Granger Centre
for time series econometrics

Granger Centre Discussion Paper Series

The local power of fixed-*T* panel unit root tests allowing for serially correlated error terms

by

Yiannis Karavias and Elias Tzavalis

Granger Centre Discussion Paper No. 12/01

The University of Nottingham

The local power of fixed-T panel unit root tests allowing for serially correlated error terms

Yiannis Karavias[a] and Elias Tzavalis[b]

a :School of Economics, University of Nottingham

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

Abstract

The asymptotic power properties of fixed-T panel unit root tests allowing for serially correlated error terms are examined by deriving their asymptotic local power functions. This is done for dynamic panel data models allowing for individual effects or individual effects and incidental trends. For the first model, the paper shows that an instrumental variables (IV) based test statistic, which exploits orthogonal moment conditions of the demeaned by their initial observations individual series of the panel, performs better than least squares (LS) tests based on the "within group" transformation of the series. Allowing for serial correlation reduces the power of the IV based test. This reduction however is unimportant in the case of positive serial correlation of the error terms. For the panel data model with incidental trends, the paper shows that LS based test statistics relying on "within group" or forward deviations transformations of the data have non-trivial power in the natural root-N neighborhood of unity, if the errors terms are negatively correlated. This power is retained even in panels with small N. For the IV based test statistic, the asymptotic local power function constitutes a poor approximation of its true power, even in large N panels.

JEL classification: C22, C23

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

The authors would like to thank Tassos Magdalinos and Robert Taylor for their helpful comments. They also thank seminar participants of the Granger Centre for Time Series Econometrics for useful comments on a previous version of the paper.

a :Yiannis Karavias: ioannis.karavias@nottingham.ac.uk

b :Elias Tzavalis: etzavalis@aueb.gr

1 Introduction

Panel unit root test statistics assuming fixed (finite) time dimension (T) and large cross-sectional dimension (N) have received much interest in the literature over the last decade because of their very good small sample properties. Early contributions in this area include the test statistics suggested by Sargan and Bhargava (1983), Breitung and Meyer (1994), Harris and Tzavalis (1999, 2004), Kruiniger and Tzavalis (2002), Bond et al. (2005), De Wachter et al. (2007), Kruiniger (2008), Han and Phillips (2010) and De Blander and Dhaene (2011).

In this paper, we derive analytically the limiting distribution of fixed-T panel unit root tests allowing for serial correlation under local alternatives and, then, we study the asymptotic power properties of these tests. Despite the plethora of studies for large-T panel unit root tests,[1] there are a few studies in the literature investigating the asymptotic local power properties of fixed-T panel unit root tests (see, e.g., Bond et al. (2005) and Madsen (2010)). These studies are focused on panel data unit root tests which assume white noise error terms and consider panel data models without incidental trends. Allowing for serial correlation of the error terms, or higher order dynamics of panel data models, can affect the power performance of fixed-T panel unit root tests in small samples. The effects of serial correlation on the power of these tests have been studied through Monte Carlo simulations by De Blander and Dhaene (2011). Our paper considers fixed-T unit root tests for the panel data autoregressive model with individual effects and for that allowing also for incidental trends.

Several contributions are made by the paper. First, it is shown that, for the model with individual effects, the instrumental variables (IV) based test statistic suggested by De Wachter et al. (2007) is a very powerful test statistic. To allow for serial correlation and to remove the panel data initial conditions nuisance effects on testing for unit roots, this test statistic exploits orthogonal moments of panel data individual series demeaned by their initial observations under the null hypothesis of unit roots. It is found to have higher power than the least squares (LS) based test statistic suggested by Kruiniger and Tzavalis (2002).[2] The latter relies on the "within group" transformation matrix to become invariant to initial conditions. For large T, scaled appropriately with T the IV based test statistic reaches its maximum power, which equals that of the common-point optimal test of Moon et al. (2007).

Second, allowing for serial correlation has a different impact on the power of each of the fixed-T panel unit root tests examined. This was expected, since the moments used in estimation and testing procedures under serial correlation of the error terms are exploited differently in every test. The power loss of the tests is more severe when the degree of serial

[1] See, e.g., Moon and Phillips (1999), Breitung (2000), Moon and Perron (2004), Moon et al. (2007), Moon and Perron (2008), Harris et al. (2010).

[2] A similar statistic is also presented by Moon and Perron (2004) for large-T panels.

correlation is large and negative. In this case, the "within group" LS based test statistic becomes biased. In the case of positive serial correlation, the power reductions of the IV based test statistic are unimportant, while the "within group" LS based test displays power gains.

