Local power of fixed-Tpanel unit root tests with serially correlated errors and incidental trends

Karavias, Yiannis; Tzavalis, Elias

DOI:
10.1111/jtsa.12144

Local power of fixed-T panel unit root tests with serially correlated errors and incidental trends

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

a :School of Economics, University of Nottingham

b :Department of Economics, Athens University of Economics & Business

March 2015

Abstract

The asymptotic local power properties of various fixed T panel unit root tests with serially correlated errors and incidental trends are studied. Asymptotic (over N) local power functions are analytically derived and through them the effects of general forms of serial correlation are examined. We find that a test based on an instrumental variables estimator dominates the tests based on the within groups estimator. These functions also show that in the presence of incidental trends an IV test based on the first differences of the model has non-trivial local power in a $N^{-1/2}$ neighbourhood of unity. Furthermore, for a test based on the within groups estimator, although it is found that it has trivial power in the presence of incidental trends, this ceases to be the case if there is serial correlation as well.

JEL classification: C22, C23

Keywords: Panel data; unit root; local power functions; serial correlation; incidental trends

Acknowledgements: We would like to thank the Editor Robert Taylor, an Associate Editor, two anonymous referees, Tassos Magdalinos, the seminar participants of the Granger Centre for Time Series Econometrics, the participants of the 12th Conference on Research on Economic Theory and Econometrics and the participants of the 20th International Panel Data Conference for for their helpful comments and suggestions.

*Corresponding author. E-mail: ioannis.karavias@nottingham.ac.uk (Y. Karavias), etzavalis@aueb.gr (E. Tzavalis).

1 Introduction

The power of time series unit root tests is greatly affected by the presence of serial correlation and linear trends. These phenomena are frequently encountered in empirical econometric work on unit root testing, but in many cases economic theory also predicts their existence.[1] Panel unit root tests are not very different in this respect from their time series counterparts, yet there is very little theoretical work in the area. The impact of dependence in the innovations has not been previously analysed while the effect of individual linear trends has been recently documented by Moon et al. (2007), for panel data unit root tests where the number of time series observations (T) is asymptotic.

This paper examines local power of finite-T panel unit root tests with serially dependent errors and individual linear trends. Finite-T tests advance the use of cross sectional information over time series information for inference and this characteristic makes them particularly interesting for time series analysis for the following two reasons. First, their sophisticated exploitation of the cross section dimension can improve large-T panel unit root tests. The latter are mostly based on averages of single time series tests and therefore don't use cross section information efficiently.[2] Second, finite-T panel unit root tests can have better performance than large-T tests even in large-T settings. This usually happens when the number of cross section units (N) is moderate or large, see e.g., Karavias and Tzavalis (2014a). This can be attributed to the fact that the estimation of the long-run variance of the errors, which is a difficult econometric task, can be avoided. For these reasons, the finite-T framework is relevant to both theoretical and empirical time series analysis.

The first contribution of this paper is that it derives local power functions which allow for general forms of serial dependence. This is novel for both the finite and large-T literatures on panel unit rot tests given that previous studies only consider independent innovations. It is shown that the effect of serial correlation on the power can be positive and depends on the estimation method used. The second contribution is related to the impact of individual linear trends in the data generating process. In the large-T literature (see e.g. Moon and Perron (2004), Moon et al. (2007)), the local power of various tests was examined and one important finding is that, when an AR(1) model with individual trends is used, local power is trivial in a $N^{-1/2}T^{-1}$ neighbourhood of unity and non-trivial only in a $N^{-1/4}T^{-1}$ neighbourhood. This paper finds that a test based on double differences of the data is powerful in the natural $N^{-1/2}$ neighbourhood of the null hypothesis. It is also shown that, in the presence of linear trends, the existence of serial correlation may also lead to non-trivial power.

Studies which examine local power in the fixed-T framework are those of Bond et al.

[1] See e.g. the works of Phillips (1987) and Schwert (1989) for the impact of serial correlation. The effect of a linear trend is studied in Elliot et al. (1996).

[2] A simple example is the now textbook large-T test of Breitung (2000) which employs the Helmert transformation frequently found in the analysis of short panels (see e.g. Arellano and Bover (1995)).

(2005), Kruiniger (2008) and Madsen (2010), but they only focus on models with individual intercepts and i.i.d. innovations. Since the focus of this paper is the local power of panel unit root tests which allow for individual trends and serial correlation, we need to choose tests which allow for them. Such available tests in the literature are those of Kruiniger and Tzavalis (2002) and De Blander and Dhaene (2012), or the test of De Wachter et al. (2007) for the case of individual intercepts. We focus on the tests of Kruiniger and Tzavalis (2002) and De Wachter et al. (2007), since we can derive analytical results.[3] Furthermore, although the testing methodologies between Kruiniger and Tzavalis (2002) and De Wachter et al. (2007) are very different, their assumptions are very similar. Thus, meaningful and fair comparrisons can be made. The Kruiniger's and Tzavalis (2002) tests are based on the within groups estimator and, in this way, they are similar to the Harris's and Tzavalis (1999) tests which are analysed by Madsen (2010) for the model with individual intercepts. They are however very different as, unlike the Harris' and Tzavalis (1999) tests, the Kruiniger and Tzavalis (2002) tests correct the estimator only for the bias of the numerator and thus, have a different behaviour, as it has been shown for their large-T counterparts by Moon and Perron (2008).

One issue arising within the general panel data unit root testing framework is homogeneity assumptions on the leading AR root. The null hypothesis of non-stationarity implies that all series have the same AR coefficient, i.e., for the AR(1) model with autoregressive parameters φ_i, $H_0 : \varphi_i = 1$, for all $i = 1, ..., N$. This assumption is reasonable in many frameworks where the null hypothesis is dictated by economic theory, i.e., when testing convergence hypotheses or hypotheses about hysteretic behaviours. Quah (1992) provides convincing arguments why pooling time series is useful and empirically relevant. Maddala and Wu (1999) are particularly critical on homogeneity assumptions in panel unit root tests; however, they also acknowledge that common coefficients under the null hypothesis are plausible in many cases (p. 635).[4] The alternative hypothesis of Kruiniger and Tzavalis (2002) and De Wachter et al. (2007) tests is also homogeneous, i.e., $H_1 : \varphi_i = \varphi < 1$, for all $i = 1, ..., N$, but in this paper we show that both tests have power against heterogeneous alternatives as well.

The paper is organized as follows. Section 2 introduces the finite-T panel unit root test statistics employed in our analysis and presents the required assumptions for the derivation of the asymptotic results. Section 3 derives the asymptotic local power functions and provides results on the behaviour of the tests. Section 4 conducts a Monte Carlo exercise to examine the small sample performance of the asymptotic theory and Section 5 concludes the paper.

[3] The tests suggested by De Blander and Dhaene (2012) are based on an Mean Unbiased Estimator for which analytical results are very difficult to derive.

[4] Specific examples of tests which are based on small T panel data sets can be found in Baltagi et al. (2007), with $N = 1000$ and $T = 14$, who test growth convergence, Canarella, et al. (2013), with $N = 1092$ and $T = 10$, who test for the existence of a competitive environment and in Nagayasou and Inakura (2009), with $N = 47$ and $T = 14$, who test the PPP, inter alia.

Appendix A contains some analytical results for MA(1) error terms. Appendix B contains all the proofs and appears in the Supporting Information which can be found online. In the following, we name the main diagonal of a matrix as "*diagonal* 0", the first upper diagonal as "*diagonal* +1", the first lower diagonal as "*diagonal* −1" etc.

2 Models and Assumptions

Consider the following first order autoregressive panel data models with individual effects:

$$M1 \quad : \quad y_i = \varphi y_{i-1} + (1-\varphi)\alpha_i e + u_i, \quad i = 1, ..., N. \tag{1}$$

$$M2 \quad : \quad y_i = \varphi y_{i-1} + (1-\varphi)\alpha_i e + \varphi \beta_i e + (1-\varphi)\beta_i \tau + u_i. \tag{2}$$

where $y_i = (y_{i1}, ..., y_{iT})'$ and $y_{i-1} = (y_{i0}, ..., y_{iT-1})'$ are $T \times 1$ vectors, u_i is the $T \times 1$ vector of error terms u_{it}, and α_i and β_i are the individual coefficients of the deterministic components of the models. α_i coefficients reflect individual effects of the panel, while β_i capture the slopes of individual linear trends, referred to as incidental trends. The $T \times 1$ vectors e and τ have elements $e_t = 1$ and $\tau_t = t$ for $t = 1, ..., T$, i.e. a constant and a linear trend.

To study the asymptotic local power of fixed-T unit root tests, define the autoregressive coefficient φ as $\varphi_N = 1 - c/\sqrt{N}$. Then, the hypothesis of interest becomes

$$H_0 \quad : \quad c = 0 \tag{3}$$

$$H_1 \quad : \quad c > 0, \tag{4}$$

where c is the local to unity parameter. The asymptotic distributions of fixed-T panel unit root test statistics allowing for serial correlation or heteroscedasticity in error terms u_{it} under the sequence of local alternatives φ_N can be derived by making the following assumptions.

