrt{N}(\hat{\varphi}^{(\lambda)} - 1 - \hat{b}^{(\lambda)}/\hat{\delta}^{(\lambda)})$. Noticing that $p\lim_N \hat{\delta}^{(\lambda)} = tr(\Lambda' Q^{(\lambda)}\Lambda)$ and that in the HT test, $V_{HT}^{(\lambda)} = 2tr(A_{HT}^{(\lambda)^2})/tr(\Lambda' Q^{(\lambda)}\Lambda)^2$, the KT test maybe written as $\sqrt{\hat{\delta}^{(\lambda)2}/\left(2tr(A_{KT}^{(\lambda)^2})\right)}\sqrt{N}(\hat{\varphi}^{(\lambda)} - 1 - \hat{b}^{(\lambda)}/\hat{\delta}^{(\lambda)})$.

The formulas of k_{KT}^* and $V_{KT}^{*(\lambda)}$, given by the corollary for the large-T version of the KT test, can be derived by following similar steps to the above. Then, using the results of identities (40)-(45), we can obtain

$$k_{KT}^* \equiv \lim_T \frac{1}{T} k_{KT} = 0 \text{ and } V_{KT}^{*(\lambda)} \equiv \lim_T T^2 V_{KT}^{(\lambda)}/\left(p\lim_N \hat{\delta}^{(\lambda)}\right) = \frac{36(2\lambda^4 - 4\lambda^3 + 3\lambda^2 - \lambda)}{12(\lambda-1)\lambda(2\lambda^2 - 2\lambda + 1)^2}.$$

Proof of Theorem 2: To prove the theorem, we will follow analogous steps to those for the proof of Theorem 1. We now will rely on relationships (8) and (38), where now vector $\gamma_i^{(\lambda)}$ is defined as

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

due to the presence of individual trends under $\varphi_{N,T} = 1 - c/\sqrt{N}$, where $\mu_i = (\alpha_i^{(1)} - \beta_i^{(1)}, \alpha_i^{(2)} - \beta_i^{(2)}, \beta_i^{(1)}, \beta_i^{(2)})'$, $e_* = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$ and $\nu_i = (\beta_i^{(1)}, \beta_i^{(2)})'$. The non-standardized HT test statistic for model $M2$ can be written as follows:

$$\begin{aligned}
&\sqrt{N}(\hat{\varphi}^{(\lambda)} - \varphi_N - B(\lambda)) \\
= \; & \sqrt{N}\left(\frac{\frac{1}{N}\sum_{i=1}^{N} y_{i,-1}' Q^{(\lambda)}(\varphi_N y_{i,-1} + X^{(\lambda)}\gamma_i^{*(\lambda)} + u_i)}{\frac{1}{N}\sum_{i=1}^{N} y_{i,-1}' Q^{(\lambda)} y_{i,-1}} - \varphi_N - B(\lambda) \right) \\
= \; & \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i,-1}' Q^{(\lambda)} u_i - \frac{1}{\sqrt{N}} B(\lambda)\sum_{i=1}^{N} y_{i,-1}' Q^{(\lambda)} y_{i,-1}}{\frac{1}{N}\sum_{i=1}^{N} y_{i,-1}' Q^{(\lambda)} y_{i,-1}} = \frac{(A') - (B')}{(C')},
\end{aligned}$$

where $(A') \equiv \left(1/\sqrt{N}\right) \sum_{i=1}^{N} y_{i,-1}' Q^{(\lambda)} u_i$, $(B') \equiv B(\lambda) \left(1/\sqrt{N}\right) \sum_{i=1}^{N} y_{i,-1}' Q^{(\lambda)} y_{i,-1}$ and $(C') \equiv (1/N) \sum_{i=1}^{N} y_{i,-1}' Q^{(\lambda)} y_{i,-1}$. As in the proof of Theorem 1, next we derive asymptotic results of (A'), (B') and (C'), using $\gamma_i^{*(\lambda)} = e_* \nu_i + (1 - \varphi_N)\mu_i$. The most important ones are the following:

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} \nu_i' e_*' X^{(\lambda)\prime} \Lambda' Q^{(\lambda)} \Lambda X^{(\lambda)} e_* \nu_i - tr(e_*' X^{(\lambda)\prime} \Lambda' Q^{(\lambda)} \Lambda X^{(\lambda)} e_* E(\nu_i \nu_i')) \right) \xrightarrow{d} N(0, V_{HT,4})$$

$$\frac{c}{N}\sum_{i=1}^{N} \nu_i' e_*' X^{(\lambda)\prime} F' Q^{(\lambda)} \Lambda e_* \nu_i \xrightarrow{p} ctr(e_*' X^{(\lambda)\prime} F' Q^{(\lambda)} \Lambda e_* E(\nu_i \nu_i'))$$

$$\frac{c}{N}\sum_{i=1}^{N} \nu_i' e_*' X^{(\lambda)\prime} \Lambda' Q^{(\lambda)} F X^{(\lambda)} e_* \nu_i \xrightarrow{p} ctr(e_*' X^{(\lambda)\prime} \Lambda' Q^{(\lambda)} F e_* E(\nu_i \nu_i'))$$