Third, fixed-T panel unit root tests suffer from the "incidental trends problem", as their corresponding large-T tests. However, this problem appears in the case of no serially correlated error terms. In this case, the asymptotic power of the LS based test statistics relying on either the "within group" transformation of the individual series of the panel or on their forward orthogonal deviations transformation, suggested by Breitung (2000), have both trivial power. However, under negative serial correlation of the error terms, both tests above have non-trivial power. This power is retained in small samples even under positive serial correlation. The IV based test statistic, which relies on a first-difference transformation of the data to avoid estimating incidental trends, is found to have asymptotic local power even in the case of no serially correlated error terms. But, as shown by Monte Carlo simulations, the asymptotic local power function constitutes a very bad approximation of the true power of this test.

The paper is organized as follows. Section 2 introduces the fixed-T test statistics 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 behavior of the tests. Section 4 conducts a small Monte Carlo exercise to examine the small sample performance of the asymptotic results and Section 6 concludes the paper. All proofs are relegated to the Appendix. 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:

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

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

where $y_i = (y_{i1}, ..., y_{iT})'$ and $y_i = (y_{i0}, ..., y_{iT-1})'$ are $(TX1)$ vectors, u_i is the $(TX1)$ vector of error terms u_{it}, and a_i and β_i are the individual coefficients of the deterministic components of the models. a_i coefficients reflect individual effects of the panel, while β_i capture the slopes of individual linear trends, referred to as incidental trends. The $(TX1)$ vector e has elements $e_t = 1$, for $t = 1...T$, and $\tau_t = t$ is the time trend.

To study the asymptotic local power of fixed-T unit root tests, define the autoregressive coefficient φ as $\varphi_N = 1 - \frac{c}{\sqrt{N}}$. 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: (1.a)$\{u_i\}$ constitutes a sequence of independent normal random vectors of dimension $(TX1)$ 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 $p_{\max} \leq T-2$. (1.b) $\gamma_{tt} > 0$ for at least one $t = 1, ..., T$. (1.c) The $4+\delta-th$ population moments of Δy_i, $i = 1, ..., N$ are uniformly bounded. That is, for every $l \in R^T$ such that $l'l = 1$, $E(|l'\Delta y_i|^{4+\delta}) < B < +\infty$ for some B, where Δ is the difference operator. (1.d) $l'Var(vec(\Delta y_i \Delta y_i')l > 0$ for every $l \in R^{0.5T(T+1)}$ such that $l'l = 1$.

Assumption 2: The individual coefficients a_i and β_i, and the initial observations of models $M1$ and $M2$, y_{i0}, satisfy the following conditions: $E(u_{it}a_i) = 0$, $E(u_{it}\beta_i) = 0$ and $E(u_{it}y_{i0}) = 0$, for $t = 1, ..., T$ and $i = 1, ..., N$, and $Var(y_{i0}) < +\infty$.

Assumption (1.a) 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, $\gamma_{1T} = \gamma_{T1} = 0$. This assumption can be strengthened to allow for a smaller order of serial correlation. If p is the order of serial correlation assumed by the researcher and p^* the true order, then the limiting distribution of φ_N is valid as long as $p \geq p^*$. Choosing $p > p^*$ means selecting fewer than possible moments for inference. For a discussion, on how to estimate the true order of serial correlation, p^*, see Hayakawa (2010). Assuming normality in the error terms allows for closed form representations of the variances of the limiting distributions of the tests.

Assumption (1.b) imposes finite fourth moments on initial conditions y_{i0}, error terms u_{it} and individual coefficients a_i and β_i of models $M1$ and $M2$. Along with assumptions (1.c) and (1.d), they allow application of the Markov LLN and the Lindeberg -Levy CLT, and ensure that all quantities in the denominators of the estimators of φ_N are non-zero.

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, FOD and $FDIV$ statistics, in next section),[3] or by the "within group"

[3] This approach is suggested by Schmidt and Phillips (1992), for single time series, and Breitung and

transformation of y_{it} (see WG and WGT statistics).[4] 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 stationary of y_{i0}, implying $Var(y_{i0}) = \frac{\sigma^2}{1-\varphi_N^2}$ (see Kruiniger (2008) and Madsen (2010)) is not considered. This is because, as is also noted by Moon et al. (2007), this assumption implies that $Var(y_{i0}) \to \infty$ when $\varphi_N \to 1$, which means that the variance of the initial condition increases with the number of cross-section units, which is not meaningful for cross-section data sets.