Assumption 1: (1a)$\{u_i\}$ constitutes a sequence of independent, identical, normal random vectors of dimension $T \times 1$ with mean $E(u_i) = 0$ and variance-autocovariance matrix $E(u_i u_i') = \Gamma \equiv [\gamma_{ts}]$, where $\gamma_{ts} = E(u_{it} u_{is}) = 0$ for $s = t + p_{\max} + 1, ..., T$, and $p_{\max} \leq T - 2$. (1b) Δy_i are independent across $i = 1, ..., N$, have finite $(4 + \delta)$-th population moments, for $\delta > 0$, and $Var(vec(\Delta y_i \Delta y_i'))$ is a positive definite matrix.

Assumption 2: The errors u_{it} are independent of α_i, β_i and y_{i0} for $t = 1, ..., T$ and $i = 1, ..., N$ and $Var(y_{i0}) < +\infty$.

Assumption (1a) implies that the order of serial correlation of error term u_{it} can be at most $T - 2$. It requires the existence of at least one moment condition, in conducting inference about the true value of φ_N, which is free of correlation nuisance parameters. That is, it implies that at least $\gamma_{1T} = \gamma_{T1} = 0$. Assuming normality in the error terms allows for closed form representations of the variances of the limiting distributions of the tests.[5] Assumption (1.b) imposes finite fourth moments on initial conditions y_{i0}, error terms u_{it} and the individual specific coefficients α_i and β_i of models $M1$ and $M2$.

Assumption 2 is required only when $c > 0$. Under null hypothesis H_0: $c = 0$, all test statistics considered in the paper are invariant to y_{i0} and/or coefficients α_i and β_i. This is achieved either by subtracting y_{i0} from the levels of all individual series y_{it} of models $M1$ and $M2$ (see IV and FDIV statistics),[6] or by the within groups transformation of y_{it} (see WG and WGT statistics).[7] Under the local alternative hypothesis H_1: $c > 0$, the assumption that $Var(y_{i0}) < +\infty$ allows for constant, random and mean stationary initial conditions. Covariance stationarity of y_{i0} (see Kruiniger (2008) and Madsen (2010)), although it does not affect the limiting distributions under the null, is not allowed under the local alternatives. This is because, as also noted by Moon et al. (2007), it implies that $Var(y_{i0}) \to \infty$ when $\varphi_N \to 1$, which means that the initial condition will dominate the sample data $y_{it.}$

To study the asymptotic local power of the tests, we employ a "slope" parameter, denoted as k, which is found in 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 if it is negative they will be biased.

3 Asymptotic local power functions

This section presents the fixed-T panel unit root test statistics considered and it derives their limiting distributions under the sequence of local alternatives. The first part of the section presents results for model $M1$, while the second for model $M2$.

[5] General representations of the asymptotic local power functions can be straightforwardly derived under non-normality or cross section heterogeneity of the error term. The intuition and analytic results are similar. Thus, for ease of exposition we do not consider these cases.

[6] This approach is suggested by Schmidt and Phillips (1992), for single time series, and Breitung and Meyer (1994) panel data models with individual effects.

[7] This transformation means that one subtracts the means of the individual series of the panel from their levels, across all units i. It is also employed by the panel unit root tests of Harris and Tzavalis (1999), and Levin et al. (2002).

3.1 Individual intercepts

The IV panel unit root test statistic (see De Wachter et al. (2007)): This test statistic assumes an order of serial correlation $p \leq T-2$ and it is based on transformation of the individual series of the panel in deviations from their initial conditions, given as $z_{it} = y_{it} - y_{i0}$. The statistic becomes invariant to the serial correlation effects by exploiting the following moment conditions:

$$E\left[\sum_{t=1}^{T-p-1} z_{it} u_{i,t+p+1}(\varphi)\right] = 0, \ i = 1, ..., N, \tag{5}$$

and it is based on the IV estimator

$$\hat{\varphi}_{IV} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T-p-1} z_{it} z_{it+p}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T-p-1} z_{it} z_{it+p+1}\right). \tag{6}$$

The moments given by (5) can be rewritten in matrix notation as follows:

$$E(z'_{i-1}\Pi_p u_i) = 0, \tag{7}$$

where Π_p is a $T \times T$ matrix selecting zero-mean moments, according to (5), and $z_{i-1} = y_{i-1} - y_{i0}e$.[8] In particular, Π_p has ones in the "*diagonal* $+p$" and zeroes everywhere else.[9] Given the definition of Π, the above IV estimator can be rewritten as

$$\hat{\varphi}_{IV} = \left(\sum_{i=1}^{N} z'_{i-1}\Pi_p z_{i-1}\right)^{-1}\left(\sum_{i=1}^{N} z'_{i-1}\Pi_p z_i\right) \tag{8}$$

The asymptotic distribution of the IV based unit root test statistic under the sequence of local alternatives $\varphi_N = 1 - c/\sqrt{N}$ is derived in the next theorem.

Theorem 1 *Under Assumptions 1 and 2, we have*

$$\sqrt{N}\hat{V}_{IV}^{-1/2}(\hat{\varphi}_{IV} - 1) \xrightarrow{d} N(-ck_{IV}, 1), \tag{9}$$

[8] De Wachter et al. (2007) also propose a GMM counterpart of this statistic, but this general set-up is not considered as the results are less tractable because of the optimal weight matrix. Furthermore, as argued in Han and Phillips (2010), the efficiency gains of GMM are marginal.

[9] For example, if $T = 4$ we have

$$\Pi_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \text{ and } \Pi_2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

as $N \to \infty$, where

$$k_{IV} = \frac{1}{\sqrt{V_{IV}}} \tag{10}$$

and $V_{IV} = 2tr((A_{IV}\Gamma)^2)/tr(\Lambda'\Pi_p\Lambda\Gamma)^2$, with $A_{IV} = (\Lambda'\Pi_p + \Pi_p'\Lambda)/2$, is the variance of the limiting distribution of $\hat{\varphi}_{IV}$. Λ is a $T \times T$ matrix defined as $\Lambda = [\lambda_{\mu\nu}]$, where $\lambda_{\mu\nu} = 1$ if $\mu < \nu$ and $\lambda_{\mu\nu} = 0$ otherwise, with $\mu, \nu = 1, ..., T$.

The limiting distribution of the IV test statistic given by Theorem 1 nests the distributions under the null and alternative hypotheses H_0: $c = 0$ and H_1: $c > 0$, respectively. For $c = 0$, (9) gives the distribution of the test statistic under H_0, derived by De Wachter et al. (2007). The test statistic of Breitung and Meyer (1994) can be seen as a special case of the IV test, for $p = 0$.[10] The only unknown quantity in the variance function V_{IV} is Γ. If $\Gamma = \sigma_u^2 I_T$, where I_T is the $T \times T$ identity matrix, then no estimation is needed since σ_u^2 cancels out from both the nominator and denominator, i.e., $2tr((A_{IV}\Gamma)^2)/tr(\Lambda'\Pi_p\Lambda\Gamma)^2 = 2tr((A_{IV})^2)/tr(\Lambda'\Pi_p\Lambda)^2$. In the more general case that $\Gamma \neq \sigma_u^2 I_T$, an estimator of Γ can be obtained under null hypothesis H_0: $c = 0$ as

$$\hat{\Gamma} = \frac{1}{N}\sum_{i=1}^{N} \Delta y_i \Delta y_i', \tag{11}$$

since $\Delta y_i = u_i$ under this hypothesis.

The results of Theorem 1 show that the IV test statistic has always non-trivial power, since the slope parameter of the local power function k_{IV} is always positive. This parameter depends on the time dimension of the panel T, the order of serial correlation p and the form of serial correlation considered by variance-covariance matrix Γ.

Next, we present more analytically how serial correlation in u_{it} affects the slope parameter k_{IV}. This is done for the case that u_{it} follows a MA(1) model, i.e.,

$$u_{it} = v_{it} + \theta v_{it-1}, \text{ for all } i, \tag{12}$$

with $v_{it} \sim NIID(0, \sigma_u^2)$ and $|\theta| < 1$. MA(q) models are particularly interesting because they are documented in many economic series, see Schwert (1989) and Phillips (1987). In Section 4, we also analyze the case of an AR(1) model of u_{it}. For the above MA(1) model, the closed form of k_{IV}, defined as $k_{IV}(p, \theta)$, for different values of p and θ, is given in Appendix A. This can be employed to examine how the values of nuisance parameter θ affect the local power of the IV based panel unit root test statistic. To this end, Figure 1 presents values of $k_{IV}(p, \theta)$ across T, for $p \in \{0, 1\}$ and $\theta \in \{-0.9, -0.5, 0, 0.5, 0.9\}$. Inspection of Figure 1 clearly indicates that the IV test statistic has its maximum asymptotic local power, when

[10] As Bond et al. (2005) show, in this case $\hat{\varphi}_{IV}$ can be also seen as a maximum likelihood estimator of φ.