$$\frac{c}{N}\sum_{i=1}^{N} \mu_i' X^{(\lambda)\prime} \Omega' Q^{(\lambda)} \Omega X^{(\lambda)} e_* \nu_i \xrightarrow{p} ctr(X^{(\lambda)\prime} \Lambda' Q^{(\lambda)} \Lambda X^{(\lambda)} e_* E(\nu_i \mu_i'))$$

$$\frac{c}{N}\sum_{i=1}^{N} \nu_i' e_*' X^{(\lambda)\prime} \Lambda' Q^{(\lambda)} \Lambda X^{(\lambda)} \mu_i \xrightarrow{p} ctr(e_*' X^{(\lambda)\prime} \Lambda' Q^{(\lambda)} \Lambda X^{(\lambda)} E(\mu_i \nu_i'))$$

Given these results, the proof of Theorem 2 for the test statistic HT follows immediately, after using the following identities:

$$
\begin{aligned}
tr(e_*'X^{(\lambda)\prime}\Lambda' Q^{(\lambda)}\Lambda X^{(\lambda)}e_*E(\nu_i\nu_i')) &= 0 \\
tr(e_*'X^{(\lambda)\prime}F'Q^{(\lambda)}\Lambda e_*E(\nu_i\nu_i')) - tr(e_*'X^{(\lambda)\prime}\Lambda' Q^{(\lambda)}Fe_*E(\nu_i\nu_i')) &= 0 \\
\text{and } tr(X^{(\lambda)\prime}\Lambda' Q^{(\lambda)}\Lambda X^{(\lambda)}e_*E(\nu_i\mu_i')) - tr(e_*'X^{(\lambda)\prime}\Lambda' Q^{(\lambda)}\Lambda X^{(\lambda)}E(\mu_i\nu_i')) &= 0.
\end{aligned}
$$

The proof of the second result of the theorem, i.e., $k_{KT} = 0$, can be proved by following analogous steps to the above and using the following identities:

$$
\begin{aligned}
tr(e_*'X^{(\lambda)\prime}\Theta^{(\lambda)}\Lambda X^{(\lambda)}e_*E(\nu_i\nu_i')) - tr(e_*'X^{(\lambda)\prime}\Lambda'\Theta^{(\lambda)}X^{(\lambda)}e_*E(\nu_i\nu_i')) &= 0 \\
\text{and} \quad tr(X^{(\lambda)\prime}\Theta^{(\lambda)}X^{(\lambda)}e_*E(\nu_i\mu_i')) - tr(e_*'X^{(\lambda)\prime}\Theta^{(\lambda)}X^{(\lambda)}E(\mu_i\nu_i')) &= 0.
\end{aligned}
$$

Proof of Corollary 2: The proof comes directly from Theorem 2 by scaling the results with T.

Proof of Theorem 3: This can be proved by following analogous steps to the proof of Theorem 1, for the KT test statistic, by inserting $\bar{\Gamma}$ instead of $\sigma^2 I$ and by using the corresponding asymptotic theorems (see Karavias and Tzavalis (2014b) or proof of Theorem 7 for an example).

Proof of Theorem 4: This can be proved by following analogous steps to the proof of Theorems 2 and 3, for the KT test statistic.

Proof of Theorem 5: Under the null hypothesis, model $\tilde{M}1$ becomes:

$$
y_i = y_{i,-1} + \xi_i^{(1)} f^{(1)} + \xi_i^{(2)} f^{(2)} + u_i, \tag{47}
$$

for $i = 1, ..., N$. Solving backwards last rlationship yields

$$y_{i,-1} = y_{i,0}e + \xi_i^{(1)}\Lambda f^{(1)} + \xi_i^{(2)}\Lambda f^{(2)} + \Lambda u_i. \tag{48}$$

The following proof is as in Karavias and Tzavalis (2014a). Equation (48) corresponds to equation (8), under the null hypothesis and the presence of common factors. Notice that multiplying (47) and (48) with $Q^{(\lambda)}$ (based on the augmented $X^{(\lambda)} = \left(e^{(1)}, e^{(2)}, f^{(1)}, f^{(2)}, \Lambda f^{(1)}, \Lambda f^{(2)}\right)$) removes the nuisance parameters such that:

$$\begin{aligned} Q^{(\lambda)}y_i &= Q^{(\lambda)}y_{i,-1} + Q^{(\lambda)}u_i \\ Q^{(\lambda)}y_{i,-1} &= Q^{(\lambda)}\Lambda u_i. \end{aligned} \tag{49}$$

Substituting (49) in the inconsistency of $\hat{\varphi}^{(\lambda)}$:

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

By applying standard properties of the quadratic forms:

$$\begin{aligned} E(u_i'\Lambda' Q^{(\lambda)}u_i) &= \sigma^2 tr(\Lambda' Q^{(\lambda)}), \\ E(u_i'\Lambda' Q^{(\lambda)}\Lambda u_i) &= \sigma^2 tr(\Lambda' Q^{(\lambda)}\Lambda) \end{aligned}$$

and thus

$$B(\lambda) = p\lim_{N}(\hat{\varphi}^{(\lambda)} - 1) = \frac{tr(\Lambda' Q^{(\lambda)})}{tr(\Lambda' Q^{(\lambda)}\Lambda)}.$$