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 $M2$.

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 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:

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

Meyer (1994) for the individual series of panel data models with individual effects.

[4] This transformation means that one subtracts the means of the individual series of the panel from their levels, across all units. This transformation is also made by Dickey and Fuller (1979) in their unit root test, for single time series. It is also employed by the panel unit root tests of Harris and Tzavalis (1999), and Levin et al. (2002).

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 (TXT) matrix selecting zero-mean moments, according to (5), and $z_{i-1} = y_{i-1} - y_{i0}e$. In particular, Π_p has ones in the *pth* diagonal and zeros everywhere else. Given the definition of Π, the above IV estimator can be rewritten as

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

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

Theorem 1 *Under Assumptions 1, 2 and the assumption that the order of serial correlation is at most p, we have*

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

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

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

and $V_{IV} = \frac{2tr((A_{IV}\Gamma)^2)}{tr(\Lambda' \Pi_p \Lambda \Gamma)^2}$, *with* $A_{IV} = \frac{1}{2}(\Lambda' \Pi_p + \Pi'_p \Lambda)$, *is the variance of the limiting distribution of* $\hat{\varphi}_{IV}$. *The definition of matrix* Λ *is given in the appendix (see proof of the theorem).*

The limiting distribution of the IV test statistic given by Theorem 1 nests the distributions of it 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 $p = 0$.[5] The only unknown quantity in the variance is Γ, which is required for the estimation of the variance of the limiting distribution of $\hat{\varphi}_{IV}$, V_{IV}. If $\Gamma = \sigma_u^2 I_T$, where I_T is the (TXT) identity matrix, then no estimation of Γ is needed since σ_u^2 is cancel out from both the nominator and denominator of $\hat{\varphi}_{IV}$. 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

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

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

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

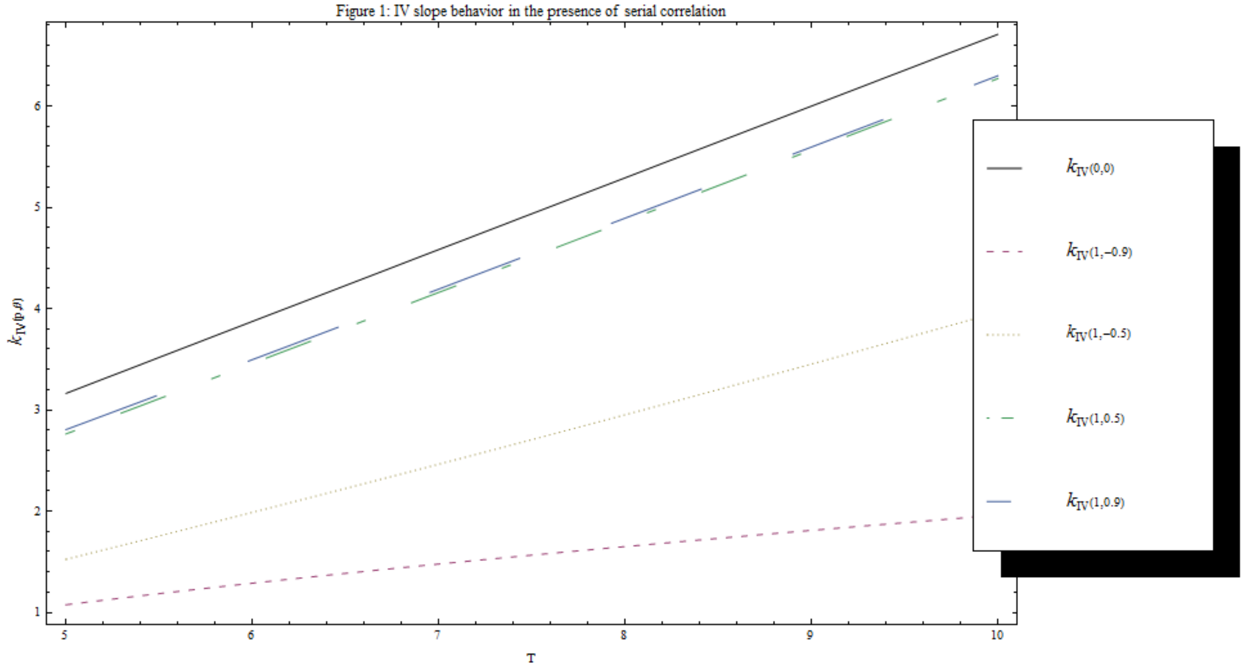

Figure 1: IV slope behavior in the presence of serial correlation

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 assumed order of serial correlation p and the form of serial correlation considered by variance-covariance matrix Γ. In the case where error terms u_{it} follow MA(1) process

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

with $v_{it} \sim NIID(0, \sigma_u^2)$, then an closed form of k_{IV}, defined as $k_{IV}(p, \theta)$ for different values of p and θ, is given in the next corollary.