$p = 0$ and $\theta = 0$. This can be attributed to the fact that, in this case, the test exploits the maximum number of possible moment conditions in (5). If $p = 1$ (implying that one moment condition is lost), then the power of the test decreases. Finally, the test has much higher power if $\theta > 0$ than $\theta < 0$. This can be attributed to the fact that $\theta > 0$ increases the variability of y_{it}, thus making it easier for the test to distinguish between hypotheses H_0: $c = 0$ and H_1: $c > 0$. In this case, the variance of estimator $\hat{\varphi}_{IV}$ decreases. On the other hand, $\theta < 0$ reduces the variability of y_{it} and thus, the IV test statistic is harder to distinguish H_0: $c = 0$ from H_1: $c > 0$. Independently of the sign of θ, the plotted values of $k_{IV}(p, \theta)$, given by Figure 1, clearly indicate that the power of the IV test increases with T.

The WG panel unit root test statistic (see Kruiniger and Tzavalis (2002)): This test statistic becomes invariant to initial conditions y_{i0} of the panel by taking the within groups transformation of the individual series y_{it}, using the annihilator matrix $Q = I_T - e(e'e)^{-1}e'$, where I_T is the $T \times T$ identity matrix. Then, the least squares estimator of the transformed series is given as

$$\hat{\varphi}_{WG} = \left(\sum_{i=1}^{N} y'_{i-1} Q y_{i-1} \right)^{-1} \left(\sum_{i=1}^{N} y'_{i-1} Q y_i \right). \tag{13}$$

Since $\hat{\varphi}_{WG}$ is not a consistent estimator of φ, due to the above transformation of y_{it} and the presence of serial correlation in error terms u_{it}, Kruiniger and Tzavalis (2002) suggested the following fixed-T WG test statistic:

$$\sqrt{N}\hat{\delta}_{WG} \left(\hat{\varphi}_{WG} - 1 - \frac{\hat{b}_{WG}}{\hat{\delta}_{WG}} \right) \xrightarrow{d} N(0, V_{WG}), \tag{14}$$

$$\text{or} \quad \sqrt{N} V_{WG}^{-1/2} \hat{\delta}_{WG} \left(\hat{\varphi}_{WG} - 1 - \frac{\hat{b}_{WG}}{\hat{\delta}_{WG}} \right) \xrightarrow{d} N(0, 1),$$

which corrects estimator $\hat{\varphi}_{WG}$ for the above two sources of its inconsistency, where $\hat{\delta}_{WG} = (1/N) \sum_{i=1}^{N} y'_{i-1} Q y_{i-1}$ is the denominator of estimator $\hat{\varphi}_{WG}$ scaled by N, $\hat{b}_{WG} = tr(\Psi_{p,WG}\hat{\Gamma})$, $\hat{b}_{WG}/\hat{\delta}_{WG}$ is a consistent estimator of the inconsistency of $\hat{\varphi}_{WG}$ given as $tr(\Lambda' Q \Gamma)/tr(\Lambda' Q \Lambda \Gamma)$ and $\Psi_{p,WG}$ is a $T \times T$ dimension selection matrix having in its $-p, ..., 0, ..., p$ diagonals the corresponding elements of matrix $\Lambda' Q$, and zero everywhere else. $\hat{\Gamma} = (1/N) \sum_{i=1}^{N} \Delta y_i \Delta y'_i$ and $V_{WG} = 2tr((A_{WG}\Gamma)^2)$ is the variance of the limiting distribution of the corrected for its inconsistency WG estimator $\hat{\varphi}_{WG}$, where $A_{WG} = (\Lambda' Q + Q\Lambda - \Psi_{p,WG} - \Psi'_{p,WG})/2$.[11] This

[11] Note that the WG test statistic, given by (14), has been reformulated to avoid computing selection matrix S of Kruiniger and Tzavalis (2002), which is very demanding. The relationship between the two alternative formulations of the test statistics can be seen by noticing that

$$tr(\Psi_{p,WG}\hat{\Gamma}) = vec(Q\Lambda) S \left(\frac{1}{N} \sum_{i=1}^{N} vec(\Delta y_i \Delta y'_i) \right)$$

and

variance can be consistently estimated provided consistent estimates of Γ. As for the IV test statistic, this can be done based on (11). Note that the WG test is different from Harris' and Tzavalis (1999). Although both are based on the same estimator, the WG test corrects the estimator only for the bias of the numerator and not for the bias of both the numerator and the denominator, as the Harris and Tzavalis (1999) test does. For the implications of this difference, see also Moon and Perron (2008) and Hahn and Kuersteiner (2002).

The WG unit root test statistic is based on the same testing principle with the IV test statistic, described above. It exploits moments of the numerator of $\hat{\varphi}_{WG}$ which have zero mean under H_0: $c = 0$. But, this now is done for the corrected for its inconsistency estimator $\hat{\varphi}_{WG} - 1 - \hat{b}_{WG}/\hat{\delta}_{WG}$ through the selection matrix $\Psi_{p,WG}$.[12] Moon and Perron (2008) have suggested a version of the WG test statistic for the case that both N and T go to infinity. The next theorem gives the limiting distribution of the WG statistic under the sequence of local alternatives $\varphi_N = 1 - c/\sqrt{N}$.

Theorem 2 *Under Assumptions 1 and 2, we have*

$$\sqrt{N}\hat{V}_{WG}^{-1/2}\hat{\delta}_{WG}\left(\hat{\varphi}_{WG} - 1 - \frac{\hat{b}_{WG}}{\hat{\delta}_{WG}}\right) \xrightarrow{d} N(-ck_{WG}, 1), \tag{15}$$

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

$$k_{WG} = \frac{tr(\Lambda' Q \Lambda \Gamma) + tr(F' Q \Gamma) - tr(\Psi_{p,WG}\Lambda\Gamma) - tr(\Lambda'\Psi_{p,WG}\Gamma)}{\sqrt{V_{WG}}} \tag{16}$$

and $F = \frac{d\Omega}{d\varphi}\mid_{\varphi=1}$, *where* Ω *is given in Appendix B.*

The results of Theorem 2 indicate that annihilator matrix Q and the inconsistency correction of estimator $\hat{\varphi}_{WG}$, $\hat{b}_{WG}/\hat{\delta}_{WG}$, based on $\Psi_{p,WG}$, complicate the local power function. As equation (16) shows, the slope parameter of this function k_{WG} depends on the following quantities: $tr(\Lambda' Q \Lambda \Gamma)$, $tr(F' Q \Gamma)$, $tr(\Psi_{p,WG}\Lambda\Gamma)$ and $tr(\Lambda'\Psi_{p,WG}\Gamma)$. The first two quantities come from the annihilator matrix Q and the last two from selection matrix $\Psi_{p,WG}$. Note again that k_{WG} depends on T, p and Γ, but we supress notation for ease of exposition until the specific case of MA(1) errors is discussed. For $p = 0$, the effects of matrix $\Psi_{p,WG}$ disappear, since $tr(\Psi_{p,WG}\Lambda\Gamma) = tr(\Lambda'\Psi_{p,WG}\Gamma) = 0$.

$$2tr((A_{WG}\Gamma)^2) = vec(Q\Lambda)'(I_{T^2} - S)Var(vec(\Delta y_i \Delta y_i'))(I_{T^2} - S)vec(Q\Lambda),$$

where I_{T^2} is the (T^2XT^2) identity matrix and S is a (T^2XT^2) diagonal selection matrix, with elements s_{st} defined as $s_{(s-1)T+t,(s-1)T+t} = 1 - d(\gamma_{ts} = 0)$ with $s, t = 1, 2, ..., T$ and $d(.)$ is the Dirac function.

[12] To understand more clearly the role of selection matrix $\Psi_{p,WG}$, assume $T = 3$ and consider that error terms u_{it} follow $MA(1)$ process (12). Then, matrix Γ becomes $\Gamma = \begin{pmatrix} \sigma_u^2(1+\theta^2) & \sigma_u^2\theta & 0 \\ \sigma_u^2\theta & \sigma_u^2(1+\theta^2) & \sigma_u^2\theta \\ 0 & \sigma_u^2\theta & \sigma_u^2(1+\theta^2) \end{pmatrix}$ and $\Psi_{1,WG}$ is given as $\Psi_{1,WG} = \begin{pmatrix} -\frac{2}{3} & -\frac{1}{3} & 0 \\ \frac{1}{3} & -\frac{1}{3} & 0 \\ 0 & \frac{2}{3} & 0 \end{pmatrix}$.