To derive the limiting distribution under the null

$$\sqrt{N}(\hat{\varphi}^{(\lambda)} - 1 - B(\lambda)) = \frac{\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[u_i' \Lambda' Q^{(\lambda)} u_i - B(\lambda) u_i' \Lambda' Q^{(\lambda)} u_i \right]}{\frac{1}{N} \sum_{i=1}^{N} u_i' \Lambda' Q^{(\lambda)} u_i}.$$

Then, $p\lim_N (1/N) \sum_{i=1}^{N} u_i' \Lambda' Q^{(\lambda)} u_i = \sigma^2 tr(\Lambda' Q^{(\lambda)} \Lambda)$. Also, $N^{-1/2} \sum_{i=1}^{N} \left[u_i' \Lambda' Q^{(\lambda)} u_i - B(\lambda) u_i' \Lambda' Q^{(\lambda)} u_i \right] = (1/\sqrt{N}) \sum_{i=1}^{N} u_i' A_{HT}^{(\lambda)} u_i$ where $E(u_i' A_{HT}^{(\lambda)} u_i) = 0$ and $Var(u_i' A_{HT}^{(\lambda)} u_i) = 2\sigma^4 tr(A_{HT}^{(\lambda)2})$. The result follows from the Lindeberg-Levy CLT and the CMT. The proof of the distribution under the local alternatives is the same with the proof of Theorem 1.

Proof of Theorem 6: Under the null hypothesis, model $\tilde{M}2$ becomes:

$$y_i = y_{i,-1} + \beta_i^{(1)} e^{(1)} + \beta_i^{(2)} e^{(2)} + \xi_i^{(1)} f^{(1)} + \xi_i^{(2)} f^{(2)} + u_i, \tag{50}$$

for $i = 1, ..., N$. Solving backwards last relationship gives

$$y_{i,-1} = y_{i,0} e + \beta_i^{(1)} \Lambda e^{(1)} + \beta_i^{(2)} \Lambda e^{(2)} + \xi_i^{(1)} \Lambda f^{(1)} + \xi_i^{(2)} \Lambda f^{(2)} + \Lambda u_i. \tag{51}$$

Then, by multiplying with $Q^{(\lambda)}$ (based on $X^{(\lambda)} = \left(e^{(1)}, e^{(2)}, \tau^{(1)}, \tau^{(2)}, f^{(1)}, f^{(2)}, \Lambda f^{(1)}, \Lambda f^{(2)} \right)$) we remove the nuisance parameters such that:

$$\begin{aligned} Q^{(\lambda)} y_i &= Q^{(\lambda)} y_{i,-1} + Q^{(\lambda)} u_i \\ Q^{(\lambda)} y_{i,-1} &= Q^{(\lambda)} \Lambda u_i. \end{aligned} \tag{52}$$

The proof then follows the steps of Theorem 5.

Proof of Theorem 7: We first begin by proving some claims in the text before Theorem 7. Under null hypothesis H_0:$c = 0$, we have $\Delta y_i = \xi_i^{(1)} f^{(1)} + \xi_i^{(2)} f^{(2)} + u_i$. Then, $\hat{\Gamma}$ can be written as

$$\begin{aligned}
\hat{\Gamma} &= \frac{1}{N}\sum_{i=1}^{N} \Delta y_i \Delta y_i' \\
&= \frac{1}{N}\sum_{i=1}^{N} \left(\xi_i^{(1)} f^{(1)} + \xi_i^{(2)} f^{(2)} + u_i\right)\left(\xi_i^{(1)} f^{(1)} + \xi_i^{(2)} f^{(2)} + u_i\right)' \\
&= \frac{1}{N}\sum_{i=1}^{N} (\xi_i^{(1)2} f^{(1)} f^{(1)\prime} + \xi_i^{(2)2} f^{(2)} f^{(2)\prime} + u_i u_i' + \xi_i^{(1)}\xi_i^{(2)} f^{(1)} f^{(2)\prime} + \xi_i^{(1)} f^{(1)} u_i' \\
&\quad + \xi_i^{(2)}\xi_i^{(1)} f^{(2)} f^{(1)\prime} + \xi_i^{(2)} f^{(2)} u_i' + \xi_i^{(1)} u_i f^{(1)\prime} + \xi_i^{(2)} u_i f^{(2)\prime})
\end{aligned}$$