Corollary 1 *If error terms u_{it} follow MA(1) process (12), and Assumptions 1 and 2 hold, then slope parameter $k_{IV}(p, \theta)$ is given as*

$$k_{IV}(0,0) = \sqrt{\frac{1}{2}(T^2 - T)} \tag{13}$$

$$\text{and} \quad k_{IV}(1,\theta) = \frac{D_{1,IV}\theta^2 + D_{2,IV}\theta + D_{1,IV}}{\sqrt{R_{1,IV}\theta^4 + R_{2,IV}\theta^3 + R_{3,IV}\theta^2 + R_{2,IV}\theta + R_{1,IV}}} \tag{14}$$

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

The results of Corollary 1 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 $p=0$ 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 group" transformation of the individual series y_{it}, using the annihilator matrix $Q=I_T-e(e'e)^{-1}e'$, where I_T is the (TXT) identity matrix. Then, the least squares estimator of the transformed series is given as

$$\hat{\varphi}_{WG}=(\sum_{i=1}^{N}y'_{i-1}Qy_{i-1})^{-1}(\sum_{i=1}^{N}y'_{i-1}Qy_i). \tag{15}$$

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:

$$\sqrt{N}\hat{\delta}_{WG}\left(\hat{\varphi}_{WG}-1-\frac{\hat{b}_{WG}}{\hat{\delta}_{WG}}\right)\xrightarrow{d}N(0,V_{WG}), \tag{16}$$

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

which corrects estimator $\hat{\varphi}_{WG}$ for the above two sources of its inconsistency, where $\hat{\delta}_{WG}=\frac{1}{N}\sum_{i=1}^{N}y'_{i-1}Qy_{i-1}$ is the denominator of estimator $\hat{\varphi}_{WG}$ scaled by N, $\frac{\hat{b}_{WG}}{\hat{\delta}_{WG}}=\frac{tr(\Psi_{p,WG}\hat{\Gamma})}{\hat{\delta}_{WG}}$ is a consistent estimator of the inconsistency of $\hat{\varphi}_{WG}$, given as $\frac{tr(\Lambda'Q\Gamma)}{tr(\Lambda'Q\Lambda\Gamma)}$, and $\Psi_{p,WG}$ is a (TXT)-dimension selection matrix having in its $-p,..,0,...p$ diagonals the corresponding elements of matrix $\Lambda'Q$, and zero everywhere else. $\hat{\Gamma}=\frac{1}{N}\sum_{i=1}^{N}\Delta y_i\Delta y_i'$ and $V_{WG}=2tr((A_{WG}\Gamma)^2)$ is the variance of the limiting distribution of the corrected for its inconsistency LS estimator $\hat{\varphi}_{WG}$, where $A_{WG} = \frac{1}{2}(\Lambda' Q + Q\Lambda - \Psi_{p,WG} - \Psi'_{p,WG})$.[6] This variance can be consistently estimated provided consistent estimates of Γ. As for the IV test statistic, this can be done based on (11).

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 $\hat{\varphi}_{WG} - 1 - \frac{\hat{b}_{WG}}{\hat{\delta}_{WG}}$ through the selection matrix $\Psi_{p,WG}$.[7] Moon and Perron (2004) have suggested a version of the WG test statistic for the case that both N and T go infinity. The next theorem gives the limiting distribution of the WG statistic under the sequence of local alternatives $\varphi_N = 1 - \frac{c}{\sqrt{N}}$.

Theorem 2 *Under Assumptions 1, 2 and the assumption that the order of serial correlation is at most p, we have*

$$\sqrt{N}\ V_{WG}^{-1/2}\hat{\delta}_{WG}\left(\hat{\varphi}_{WG} - 1 - \frac{\hat{b}_{WG}}{\hat{\delta}_{WG}}\right) \xrightarrow{d} N(-ck_{WG}, 1), \tag{17}$$

as $N \to \infty$, where

$$k_{WG} = \frac{tr(\Lambda' Q\Lambda\Gamma) + tr(F'Q\Gamma) - tr(\Psi_{p,WG}\Lambda\Gamma) - tr(\Lambda'\Psi_{p,WG}\Gamma)}{\sqrt{V_{WG}}} \tag{18}$$

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

The results of Theorem 2 indicate that annihilator matrix Q and the inconsistency correction of estimator $\hat{\varphi}_{WG}$, $\frac{\hat{b}_{WG}}{\hat{\delta}_{WG}}$, based on $\Psi_{p,WG}$, makes more complex the local power function. As equation (18) 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

[6] Note that the WG test statistic, given by 16, 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

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

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.