To study the effects of the serial correlation nuisance parameters and lag-order p on k_{WG}, in Appendix A we derive analytic formulas of k_{WG}, for $p \in \{0, 1\}$ and $\theta \in \{-0.9, -0.5, 0, 0.5, 0.9\}$, while in Figure 2 we plot values of these formulas across T. As can be seen from this figure, the effects of θ and p on the power of the WG test differ from those on the power of the IV test. This can be attributed to the within groups transformation of individual series y_{it} and the correction of estimator $\hat{\varphi}_{WG}$ for its inconsistency. For positive values of θ, the WG test statistic has more power than for $\theta = 0$. For $\theta > 0$, the power also increases with T. These results are in contrast to those for the IV test statistic. For θ negative, the WG test statistic becomes biased, something that never happens for the IV test statistic. This happens because the inconsistency correction affects slope parameter $k_{WG}(p, \theta)$ through the quantity $tr(\Psi_{p,WG}\Lambda\Gamma) + tr(\Lambda'\Psi_{p,WG}\Gamma)$. For $\theta < 0$, this quantity takes positive values and, thus, reduces the power of the WG test statistic. For $\theta > 0$, it becomes negative and thus, it moves the limiting distribution towards the critical region, increasing the power of the test. As T increases, the above sign effects of θ on the WG test statistic are amplified. That is, they lead to a test with greater power or bias, if $\theta > 0$ and $\theta < 0$, respectively. Finally, comparison between $k_{WG}(p, \theta)$ and $k_{IV}(p, \theta)$ reveals that the IV test is more powerful than the WG test statistic. This is true for all values of θ and p considered, and across T. It can be also seen by the results of Table 1, which presents values of slope parameter k for the IV and WG test statistics for $T \in \{7, 10\}$, $p \in \{0, 1\}$ and $\theta \in \{-0.9, -0.5, 0, 0.5, 0.9\}$. Extensions to higher order of serial correlation are conceptually similar but less tractable.

The limiting distributions of the IV and WG test statistics given by Theorems 1 and 2, respectively, scaled appropriately by T become invariant to the serial correlation nuisance parameters, if $T, N \to \infty$ jointly. This result is established in the next proposition, which derives the limiting distributions of the scaled by T versions of the IV and WG test statistics under the following sequence of local alternatives:

$$\varphi_{NT} = 1 - \frac{c}{T\sqrt{N}},$$

considered in the large-T panel data literature (see, e.g., Moon et al. (2007)).

Proposition 1 *Let Assumptions 1 and 2 hold. Then, under $\varphi_{NT} = 1 - c/T\sqrt{N}$, we have*

$$T\sqrt{N}(\sqrt{2})^{-1}(\hat{\varphi}_{IV} - 1) \xrightarrow{d} N(-c\bar{k}_{IV}, 1), \qquad \textit{and} \tag{17}$$

$$T\sqrt{N}(\sqrt{3})^{-1}\left(\hat{\varphi}_{WG} - 1 - \frac{\hat{b}_{WG}}{\hat{\delta}_{WG}}\right) \xrightarrow{d} N(-c\bar{k}_{WG}, 1), \tag{18}$$

where

$$\bar{k}_{IV} = \frac{1}{\sqrt{2}} \textit{ and } \bar{k}_{WG} = 0,$$

if $T, N \to \infty$ jointly and the following condition holds: $\sqrt{N}/T \to 0$.

Condition $\sqrt{N}/T \to 0$ is required only under alternative hypothesis H_1: $c > 0$. Under null hypothesis H_0: $c = 0$, it is not needed (see, e.g., Harris and Tzavalis (1999, 2004), and Hahn and Kuersteiner (2002)). The results of the proposition apply for every fixed order of serial correlation p and any form of short term serial correlation. For $c = 0$, the limiting distribution of estimator $\hat{\varphi}_{IV}$, given by (17), coincides with that derived by De Wachter et al. (2007), while the limiting distribution of estimator $\hat{\varphi}_{WG}$ adjusted for its inconsistency corresponds to that derived by Moon and Perron (2008).

For $c > 0$, the IV test reaches the maximum local power which is equal to that of the common-point optimal test of Moon et al. (2007), denoted as MPP. However, the WG test has trivial power, since $\bar{k}_{WG} = 0$. This result confirms the finding of Moon and Perron (2008) who additionally show that the WG test has non-trivial power in a $n^{-1/4}T$ neighbourhood of the null hypothesis. From our analysis, it becomes more clear that the reason behind this behaviour is the inefficient use of the time series observations. As can be seen in Figure 2, a larger T does not give proportionally larger power. For comparison, Table 2 presents the values of k of the IV and WG tests, for large-T, along with those of the MPP test, the test by Levin et al. (2002) (LLC), the test by Im et al. (2003) (denoted IPS) , and Sargan's (SGLS) test, which are derived in Moon et al. (2007), Moon and Perron (2008) and Harris et al. (2010).

Following the large-T literature, the analysis of the local power can also be done under heterogeneous alternatives i.e. under $H_1 : c_i \neq 0$ for some i's, with $\varphi_{Ni} = 1 - c_i/\sqrt{N}$ and c_i being *i.i.d.* with support in a subset of a bounded interval $[0, M_c]$, for some $M_c \geq 0$. In this case the above results change, with $E(c_i)$ taking the place of c. The new null hypothesis is $H_0 : E(c_i) = 0$ (see also Moon et al. (2007)). As the rate of convergence is $\sqrt{N}$, local power is only affected by the mean of c_i and not by higher moments of their distribution, as in Westerlund and Larsson (2013). A more thorough discussion on the fact that power only depends on the mean of c_i can be found in Westerlund and Breitung (2013). Overall, the higher the mean of c_i the more power the tests have.

3.2 Incidental trends

To study the power of fixed-T panel data unit root tests allowing for serial correlation in the case of incidental trends, first we extend the IV test presented in the previous section. This extension requires that the IV test is based on a first difference of panel data series y_{it}, and it will be denoted as $FDIV$.

FDIV panel unit root test: Taking first differences of model $M2$, and for $T \geq 4$, yields:

$$\Delta y_i = \varphi \Delta y_{i-1} + (1 - \varphi)\beta_i e^* + \Delta u_i, \quad i = 1, ..., N, \tag{19}$$

where $y_i = (y_{i2}, ..., y_{iT})'$, $y_{i-1} = (y_{i1}, ..., y_{iT-1})'$, $y_{i-2} = (y_{i0}, ..., y_{iT-2})'$, $u_i = (u_{i2}, ..., u_{iT})'$, $u_{i-1} = (u_{i1}, ..., u_{iT-1})'$ and $e^* = (1, 1, ..., 1)$ are $(T-1) \times 1$ vectors. Subtracting from both sides of model (19), the vector of the initial observation $\Delta y_{i1} e$ gives the following first differences transformation of the model:

$$y_i^* = \varphi y_{i-1}^* + (1-\varphi)\alpha_i^* e^* + u_i^*, \quad i = 1, ..., N, \tag{20}$$

where $y_i^* = \Delta y_i - \Delta y_{i1} e$, $y_{i-1}^* = \Delta y_{i-1} - \Delta y_{i1} e$, $\alpha_i^* = (\beta_i - \Delta y_{i1})$ and $u_i^* = \Delta u_i$. The above model shows that, if error terms u_{it} are serially correlated (and thus also u_{it}^*), moments similar to (7) can be exploited to test the null hypothesis of a unit root, i.e.

$$E(y_{i-1}^{*\prime} \Pi_p^* u_i^*) = 0, \tag{21}$$

where Π_p^* is a $(T-1) \times (T-1)$ matrix with unities in its $p+1$ diagonal, and zeros everywhere else. If we define $E(u_i^* u_i^{*\prime}) = \Theta$, then. a consistent estimator of Θ under H_0: $c = 0$ is given as

$$\hat{\Theta} = \frac{1}{N} \sum_{i=1}^{N} \Delta y_i^* \Delta y_i^{*\prime}, \tag{22}$$

which corresponds to (11), for $\Delta y_i = u_i$. It can be easily seen that $\Theta = 2\Gamma - \Gamma_1 - \Gamma_1'$, where $\Gamma = E(u_i u_i')$ and $\Gamma_1 = E(u_i u_{i-1}')$. But, as will be thoroughly explained later on, Γ and Γ_1 cannot be consistently estimated under H_0: $c = 0$ based on Δy_i due to the presence of incidental trends. Theorem 3 derives the limiting distribution of the IV estimator under the sequence of local alternatives $\varphi_N = 1 - c/\sqrt{N}$, exploiting the above moment conditions.