To make the exposition simpler and without loss of generality, write the top left element of $\hat{\Gamma}$ as

$$\frac{1}{N}\sum_{i=1}^{N} \left(\xi_i^{(1)2} f_1^{(1)2} + u_{i,1}^2 + 2\xi_i^{(1)} f_1^{(1)} u_{i,1}\right),$$

where $E\left(\xi_i^{(1)2} f_1^{(1)2} + u_{i,1}^2 + 2\xi_i^{(1)} f_1^{(1)} u_{i,1}\right) = E\left(\xi_i^{(1)2}\right) f_1^{(1)2} + E(u_{i,1}^2)$ by condition (e1). Also, by Condition (e2) it is straightforward that $Var(\xi_i^{(1)2} f_1^{(1)2} + u_{i,1}^2 + 2\xi_i^{(1)} f_1^{(1)} u_{i,1})$ is finite. Then, by Chebyshev's Weak Law of Large Numbers, we obtain the following result:

$$p\lim_{N} \frac{1}{N}\sum_{i=1}^{N} \left[\xi_i^{(1)2} f_1^{(1)2} + u_{i,1}^2 + 2\xi_i^{(1)} f_1^{(1)} u_{i,1} - E\left(\xi_i^{(1)2}\right) f_1^{(1)2} - E(u_{i,1}^2)\right] = 0.$$

The arguments used here apply to all elements of $\hat{\Gamma}$ and thus we have that

$$p\lim_{N}[\hat{\Gamma} - \bar{\Gamma} - \bar{\xi}^{(1)2} f^{(1)} f^{(1)\prime} + \bar{\xi}^{(2)2} f^{(2)} f^{(2)\prime})] = 0.$$

Next, we will show that $\bar{\xi}^{(1)2}$ and $\bar{\xi}^{(2)2}$ can be consistently estimated. First, write $tr(M_f^{(1)}\hat{\Gamma}) =$

$(1/N)\sum_{i=1}^{N}\Delta y_i' M_f^{(1)}\Delta y_i$. Then, we have

$$\begin{aligned} E(\Delta y_i' M_f^{(1)}\Delta y_i) &= tr\left[M_f^{(1)}E(\Delta y_i \Delta y_i')\right] \text{ and} \\ E(\Delta y_i \Delta y_i') &= E(\xi_i^{(1)2})f^{(1)}f^{(1)\prime} + E(\xi_i^{(2)2})f^{(2)}f^{(2)\prime} + \Gamma_i \end{aligned}$$

by Assumption E. The above imply that $tr\left[M_f^{(1)}E(\Delta y_i\Delta y_i')\right] = E(\xi_i^{(1)2})tr\left[M_f^{(1)}f^{(1)}f^{(1)\prime}\right] + E(\xi_i^{(2)2})tr\left[M_f^{(1)}f^{(2)}f^{(2)\prime}\right] + tr\left[M_f^{(1)}\Gamma_i\right]$. But since $M_f^{(j)}$ has zeroes in its central diagonals wherever Γ_i is non-zero, we have $tr\left[M_f^{(j)}\Gamma_i\right] = 0$ and also $tr\left[M_f^{(1)}f^{(2)}f^{(2)\prime}\right] = 0$, because the block forms that the matrices have due to the structural break. Thus, for all i, it holds that $tr\left[M_f^{(1)}E(\Delta y_i\Delta y_i')\right] = E(\xi_i^{(1)2})tr\left(M_f^{(1)}f^{(1)}f^{(1)\prime}\right)$. Since variances of $\Delta y_i\Delta y_i'$ are finite, for all i, by Assumptions A and E, the following results hold

$$\begin{aligned} p\lim_N\left[tr(M_f^{(1)}\hat{\Gamma}) - \bar{\xi}^{(1)2}tr\left(M_f^{(1)}f^{(1)}f^{(1)\prime}\right)\right] &= 0, \text{ or} \\ p\lim_N\left[\frac{tr(M_f^{(1)}\hat{\Gamma})}{tr\left(M_f^{(1)}f^{(1)}f^{(1)\prime}\right)} - \bar{\xi}^{(1)2}\right] &= 0, \end{aligned}$$

by Chebyshev's Weak Law of Large. A similar result to the above hods for $\bar{\xi}^{(j)2}$, if $j = 2$. Finally, we will show that $p\lim_N\left[tr(\Theta^{(\lambda)}\hat{\Gamma}) - tr(\Lambda' Q^{(\lambda)}\bar{\Gamma})\right] = 0$. To this end, first note that $tr(\Lambda' Q^{(\lambda)}\bar{\Gamma}) = (1/N)\sum_{i=1}^{N} tr(\Lambda' Q^{(\lambda)}\Gamma_i)$. Also, note that $tr(\Theta^{(\lambda)}\hat{\Gamma}) = (1/N)\sum_{i=1}^{N}\Delta y_i'\Theta^{(\lambda)}\Delta y_i$ with $E(\Delta y_i'\Theta^{(\lambda)}\Delta y_i) = tr(\Theta^{(\lambda)}E(\Delta y_i\Delta y_i')) = E(\xi_i^{(1)2})tr\left(\Theta^{(\lambda)}f^{(1)}f^{(1)\prime}\right) + E(\xi_i^{(2)2})tr\left(\Theta^{(\lambda)}f^{(2)}f^{(2)\prime}\right) + tr\left(\Theta^{(\lambda)}\Gamma_i\right)$. Next note the following result:

$$\begin{aligned} tr\left(\Theta^{(\lambda)}f^{(1)}f^{(1)\prime}\right) &= tr\left(\Psi^{(\lambda)}f^{(1)}f^{(1)\prime}\right) - \frac{tr(\Psi^{(\lambda)}f^{(1)}f^{(1)\prime})}{tr(M_f^{(1)}f^{(1)}f^{(1)\prime})}tr\left(M_f^{(1)}f^{(1)}f^{(1)\prime}\right) \\ &\quad - \frac{tr(\Psi^{(\lambda)}f^{(2)}f^{(2)\prime})}{tr(M_f^{(2)}f^{(2)}f^{(2)\prime})}tr\left(M_f^{(2)}f^{(1)}f^{(1)\prime}\right) \\ &= tr\left(\Psi^{(\lambda)}f^{(1)}f^{(1)\prime}\right) - tr(\Psi^{(\lambda)}f^{(1)}f^{(1)\prime}) = 0, \end{aligned}$$

which holds because $tr\left(M_f^{(2)} f^{(1)} f^{(1)\prime}\right) = 0$. Similarly, we can prove that $tr\left(\Theta^{(\lambda)} f^{(2)} f^{(2)\prime}\right) = 0$ and, hence, $tr(\Theta^{(\lambda)} E(\Delta y_i \Delta y_i')) = tr\left(\Theta^{(\lambda)}\Gamma_i\right)$. Also, note that $tr\left(\Theta^{(\lambda)}\Gamma_i\right) = tr\left(\Psi^{(\lambda)}\Gamma_i\right)$ because $tr\left(M_f^{(j)}\Gamma_i\right) = 0$, by the construction of $M_f^{(j)}$, for $j = 1, 2$. By the construction of $\Psi^{(\lambda)}$, we have $tr\left(\Psi^{(\lambda)}\Gamma_i\right) = tr(\Lambda' Q^{(\lambda)}\Gamma_i)$. Taking together the above results, we have that

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

The rest of the proof follows similar steps to those of the proof of Theorem 1. Following analogous steps to the above, we can prove the results of Theorem 7 for model $\tilde{M}2$.

8 Tables

Table 1: Values of k_{HT} and k_{KT} for model $M1$

	k_{HT}				k_{KT}			
$\lambda \backslash T$	8	10	15	20	8	10	15	20
0.25	3.18	4.12	6.11	7.75	1.85	1.86	1.96	2.10
0.50	2.93	3.62	5.32	6.99	2.12	2.23	2.34	2.39
0.75	3.18	3.81	5.78	7.75	1.85	2.04	2.09	2.10

Table 2: Values of slope parameters k^*_{HT} and k^*_{KT} for model $M1$

λ	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
k^*_{HT}	0.433	0.394	0.360	0.338	0.332	0.338	0.360	0.394	0.433
k^*_{KT}	0	0	0	0	0	0	0	0	0

Table 3: Values of k_{KT} with $u_{i,t} = \varepsilon_{i,t} + \theta\varepsilon_{i,t-1}$						
Panel A: Model $M1$						
T	T_0	$\theta = -0.8$	$\theta = -0.5$	$\theta = 0.0$	$\theta = 0.5$	$\theta = 0.8$
8	2	−1.40	−0.63	1.58	2.71	2.89
	4	0.25	0.61	1.89	2.86	3.04
	6	1.28	1.36	1.58	1.66	1.68
10	2	−1.62	−0.69	1.65	2.60	2.73
	5	0.07	0.56	2.12	3.05	3.21
	7	0.82	1.10	1.82	2.12	2.16
15	3	−1.55	−0.48	1.81	2.50	2.58
	7	−0.41	0.36	2.31	3.10	3.21
	11	0.52	1.01	1.95	2.20	2.23
20	5	−1.52	−0.31	2.00	2.61	2.68
	10	−0.54	0.38	2.38	3.02	3.10
	15	0.33	1.07	2.00	2.22	2.24
Panel B: Model $M2$						
8	4	0.08	0.070	0	−0.09	−0.11
10	5	0.20	0.15	0	−0.12	−0.14
	7	0.66	0.46	0	−0.21	−0.24
15	3	0	0	0	0	0
	7	0.47	0.32	0	−0.13	−0.15
	11	0.75	0.53	0	−0.20	−0.23
20	5	0.17	0.11	0	−0.03	−0.04
	10	0.70	0.45	0	−0.15	−0.17
	15	0.80	0.54	0	−0.17	−0.20

Table 4: Values of $\bar{k}_{HT}$ and $\bar{k}_{KT}$

Panel A: Model $\tilde{M}1$						
$\lambda \backslash T$	k_{HT}			k_{KT}		
	12	15	20	12	15	20
0.25	3.50	4.84	6.45	1.30	1.56	1.81
0.50	3.06	4.09	5.70	1.32	1.58	2.01
0.75	3.55	4.67	6.60	1.31	1.56	1.84
Panel B: Model $\tilde{M}2$						
0.35	0	0	0	1.15	1.59	2.23
0.50	0	0	0	1.00	1.51	2.20
0.65	0	0	0	1.05	1.55	2.22