[7] 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 $\Psi_{1,WG} = \begin{pmatrix} -\frac{2}{3} & -\frac{1}{3} & 0 \\ \frac{1}{3} & -\frac{1}{3} & 0 \\ 0 & \frac{2}{3} & 0 \end{pmatrix}$.

come from the annihilator matrix Q and the last two from selection matrix $\Psi_{p,WG}$. For $p = 0$, the effects of matrix $\Psi_{p,WG}$ disappear, since $tr(\Psi_{p,WG}\Lambda\Gamma) = tr(\Lambda'\Psi_{p,WG}\Gamma) = 0$. To study the effects of the serial correlation nuisance parameters and lag-order p on k_{WG}, next corollary gives analytic formulas of k_{WG}, for $p \in \{0, 1\}$ and $\theta \in \{-0.9, -0.5, 0, 0.5, 0.9\}$, while Figure 2 plots values of these formulas across T.

Corollary 2 *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*

$$k_{WG}(0,0) = \frac{\sqrt{3}(T-1)}{\sqrt{T^2 - 2T - \frac{4}{T} + 5}}, \text{ for } p = 0 \text{ and } \theta = 0, \tag{19}$$

$$\text{and} \quad k_{WG}(1,\theta) = \frac{(T-2)(T\theta^2 - \theta^2 + 3T\theta - 7\theta + T - 1)}{2T\sqrt{R_{1,WG}\theta^4 + R_{2,WG}\theta^3 + R_{3,WG}\theta^2 + R_{2,WG}\theta + R_{1,WG}}}, \tag{20}$$

where $R_{1,WG}, R_{2,WG}$ and $R_{3,WG}$ are functions of T defined in the appendix. The appendix also gives analytic formulas of $k_{WG}(p, \theta)$, for $p = 1, 2, 3$ and $\theta = 0$.

As can be seen from Figure 2, 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 group" 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 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 and 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 $\theta \in \{-0.9, -0.5, 0, 0.5, 0.9\}$.

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.