Theorem 3 *Under Assumptions 1 and 2, we have*

$$\sqrt{N} \hat{V}_{FDIV}^{-1/2} (\hat{\varphi}_{FDIV} - 1) \xrightarrow{d} N(-ck_{FDIV}, 1), \tag{23}$$

$N \to \infty$, where

$$k_{FDIV} = \frac{tr(\Lambda^{*\prime} \Pi_p^* \Lambda^* \Theta)}{\sqrt{2tr((A_{FDIV}\Theta)^2)}} \tag{24}$$

and $\hat{\varphi}_{FDIV} = \left(\sum_{i=1}^{N} y_{i-1}^{*\prime} \Pi_p^* y_{i-1}^*\right)^{-1} \left(\sum_{i=1}^{N} y_{i-1}^{*\prime} \Pi_p^* y_i^*\right)$, $V_{FDIV} = 2tr((A_{FDIV}\Theta)^2)/tr(\Lambda^{*\prime}\Pi_p^*\Lambda^*\Theta)^2$, $A_{FDIV} = (\Lambda^{*\prime}\Pi_p^* + \Pi_p^{*\prime}\Lambda^*)/2$. *$\Lambda^*$ is a $(T-1) \times (T-1)$ version of Λ.*

The results of Theorem 3 indicate that the FDIV test has non-trivial local power in the natural $(\sqrt{N})^{-1}$ neighbourhood of unity. This is a major deviation from the large-T unit roots literature and it highlights the different nature of asymptotics used in this paper. As with the IV test, the power of the $FDIV$ test statistic depends on the serial correlation nuisance parameters and lag-order p, as well as the time dimension of the panel.

In Appendix A, we derive the function of the slope parameter k_{FDIV}, if error terms u_{it} follow MA(1) process. Table 3 presents values of $k_{FDIV}(p,\theta)$, obtained through relationship (38), for $p=\{0,1\}$, $T\in\{7,10\}$ and $\theta\in\{-0.9,-0.5,0,0.5,0.9\}$. The results of the table indicate that the FDIV test has non-trivial power, for all values of p and θ considered. The power of the test increases slowly with T, as with the WG test. The asymptotic power of the FDIV test comes from the assumption that T is fixed and, additionally, by the presence of serial correlation. A positive value of θ tends to increase the power of the test, as it happens with the IV test for model $M1$.

The WG unit root statistic: The version of the WG test statistic in the case of incidental trends (denoted as WGT) considers an augmented annihilator matrix, given as $Q^*=I_T-X(X'X)^{-1}X'$, where $X=[e,\tau]$. Under null hypothesis H_0: $c=0$, multiplying model $M2$ with Q^* leads to a transformed model without individual effects and incidental trends. The WGT test statistic is based on the least squares estimator of the autoregressive coefficient φ of the transformed model, denoted as $\hat{\varphi}_{WGT}$. As with $\hat{\varphi}_{WGT}$, this estimator is adjusted for its inconsistency. The latter is due to the above transformation of individual series y_{it} and the presence of serial correlation in error terms u_{it}. To correct $\hat{\varphi}_{WGT}$ for its inconsistency coming from the serial correlation in u_{it}, we can no longer rely on the previous estimator of variance-covariance matrix Γ, $\hat{\Gamma}$, given as $\hat{\Gamma}=(1/N)\sum_{i=1}^{N}\Delta y_i\Delta y_i'$ (see (11)). This happens because Δy_i depends on the nuisance parameters of the incidental trends β_i, for model $M2$, i.e.

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

which implies

$$p\lim_{N\to\infty}\hat{\Gamma}=p\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\Delta y_i\Delta y_i'=\Gamma+E(\beta_i^2)ee'. \tag{25}$$

To remove the effects of β_i from the estimator of matrix Γ, the following selection matrix will be defined.[13] Let matrix M have elements $m_{ts}=0$ if $\gamma_{ts}\neq 0$ and $m_{ts}=1$ if $\gamma_{ts}=0$. Then, $tr(M\Gamma)=0$ and, thus, we have

$$p\lim_{N\to\infty}\frac{1}{tr(Mee')N}\sum_{i=1}^{N}\Delta y_i' M\Delta y_i=E(\beta_i^2). \tag{26}$$

The last relationship can be employed to substitute out individual effects $E(\beta_i^2)$ from (25), and thus to provide a consistent estimator of Γ and $tr(\Lambda' Q^*\Gamma)$ under null hypothesis H_0: $c=0$ which is net of β_i. Based on relationships (25) and (26), we can define selection matrix $\Phi_{p,WGT}=\Psi_{p,WGT}-(tr(\Lambda' Q^* M)/e'Me)\,M$, where $\Psi_{p,WGT}$ is a $T\times T$ matrix having in its

[13] Note that, as in case of model $M1$ (see fn 11), this selection matrix simplifies considerably the computation of the WGT test statistic, compared with the selection matrix S used by Kruiniger's and Tzavalis (2002).

diagonals $\{-p, .., 0, ..., p\}$ the corresponding elements of matrix $\Lambda' Q^*$, and zero everywhere else. This matrix has the property $tr(\Phi_{p,WGT} ee') = 0$, which leads to the following consistent estimator of $tr(\Lambda' Q^* \Gamma)$:

$$p \lim_{N \to \infty} tr(\Phi_{p,WGT} \hat{\Gamma}) = tr(\Lambda' Q^* \Gamma). \tag{27}$$

The limiting distribution of $\hat{\varphi}_{WGT}$ corrected for its inconsistency under $\varphi_N = 1 - c/\sqrt{N}$ is given in the next theorem.

Theorem 4 *Under Assumptions 1 and 2, we have*

$$\sqrt{N} \hat{V}_{WGT}^{-\frac{1}{2}} \hat{\delta}_{WGT} \left(\hat{\varphi}_{WGT} - 1 - \frac{\hat{b}_{WGT}}{\hat{\delta}_{WGT}} \right) \xrightarrow{d} N(-c k_{WGT}, 1), \tag{28}$$

$N \to +\infty$, where

$$k_{WGT} = \frac{tr(\Lambda' Q^* \Gamma) + tr(F' Q^* \Gamma) - tr(\Phi_{p,WGT} \Lambda \Gamma) - tr(\Lambda' \Phi_{p,WGT} \Gamma)}{\sqrt{V_{WGT}}}, \tag{29}$$

$\hat{\varphi}_{WGT} = \left(\sum_{i=1}^N y_{i-1}' Q^* y_{i-1} \right)^{-1} \left(\sum_{i=1}^N y_{i-1}' Q^* y_i \right)$, $\hat{b}_{WGT} / \hat{\delta}_{WGT} = tr(\Phi_{p,WGT} \hat{\Gamma}) / \left((1/N) \sum_{i=1}^N y_{i,-1}' Q^* y_{i,-1} \right)$, *and* $V_{WGT} = vec(Q^* \Lambda - \Phi_{p,WGT}')' \Xi vec(Q^* \Lambda - \Phi_{p,WGT}')$, *with* $\Xi = (1/N) \sum_{i=1}^N Var(vec(\Delta y_i \Delta y_i'))$, *is the variance of the limiting distribution of the WGT test.*

The implementation of the WG test statistic is based on the estimator of Ξ given by $\hat{\Xi} = (1/N) \sum_{i=1}^N vec(\Delta y_i \Delta y_i') vec(\Delta y_i \Delta y_i')'$. If $\beta_i = 0$ for all i, then $V_{WGT} = 2tr((A_{WGT} \Gamma)^2)$ with $A_{WGT} = (\Lambda' Q^* + Q^* \Lambda - \Phi_{p,WGT} - \Phi_{p,WGT}')/2$, and if β_i are normally distributed, then $V_{WGT} = 2tr((A_{WGT} \Gamma + E(\beta_i^2) A_{WGT} ee')^2)$. The results of Theorem 4 imply that the test statistic WGT also has non-trivial power. This result is shown in Appendix A, where a function of the slope parameter k_{WGT}, denoted $k_{WGT}(p, \theta)$, is derived under the MA process (12) of u_{it}, for different values of p and θ. Values of $k_{WGT}(p, \theta)$, for $p = \{0, 1\}$, $T \in \{7, 10\}$ and $\theta \in \{-0.9, -0.5, 0, 0.5, 0.9\}$, are given in Table 3. These indicate that test statistic WGT has asymptotic local power, if $\theta < 0$. This power is less than that of the FDIV for $\theta < 0$, and it increases slowly with T. It can be attributed to the effects of quantities $tr(\Phi_{p,WGT} \Lambda \Gamma)$ and $tr(\Lambda' \Phi_{p,WGT} \Gamma)$ on slope parameter $k_{WGT}(p, \theta)$.

The following Proposition documents the large-T behaviour of the tests FDIV and WGT.