Table 5: Simulated size and power of the HT and KT tests for $u_{i,t} \sim NIID(0, \sigma^2)$											
				Model $M1$				Model $M2$			
N				100	300	1000	TV	100	300	1000	TV
$T = 8$	$\lambda = 0.25$	$c = 0$	HT	0.048	0.060	0.059	0.050	0.047	0.040	0.051	0.050
			KT	0.054	0.050	0.050	0.050	0.056	0.062	0.057	0.050
		$c = 1$	HT	0.775	0.853	0.894	0.938	0.076	0.068	0.065	0.050
			KT	0.352	0.428	0.474	0.583	0.087	0.073	0.069	0.050
	$\lambda = 0.5$	$c = 0$	HT	0.054	0.055	0.053	0.050	0.054	0.056	0.052	0.050
			KT	0.048	0.052	0.052	0.050	0.050	0.060	0.050	0.050
		$c = 1$	HT	0.768	0.828	0.866	0.901	0.065	0.060	0.046	0.050
			KT	0.487	0.546	0.608	0.682	0.073	0.061	0.060	0.050
	$\lambda = 0.75$	$c = 0$	HT	0.064	0.055	0.051	0.050	0.057	0.053	0.047	0.050
			KT	0.063	0.055	0.051	0.050	0.060	0.057	0.057	0.050
		$c = 1$	HT	0.889	0.906	0.926	0.938	0.061	0.065	0.052	0.050
			KT	0.375	0.453	0.490	0.583	0.102	0.080	0.062	0.050
$T = 10$	$\lambda = 0.25$	$c = 0$	HT	0.059	0.053	0.053	0.050	0.055	0.050	0.042	0.050
			KT	0.058	0.049	0.047	0.050	0.056	0.062	0.058	0.050
		$c = 1$	HT	0.900	0.960	0.973	0.993	0.095	0.070	0.068	0.050
			KT	0.288	0.384	0.458	0.585	0.108	0.087	0.074	0.050
	$\lambda = 0.5$	$c = 0$	HT	0.057	0.046	0.047	0.050	0.052	0.047	0.054	0.050
			KT	0.063	0.050	0.051	0.050	0.060	0.058	0.061	0.050
		$c = 1$	HT	0.878	0.927	0.957	0.976	0.070	0.063	0.054	0.050
			KT	0.451	0.527	0.603	0.720	0.090	0.073	0.055	0.050
	$\lambda = 0.75$	$c = 0$	HT	0.056	0.060	0.053	0.050	0.060	0.047	0.045	0.050
			KT	0.052	0.048	0.044	0.050	0.069	0.052	0.051	0.050
		$c = 1$	HT	0.940	0.968	0.976	0.985	0.083	0.069	0.059	0.050
			KT	0.339	0.456	0.541	0.653	0.092	0.078	0.064	0.050

Table 6: Simulated size and power of the KT test for model $M1$ with $u_{i,t} = \varepsilon_{i,t} + \theta\varepsilon_{i,t-1}$.

T			8				10			
N			100	300	1000	TV	100	300	1000	TV
$\theta = -0.8$	$\lambda = 0.25$	$c = 0$	0.047	0.049	0.053	0.050	0.058	0.056	0.048	0.050
		$c = 1$	0.060	0.057	0.063	0.001	0.047	0.051	0.054	0
	$\lambda = 0.50$	$c = 0$	0.047	0.056	0.053	0.050	0.050	0.045	0.060	0.050
		$c = 1$	0.052	0.054	0.054	0.082	0.054	0.058	0.068	0.058
	$\lambda = 0.75$	$c = 0$	0.056	0.053	0.059	0.050	0.049	0.055	0.047	0.050
		$c = 1$	0.054	0.061	0.049	0.358	0.049	0.046	0.047	0.205
$\theta = -0.5$	$\lambda = 0.25$	$c = 0$	0.052	0.053	0.044	0.050	0.049	0.058	0.044	0.050
		$c = 1$	0.070	0.0102	0.086	0.011	0.089	0.086	0.108	0.009
	$\lambda = 0.50$	$c = 0$	0.050	0.047	0.048	0.050	0.046	0.046	0.053	0.050
		$c = 1$	0.093	0.104	0.125	0.151	0.078	0.100	0.118	0.140
	$\lambda = 0.75$	$c = 0$	0.045	0.055	0.055	0.050	0.049	0.054	0.055	0.050
		$c = 1$	0.073	0.075	0.100	0.391	0.072	0.080	0.097	0.293
$\theta = 0.5$	$\lambda = 0.25$	$c = 0$	0.047	0.041	0.057	0.050	0.054	0.054	0.041	0.050
		$c = 1$	0.375	0.477	0.580	0.858	0.278	0.391	0.505	0.830
	$\lambda = 0.50$	$c = 0$	0.050	0.044	0.044	0.050	0.062	0.051	0.050	0.050
		$c = 1$	0.678	0.769	0.825	0.888	0.681	0.789	0.856	0.921
	$\lambda = 0.75$	$c = 0$	0.046	0.049	0.042	0.050	0.053	0.050	0.058	0.050
		$c = 1$	0.544	0.644	0.652	0.509	0.580	0.693	0.783	0.683
$\theta = 0.8$	$\lambda = 0.25$	$c = 0$	0.055	0.052	0.056	0.050	0.056	0.043	0.055	0.050
		$c = 1$	0.403	0.512	0.598	0.894	0.273	0.411	0.481	0.861
	$\lambda = 0.50$	$c = 0$	0.047	0.046	0.060	0.050	0.049	0.058	0.050	0.050
		$c = 1$	0.769	0.830	0.875	0.919	0.752	0.825	0.895	0.941
	$\lambda = 0.75$	$c = 0$	0.045	0.052	0.053	0.050	0.051	0.058	0.054	0.050
		$c = 1$	0.632	0.696	0.739	0.514	0.654	0.780	0.823	0.698

Table 7: Simulated size and power of the KT test for model $M2$ with $u_{i,t} = \varepsilon_{i,t} + \theta\varepsilon_{i,t-1}$										
T			8				10			
N			100	300	1000	TV	100	300	1000	TV
$\theta = -0.8$	$\lambda = 0.50$	$c = 0$	0.047	0.040	0.045	0.050	0.044	0.047	0.052	0.050
		$c = 1$	0.047	0.050	0.056	0.059	0.046	0.048	0.060	0.075
	$\lambda = 0.75$	$c = 0$					0.051	0.048	0.045	0.050
		$c = 1$					0.058	0.066	0.072	0.164
$\theta = -0.5$	$\lambda = 0.50$	$c = 0$	0.049	0.050	0.057	0.050	0.057	0.055	0.049	0.050
		$c = 1$	0.057	0.054	0.050	0.057	0.061	0.051	0.074	0.068
	$\lambda = 0.75$	$c = 0$					0.054	0.046	0.050	0.050
		$c = 1$					0.093	0.077	0.089	0.119
$\theta = 0.5$	$\lambda = 0.50$	$c = 0$	0.048	0.054	0.042	0.050	0.050	0.060	0.053	0.050
		$c = 1$	0.054	0.048	0.043	0.041	0.066	0.044	0.047	0.038
	$\lambda = 0.75$	$c = 0$					0.051	0.054	0.052	0.050
		$c = 1$					0.071	0.052	0.038	0.031
$\theta = 0.8$	$\lambda = 0.50$	$c = 0$	0.052	0.051	0.052	0.050	0.053	0.060	0.055	0.050
		$c = 1$	0.047	0.046	0.036	0.039	0.057	0.043	0.032	0.036
	$\lambda = 0.75$	$c = 0$					0.060	0.059	0.049	0.050
		$c = 1$					0.057	0.041	0.029	0.029

Table 8: Simulated size and power of the HT and KT tests for $\tilde{M}1$ and $\tilde{M}2$, with $u_{i,t} \sim NIID(0, \sigma^2)$

				Model $\tilde{M}1$				Model $\tilde{M}2$			
N				100	300	1000	TV	100	300	1000	TV
$T = 12$	$\lambda = 0.25$	$c = 0$	HT	0.055	0.053	0.052	0.050	0.052	0.047	0.053	0.050
			KT	0.056	0.056	0.052	0.050	0.057	0.055	0.051	0.050
		$c = 1$	HT	0.432	0.477	0.515	0.67	0.074	0.059	0.053	0.050
			KT	0.227	0.267	0.296	0.366	0.233	0.269	0.301	0.313
	$\lambda = 0.5$	$c = 0$	HT	0.052	0.056	0.049	0.050	0.054	0.052	0.052	0.050
			KT	0.053	0.048	0.053	0.050	0.044	0.049	0.051	0.050
		$c = 1$	HT	0.380	0.439	0.459	0.55	0.064	0.055	0.059	0.050
			KT	0.234	0.277	0.302	0.354	0.169	0.192	0.200	0.264
	$\lambda = 0.75$	$c = 0$	HT	0.057	0.054	0.049	0.050	0.050	0.051	0.050	0.050
			KT	0.057	0.056	0.053	0.050	0.049	0.044	0.049	0.050
		$c = 1$	HT	0.508	0.539	0.557	0.69	0.069	0.060	0.060	0.050
			KT	0.279	0.309	0.334	0.384	0.175	0.188	0.194	0.278
$T = 15$	$\lambda = 0.25$	$c = 0$	HT	0.052	0.056	0.044	0.050	0.053	0.057	0.054	0.050
			KT	0.054	0.052	0.058	0.050	0.056	0.052	0.056	0.050
		$c = 1$	HT	0.604	0.623	0.644	0.92	0.093	0.078	0.064	0.050
			KT	0.252	0.292	0.346	0.458	0.299	0.336	0.381	0.483
	$\lambda = 0.5$	$c = 0$	HT	0.053	0.056	0.053	0.050	0.048	0.050	0.049	0.050
			KT	0.058	0.049	0.054	0.050	0.050	0.061	0.047	0.050
		$c = 1$	HT	0.518	0.556	0.602	0.81	0.082	0.063	0.062	0.050
			KT	0.299	0.336	0.381	0.482	0.241	0.282	0.323	0.449
	$\lambda = 0.75$	$c = 0$	HT	0.060	0.058	0.050	0.050	0.050	0.052	0.055	0.050
			KT	0.058	0.061	0.050	0.050	0.053	0.052	0.050	0.050
		$c = 1$	HT	0.616	0.640	0.652	0.90	0.084	0.071	0.062	0.050
			KT	0.284	0.324	0.375	0.466	0.242	0.273	0.290	0.462