Proposition 2 *Let Assumptions 1 and 2 hold. Then, under $\varphi_{NT} = 1 - c/T\sqrt{N}$, we have*

$$T \sqrt{N} \hat{V}_{FDIV}^{-1/2} \left(\hat{\varphi}_{FDIV} - 1 \right) \xrightarrow{d} N(-c \bar{k}_{FDIV}, 1), \qquad \textit{and} \tag{30}$$

$$T \sqrt{N} \hat{V}_{WGT}^{-\frac{1}{2}} \hat{\delta}_{WGT} \left(\hat{\varphi}_{WGT} - 1 - \frac{\hat{b}_{WGT}}{\hat{\delta}_{WGT}} \right) \xrightarrow{d} N(-c \bar{k}_{WGT}, 1), \tag{31}$$

where

$$\bar{k}_{FDIV} = 0 \text{ and } \bar{k}_{WGT} = 0,$$

if $T, N \to \infty$ *jointly and the following condition holds:* $\sqrt{N}/T \to 0$.

If $T \to \infty$, it is be shown that $k^*_{FDIV} = (T-p-3) / \left(T\sqrt{2(T-p-2)}\right) \to 0$, which means that the IV test has trivial power in the natural $N^{-1/2}T^{-1}$ neighbourhood of the null hypothesis. With respect to the time series dimension, the test has power in a $T^{-1/2}$ neighbourhood which is a considerable reduction to the T^{-1} neighbourhood usually found in the literature. As with the FDIV test, it is shown that the large-T version of the WGT test has also trivial local power. These results are in line with previous findings in the literature, see e.g. Moon et al. (2007).

4 Simulation Results

To see how well the asymptotic local power functions of the tests derived in the previous section approximate their small sample ones, this section presents the results of a Monte Carlo study based on 5000 iterations. For each iteration, we calculate the size of the tests at 5% level (i.e., for $c = 0$) and their power (i.e., for $c = 1$), assuming that error terms u_{it} follow either the MA process (12) or the AR(1) process

$$u_{it} = \rho u_{it-1} + v_{it}, \text{ for all } i, \tag{32}$$

with $v_{it} \sim NIID(0, \sigma_u^2)$ and $|\rho| < 1$. This is done for $N \in \{50, 100, 200, 300, 1000\}$, $T = 12$, $\theta \in \{-0.9, -0.5, 0.0, 0.5, 0.9\}$, $\rho \in \{-0.4, -0.2, 0.0, 0.2, 0.4\}$ and $p \in \{0, 1, 7\}$. The order of serial correlation p is zero when $\theta = \rho = 0$, otherwise it is set to $p = 1$ for $\theta \neq 0$ and $p = 7$ for $\rho \neq 0$.[14] The choices of N and T correspond to a range of datasets that can be found in the literature, see e.g. Baltagi et al. (2007), Canarella et al. (2013) and Nagayasou and Inakura (2009).

The nuisance parameters of models $M1$ and $M2$ that do not appear in the above local power functions are set to zero, i.e., $\alpha_i = 0$, $\beta_i = 0$, $y_{i0} = 0$, for all i. Setting the individual effects equal to zero does not result in loss of generality in all cases, except from the WGT. The β_i affect the WGT by entering in the denominator of k_{WGT} through the variance. Therefore, the higher the variance of β_i the greater the denominator of k_{WGT} and, hence, the lesser the power of the test. However, we do not provide simulations for β_i different than zero, here, as their impact on power is minimal (see also Karavias and Tzavalis (2014a)).

[14] For AR(p) models, more available moments are needed. For $\rho = 0.4$, we choose $p = 7$ because the 8th order autocorrelation is sufficiently small, given by 0.00078. In time series analysis similar decisions are made when Newey-West standard errors are used or when ADF type regressions are applied (see e.g. Said and Dickey (1984)).

Additional simulations (not reported here) have shown that the results are similar for non-zero values of these parameters even for N as small as 50.

Tables 4 and 5 present the findings of our simulation study. Table 4 presents the results for the test statistics based on model $M1$, while Table 5 presents those for the test statistics based on model $M2$. In the tables, "TV" denotes the theoretical values of the power of the tests obtained from their asymptotic power functions derived in the previous section. The results of Table 4 clearly indicate that, for model $M1$, the IV test has higher power than the WG test independently of T, as is predicted by the theory. For MA errors, when $\theta \geq 0$, the asymptotic power function of the test approximates sufficiently its small sample value even for small N, i.e., $N = \{50\}$. However, for $\theta < 0$, the power of the test considerably reduces. As is predicted by the theory (see Table 1), the WG test tends to have power only for $\theta \geq 0$. Note that, for $\theta \in \{-0.9, -0.5\}$, this test loses its power and becomes biased. The local power of both tests is an increasing function of ρ. In this case the WG test is never biased. Finally, note that both the IV and WG test statistics have size which is close the nominal level value 5%. The size performance of both tests improves, as N and T increases. A very interesting founding is that for negative MA terms size is excellent. This is in contrast to the single time series literature (see Schwert (1989)). The reason for this is that serial correlation does not affect the null hypothesis as was discussed earlier.

Regarding the test statistics for model $M2$, the results of Table 5 indicate that the IV based test statistic, denoted as FDIV, no longer performs satisfactorily. Its power deviates substantially from that predicted by its asymptotic local power function. This is true independently of the values of θ, T and N considered in our simulation analysis. This result can be attributed to the poor approximation of the asymptotic local power function in small samples, due to first differencing and the presence of more complicated deterministic terms (see also Moon et al. (2007) and Han and Phillips (2010)).

In contrast to the FDIV test, the WGT test is found to have some power in small samples. As is predicted by the theory, the test has power if $\theta < 0$ or $\rho < 0$. As N increases, the power of the WGT test converges to its asymptotic local power value from below. The table also indicates that the WGT test can have power in samples of small N even if $\theta \geq 0$ or $\theta \geq 0$, where their asymptotic local power indicates that should be biased, or have trivial power.

5 Conclusions

This paper examined the power properties of fixed-T panel data unit root tests under serial correlation and incidental trends, assuming that only the cross-section dimension of the panel (N) grows large. The analysis is based on two types of tests which have been proposed in the literature and which can accommodate both data generating process characteristics; the

WG tests of Kruiniger and Tzavalis (2002) and the IV tests of De Wachter et al. (2007). The latter were extended here to accommodate incidental trends. Analytic forms of the power functions were derived for general short memory error structures.

The results given by the paper lead to the following main conclusions. First, for the panel data model without incidental trends, the IV based test clearly outperforms the WG based test. This can be attributed to the fact that the last test requires an adjustment of the WG estimator for its inconsistency, due to the individual effects and the presence of serial correlation in the error terms. The power of the IV based test is higher under positive correlation of the error terms than under negative, and it is decreasing as the order of serial correlation increases.

Second, for the model with incidental trends, the FDIV model is found to have non-trivial power in a $N^{-1/2}$ neighbourhood of unity while the WG based test is found to have non-trivial power only in the presence of serial correlation. The latter has always power when the serial correlation in the error term is negative. This non-trivial power can be attributed to the impact of the inconsistency correction, required by the WG estimator, for the serial correlation nuisance parameters. For panel data models with incidental trends, the IV based test lacks power in small samples, despite its very good asymptotic properties. This is true independently of the sign of serial correlation of the error terms. The asymptotic local power of this test is found to be a very bad approximation of its true power. These results suggest employing the above WG based fixed-T panel unit root tests in mitigating the incidental trends problem in short panels with serially correlated error terms.

6 Supporting Information

The proofs of all the theorems, the propositions and the expressions of Appendix A have been relegated to Appendix B which appears in the online Supporting Information material.

References

[1] Arellano, M., Bover, O., 1995. Another look at the instrumental variable estimation of error-components models. Journal of Econometrics, 68, 29-51.

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

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

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

[5] Breitung J., Meyer W., 1994. Testing for unit roots using panel data: are wages on different bargaining levels cointegrated? Applied Economics 26, 353-361.

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

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

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

[9] Elliott, G. T. Rothenberg, and J. Stock (1996): Efficient Tests for an Autoregressive Unit Root. Econometrica 64, 813–836.

[10] 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.

[11] Han C., & Phillips, Peter C. B., 2010. GMM Estimation For Dynamic Panels With Fixed Effects And Strong Instruments At Unity. Econometric Theory, Cambridge University Press, vol. 26(01), pages 119-151.

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

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

[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] Im, K.S., M.H. Pesaran, and Y. Shin, 2003. Testing for unit roots in heterogeneous panels. Journal of Econometrics 115, 53–74.

[16] Karavias, Y., and Tzavalis, E., 2014a. A fixed-T Version of Breitung's Panel Data Unit Root Test. Economics Letters 124 (1) 83-87.

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

[18] 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.

[19] 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.

[20] Levin, A., Lin, F., and Chu, C., 2002. Unit root tests in panel data: asymptotic and finite-sample properties. Journal of Econometrics 122, 81–126.

[21] Maddala, G.S., Wu, S., 1999. A comparative study of unit root tests with panel data and a new simple test. Oxford Bulletin of Economics and Statistics 61, 631–651.

[22] 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.

[23] Moon, H.R., Perron, B., 2004. Testing for a unit root in panels with dynamic factors. Journal of Econometrics 122, 81–126.

[24] 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.

[25] 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.