Table 9: Simulated size and power of the HT and KT tests when T is large and $u_{i,t} \sim NIID(0, \sigma^2)$

		N/T	10/100	10/200	10/300	10/50	20/50	50/50	TV
Model $M1$									
$\lambda = 0.25$	$c = 0$	HT	0.088	0.089	0.092	0.080	0.067	0.055	0.05
		KT	0.040	0.032	0.036	0.038	0.046	0.053	0.05
	$c = 1$	HT	0.140	0.135	0.148	0.138	0.122	0.109	0.10
		KT	0.033	0.031	0.032	0.042	0.049	0.056	0.05
$\lambda = 0.5$	$c = 0$	HT	0.075	0.074	0.077	0.067	0.062	0.055	0.05
		KT	0.038	0.038	0.036	0.048	0.050	0.054	0.05
	$c = 1$	HT	0.131	0.131	0.124	0.111	0.095	0.096	0.09
		KT	0.041	0.032	0.032	0.050	0.053	0.055	0.05
$\lambda = 0.75$	$c = 0$	HT	0.091	0.086	0.097	0.082	0.068	0.057	0.05
		KT	0.037	0.028	0.031	0.040	0.048	0.050	0.05
	$c = 1$	HT	0.146	0.143	0.157	0.134	0.115	0.108	0.10
		KT	0.036	0.033	0.029	0.044	0.051	0.055	0.05
Model $M2$									
$\lambda = 0.25$	$c = 0$	HT	0.080	0.077	0.085	0.069	0.067	0.065	0.05
		KT	0.038	0.034	0.034	0.044	0.050	0.056	0.05
	$c = 1$	HT	0.077	0.080	0.084	0.081	0.069	0.059	0.05
		KT	0.033	0.028	0.029	0.046	0.056	0.056	0.05
$\lambda = 0.5$	$c = 0$	HT	0.074	0.070	0.070	0.065	0.058	0.058	0.05
		KT	0.046	0.037	0.038	0.056	0.052	0.051	0.05
	$c = 1$	HT	0.062	0.075	0.071	0.068	0.060	0.061	0.05
		KT	0.044	0.036	0.046	0.054	0.054	0.056	0.05
$\lambda = 0.75$	$c = 0$	HT	0.075	0.085	0.077	0.080	0.067	0.060	0.05
		KT	0.042	0.033	0.032	0.052	0.052	0.057	0.05
	$c = 1$	HT	0.076	0.082	0.083	0.078	0.068	0.051	0.05
		KT	0.042	0.033	0.030	0.047	0.052	0.050	0.05

Figure 1: HT slopes in the absence of serial correlation

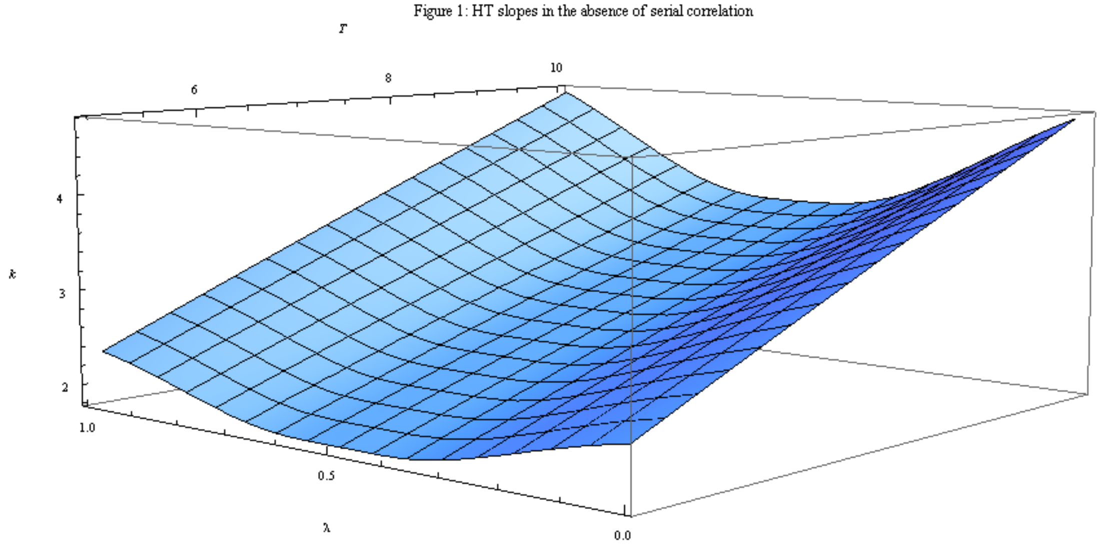

Figure 2: KT slopes in the absence of serial correlation