[26] Nagayashou, J., Inakura, N., 2009. PPP: Further evidence from Japanese regional data. International Review of Economics and Finance. 18, 419-427.

[27] Phillips, P.C.B., 1987. Time series regression with a unit root. Econometrica, 55, 277-301.

[28] Quah, D., 1992. International Patterns of Growth: II. Persistence, path dependence and sustained take-off in growth transistion. Mimeo.

[29] Said, S.E., Dickey, D., 1984. Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika. 71(3), 599-607.

[30] Schott J.R. 1996. Matrix Analysis for Statistics, Wiley-Interscience.

[31] Schmidt P., and Phillips P.C.B., 1992. Testing for a unit root in the presence of deterministic trends. Oxford Bulletin for Economics and Statistics 54(3), 257-287.

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

[33] Westerlund, J., Larsson, R., 2013. New tools for understanding the local asymptotic power of panel unit root tests. Mimeo.

[34] Westerlund, J., Breitung, J., 2013. Lessons from a Decade of IPS and LLC. Econometric Reviews, 32(5-6), 547-591.

7 Appendix A

In this Appendix, we provide analytical expressions of the slope parameters for MA(1) error processes for the IV, WG, FDIV and WGT tests. Figures 1 and 2 are derived using these results. The proofs of these functions appear in Appendix B. They are based on new trace identities of frequently used matrices which may be useful for the derivation of analytical results in dynamic panel data models.

Local Power for IV test when errors are MA(1): If error terms u_{it} follow MA(1) process (12), and Assumptions 1 and 2 hold, then the slope parameter $k_{IV}(p,\theta)$ is given as

$$k_{IV}(0,0) = \sqrt{\frac{1}{2}(T^2 - T)} \tag{33}$$

$$\text{and} \quad k_{IV}(1,\theta) = \frac{D_{1,IV}\theta^2 + D_{2,IV}\theta + D_{1,IV}}{\sqrt{R_{1,IV}\theta^4 + R_{2,IV}\theta^3 + R_{3,IV}\theta^2 + R_{2,IV}\theta + R_{1,IV}}}, \tag{34}$$

where $D_{i,IV}$ and $R_{j,IV}$, for $i = 1, 2$ and $j = 1, 2, 3$, are functions of T given in Appendix B. Closed form solutions of $k_{IV}(2,0)$ and $k_{IV}(3,0)$ are also provided there.

Local Power for WG test when errors are MA(1): If error terms u_{it} follow the MA(1) process in (12), and Assumptions 1 and 2 hold, then slope parameter $k_{IV}(p,\theta)$ is given as

$$k_{WG}(0,0) = \frac{\sqrt{3}(T-1)}{\sqrt{T^2 - 2T - \frac{4}{T} + 5}}, \text{ for } p = 0 \text{ and } \theta = 0, \tag{35}$$

$$\text{and} \quad k_{WG}(1,\theta) = \frac{(T-2)(T\theta^2 - \theta^2 + 3T\theta - 7\theta + T - 1)}{2T\sqrt{R_{1,WG}\theta^4 + R_{2,WG}\theta^3 + R_{3,WG}\theta^2 + R_{2,WG}\theta + R_{1,WG}}}, \tag{36}$$

where $R_{1,WG}$, $R_{2,WG}$ and $R_{3,WG}$ are functions of T defined in Appendix B. Analytic formulas of $k_{WG}(p,\theta)$, for $p = 1, 2, 3$ and $\theta = 0$ can also be found there.

Local Power for FDIV test when errors are MA(1): If error terms u_{it} follow MA(1) process (12), and Assumptions 1 and 2 hold, then slope parameter $k_{FDIV}(p, \theta)$ is given as

$$k_{FDIV}(p,0) = \frac{T-p-3}{\sqrt{2(T-p-2)}} \tag{37}$$

$$\text{and } k_{FDIV}(1,\theta) = \frac{(T-4)\theta^2 - \theta + T - 4}{\sqrt{2(P_1\theta^4 + P_2\theta^3 + P_3\theta^2 + P_2\theta + P_1)}}, \tag{38}$$

where polynomials P_1, P_2, and P_3 are defined Appendix B.

Local Power for WGT test when errors are MA(1): If error terms u_{it} follow MA(1) process (12), and Assumptions 1 and 2 hold, then, the values of slope parameter $k_{WGT}(p, \theta)$ are given as

$$k_{WGT}(p,0) = 0, \text{ for } p = 0, 1, 2, ..., T-2, \tag{39}$$

$$\text{and } k_{WGT}(1,\theta) \neq 0, \text{ for } \theta \neq 0. \tag{40}$$

Figure 1: IV slope behavior in the presence of serial correlation

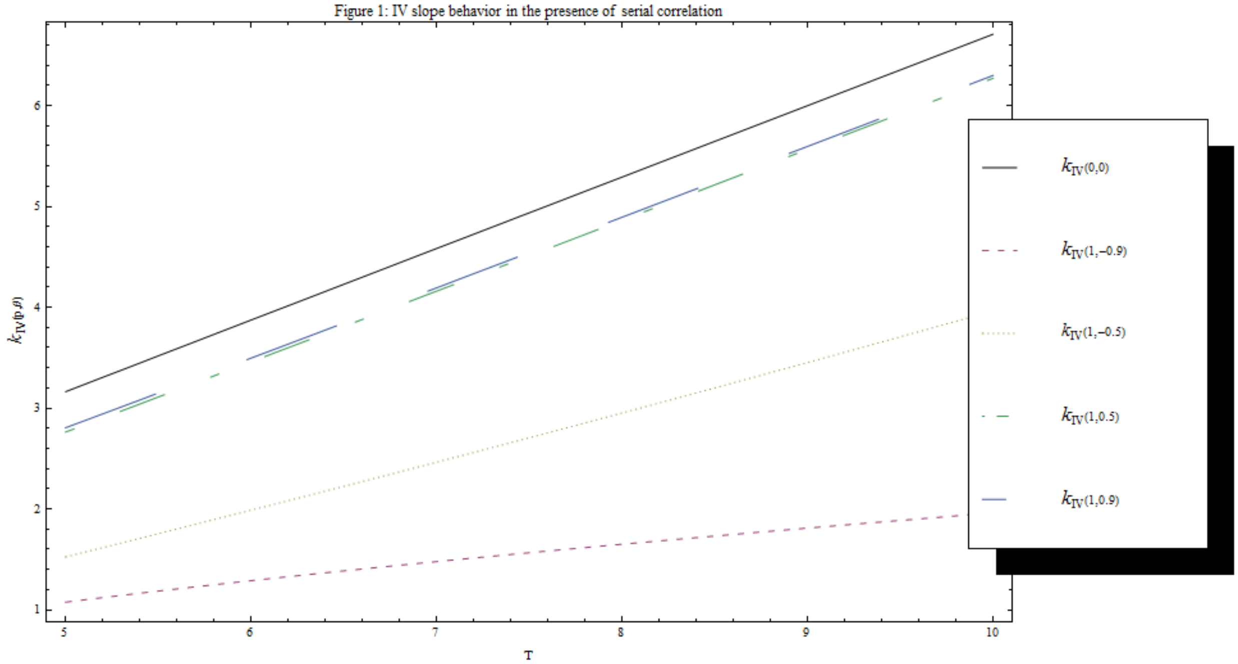

Figure 2: WG slope behavior in the presence of serial correlation

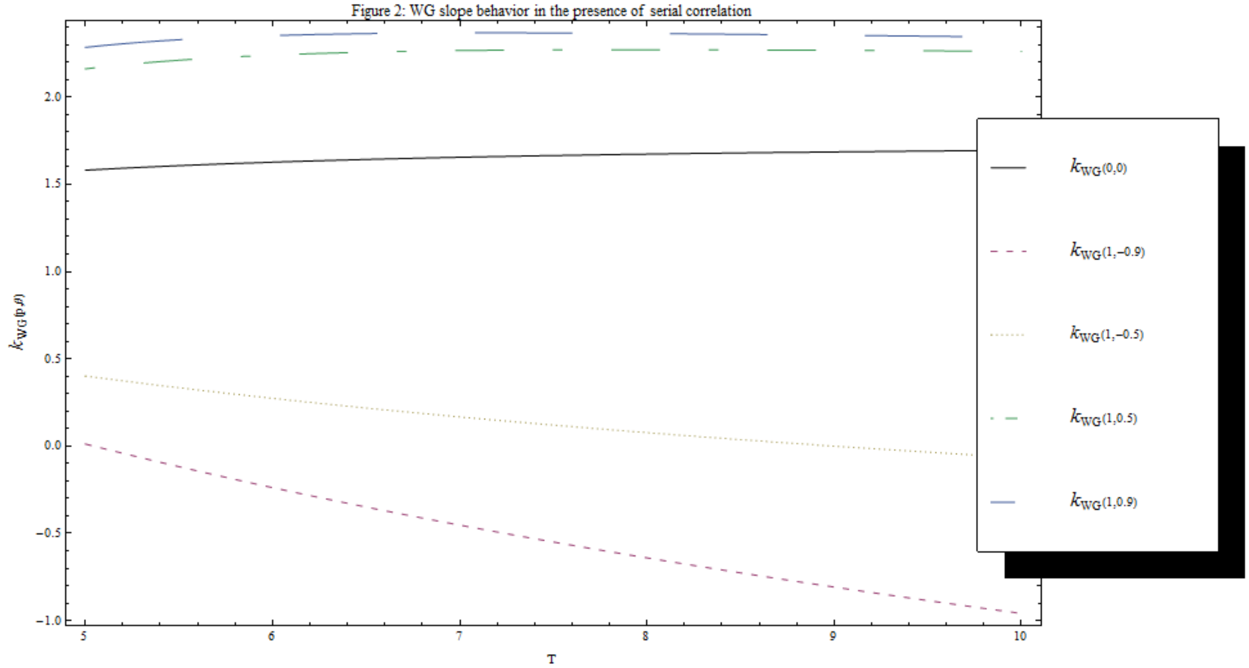

Table 1: Values of slope parameter k for model $M1$

MA(1), $T = 12$, $p = 1$					
θ	-0.9	-0.5	0.0	0.5	0.9
k_{IV}	2.2394	5.0224	7.4162	7.6823	7.7087
k_{WG}	-1.2256	-0.1812	1.7014	2.2465	2.3209
AR(1) $T = 12$, $p = 7$					
ρ	-0.4	-0.2	0.0	0.2	0.4
k_{IV}	2.5402	2.8971	3.1623	3.4380	3.8417
k_{WG}	0.7593	1.1600	1.5811	2.0659	2.7074

Table 2: Slopes of large-T tests.

IV	MPP	LLC/HT	SGLS	IPS	WG
$1/\sqrt{2}$	$1/\sqrt{2}$	$(3/2)\sqrt{(5/51)}$	$1/\sqrt{3}$	0.282	0.0

Table 3: Values of slope parameter k for model $M2$

MA(1), $T = 12$, $p = 1$					
θ	-0.9	-0.5	0.0	0.5	0.9
k_{FDIV}	1.3218	1.4078	1.8856	2.4033	2.4335
k_{WGT}	1.0883	0.6673	0	-0.2053	-0.2349
AR(1) $T = 12$, $p = 7$					
ρ	-0.4	-0.2	0.0	0.2	0.4
k_{FDIV}	0.7421	0.7834	0.8165	0.8388	0.8572
k_{WGT}	0.3793	0.1938	0	-0.2104	-0.4554

Table 4: Size and power of the IV and WG tests for model $M1$.

T=12		MA(1)						AR(1)					
N		50	100	200	300	1000	TV	50	100	200	300	1000	TV
		$\theta = -0.9$						$\rho = -0.4$					
c=0	IV	0.046	0.043	0.043	0.045	0.049	0.050	0.048	0.051	0.045	0.050	0.043	0.050
	WG	0.048	0.051	0.053	0.054	0.052	0.050	0.048	0.049	0.047	0.052	0.050	0.050
c=1	IV	0.432	0.759	0.931	0.972	0.997	0.723	0.159	0.263	0.371	0.433	0.597	0.814
	WG	0.013	0.005	0.004	0.005	0.003	0.002	0.060	0.077	0.093	0.106	0.136	0.187
		$\theta = -0.5$						$\rho = -0.5$					
c=0	IV	0.067	0.057	0.051	0.048	0.052	0.050	0.049	0.054	0.058	0.046	0.050	0.050
	WG	0.052	0.048	0.053	0.053	0.050	0.050	0.042	0.045	0.050	0.051	0.047	0.050
c=1	IV	0.935	0.989	0.997	0.999	1	0.999	0.208	0.327	0.495	0.563	0.728	0.894
	WG	0.024	0.021	0.020	0.018	0.024	0.033	0.076	0.093	0.142	0.154	0.21	0.313
		$\theta = 0$						$\rho = 0$					
c=0	IV	0.062	0.061	0.050	0.051	0.050	0.050	0.048	0.045	0.050	0.053	0.051	0.050
	WG	0.050	0.047	0.050	0.046	0.051	0.050	0.046	0.050	0.053	0.049	0.057	0.050
c=1	IV	0.998	1	0.999	0.999	1	1	0.236	0.398	0.568	0.641	0.808	0.935
	WG	0.083	0.100	0.154	0.191	0.298	0.522	0.099	0.14	0.191	0.231	0.316	0.474
		$\theta = 0.5$						$\rho = 0.2$					
c=0	IV	0.056	0.061	0.055	0.055	0.057	0.050	0.053	0.050	0.045	0.048	0.040	0.050
	WG	0.055	0.046	0.049	0.053	0.053	0.050	0.046	0.049	0.05	0.041	0.047	0.050
c=1	IV	0.999	1	1	1	1	1	0.289	0.479	0.649	0.738	0.864	0.963
	WG	0.136	0.190	0.264	0.338	0.474	0.726	0.127	0.195	0.279	0.328	0.462	0.663
		$\theta = 0.9$						$\rho = 0.4$					
c=0	IV	0.063	0.057	0.057	0.054	0.051	0.050	0.04	0.047	0.051	0.045	0.050	0.050
	WG	0.047	0.053	0.054	0.046	0.056	0.050	0.048	0.050	0.042	0.052	0.041	0.050
c=1	IV	0.999	1	1	1	1	1	0.385	0.595	0.77	0.833	0.935	0.986
	WG	0.142	0.201	0.288	0.347	0.499	0.750	0.181	0.302	0.438	0.511	0.661	0.856

Table 5: Size and local power of FDIV and WGT tests for model $M2$.

T=12		MA(1)						AR(1)					
N		50	100	200	300	1000	TV	50	100	200	300	1000	TV
		$\theta=-0.9$						$\rho=-0.4$					
c=0	FDIV	0.043	0.048	0.048	0.051	0.045	0.050	0.042	0.042	0.054	0.052	0.054	0.050
	WGT	0.054	0.053	0.053	0.051	0.050	0.050	0.053	0.051	0.052	0.049	0.055	0.050
c=1	FDIV	0.058	0.059	0.063	0.056	0.059	0.373	0.049	0.053	0.045	0.049	0.055	0.183
	WGT	0.108	0.145	0.169	0.194	0.234	0.288	0.061	0.077	0.084	0.087	0.086	0.102
		$\theta=-0.5$						$\rho=-0.2$					
c=0	FDIV	0.042	0.050	0.049	0.046	0.049	0.050	0.049	0.040	0.049	0.049	0.049	0.050
	WGT	0.052	0.045	0.050	0.053	0.048	0.050	0.049	0.048	0.055	0.054	0.050	0.050
c=1	FDIV	0.049	0.054	0.053	0.049	0.057	0.406	0.049	0.050	0.051	0.046	0.054	0.194
	WGT	0.138	0.139	0.160	0.164	0.165	0.164	0.064	0.070	0.073	0.077	0.081	0.073
		$\theta=0$						$\rho=0$					
c=0	FDIV	0.045	0.045	0.043	0.048	0.050	0.050	0.042	0.044	0.048	0.048	0.047	0.050
	WGT	0.069	0.057	0.058	0.054	0.053	0.050	0.053	0.049	0.048	0.051	0.053	0.050
c=1	FDIV	0.039	0.048	0.047	0.047	0.042	0.595	0.043	0.046	0.050	0.049	0.044	0.203
	WGT	0.169	0.168	0.136	0.134	0.089	0.050	0.061	0.059	0.068	0.064	0.057	0.050
		$\theta=0.5$						$\rho=0.2$					
c=0	FDIV	0.038	0.042	0.046	0.045	0.047	0.050	0.046	0.044	0.044	0.052	0.048	0.050
	WGT	0.059	0.055	0.055	0.056	0.060	0.050	0.052	0.053	0.047	0.047	0.053	0.050
c=1	FDIV	0.020	0.036	0.031	0.035	0.044	0.775	0.045	0.051	0.052	0.049	0.051	0.210
	WGT	0.200	0.162	0.125	0.101	0.073	0.032	0.059	0.057	0.049	0.039	0.046	0.031
		$\theta=0.9$						$\rho=0.4$					
c=0	FDIV	0.034	0.040	0.046	0.044	0.048	0.050	0.047	0.053	0.050	0.046	0.051	0.050
	WGT	0.063	0.058	0.056	0.054	0.058	0.050	0.050	0.050	0.050	0.050	0.054	0.050
c=1	FDIV	0.023	0.029	0.034	0.035	0.039	0.784	0.045	0.041	0.044	0.047	0.048	0.215
	WGT	0.200	0.160	0.124	0.102	0.064	0.030	0.049	0.047	0.039	0.036	0.026	0.017