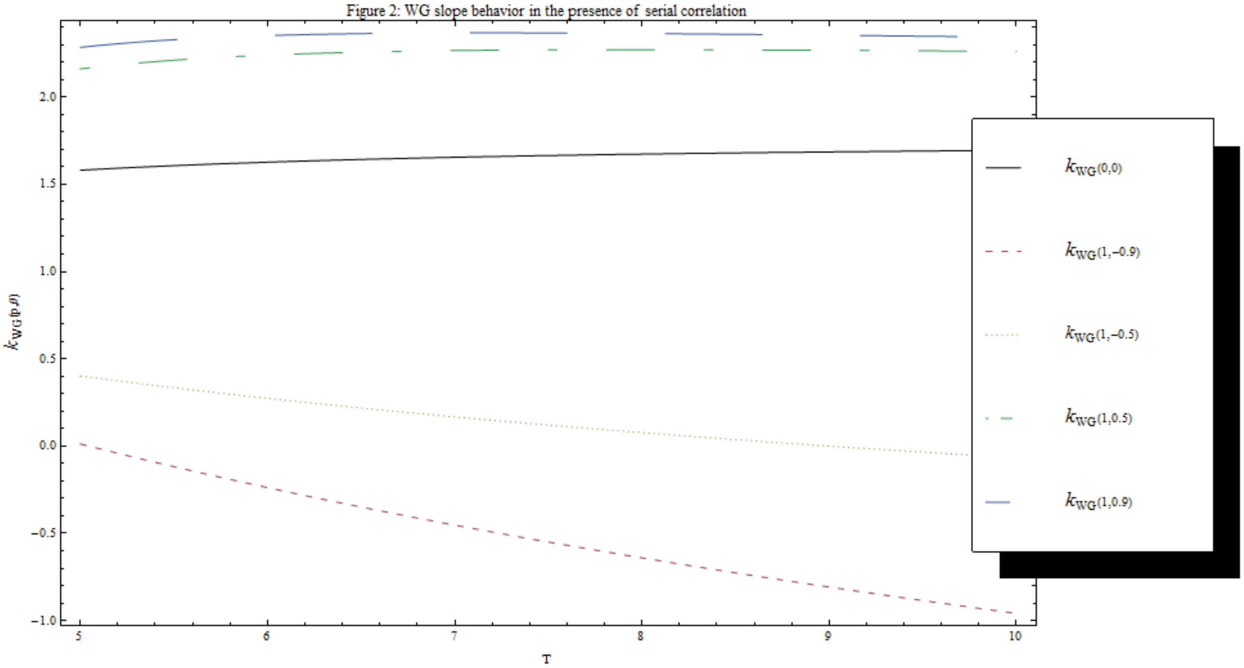

Figure 2: WG slope behavior in the presence of serial correlation

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

$$T\sqrt{N}(\sqrt{2})^{-1}(\hat{\varphi}_{IV} - 1) \xrightarrow{d} N(-c\frac{1}{\sqrt{2}}, 1), \qquad \text{and} \tag{21}$$

$$T\sqrt{N}(\sqrt{3})^{-1}\hat{\delta}_{WG}\left(\hat{\varphi}_{WG} - 1 - \frac{\hat{b}_{WG}}{\hat{\delta}_{WG}}\right) \xrightarrow{d} N(-c0, 1), \tag{22}$$

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 (21), 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).

Table 1: Values of slope parameter k

			T=7		
θ	-0.9	-0.5	0.0	0.5	0.9
k_{IV}	1.477	2.463	4.582	4.159	4.192
k_{WG}	-0.452	0.167	1.655	2.266	2.367
			T=10		
$p\backslash\theta$	-0.9	-0.5	0.0	0.5	0.9
k_{IV}	1.960	3.965	6.708	6.271	6.299
k_{WG}	-0.958	-0.067	1.694	2.261	2.343

For $c > 0$, the IV test reaches its 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 $k_{WG} = 0$. This happens because the last test adjusts only the numerator of $\hat{\varphi}_{WG}$ for its inconsistency, in contrast to Harris' and Tzavalis (1999) (denoted HT) and Levin's et al. (2002) (denoted LLC) tests. The latter tests adjust both the numerator and denominator of $\hat{\varphi}_{WG}$ for its inconsistency. Moon and Perron (2008) show that the WG test has non-trivial power in a $n^{-1/4}T$ neighborhood of the null hypothesis. Values of the slope parameter of the power function of the above tests, for large T, are reported in the following table:

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

For comparisons, the table also reports values of k for the large-T panel unit root tests of Im et al. (2003) (denoted IPS) , and Sargan's (SGLS) test statistic (see Moon and Perron (2008)). Values of k for these tests are obtained in Moon et al. (2007), Moon and Perron (2008) and Harris et al. (2010).

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, this section extends the IV test presented in the previous section and gives a fixed-T version of Breitung's (2000) test which also allows for serial correlation. As said before, the latter is based on forward orthogonal deviations transformation of individual series of the panel y_{it} to overcome the problem of estimating the incidental trends' nuisance parameters. Thus, it will be henceforth denoted as FOD. To overcome this problem, 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 $M2$ yields

$$\Delta y_i = \varphi \Delta y_{i-1} + (1-\varphi)\beta_i e^* + \Delta u_i, \quad i = 1, ..., N, \tag{23}$$

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 $(T-1)X1$ vectors. Subtracting from both sides of model (23), the vector of the first difference of initial observation $\Delta y_{i1} e$ gives the following first differences transformation of the model:

$$y_i^* = \varphi y_{i-1}^* + (1-\varphi)a_i^* + u_i^*, \quad i = 1, ..., N, \tag{24}$$

where $y_i^* = \Delta y_i - \Delta y_{i1} e$, $y_{i-1}^* = \Delta y_{i-1} - \Delta y_{i1} e$, $a_i^* = (\beta_i - \Delta y_{i1})$ and $u_i^* = \Delta u_i$. Model (24) clearly shows that, if error terms u_{it} are serially correlated, moments similar to (7) can be exploited to test the null hypothesis of a unit root, i.e.

$$E(y_{i-1}^{*\prime} \Pi_p^* u_i^*) = 0, \tag{25}$$

where Π_p^* is a $(T-1)X(T-1)$ 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

$$\hat{\Theta} = \frac{1}{N} \sum_{i=1}^{N} \Delta y_i^* \Delta y_i^{*\prime}, \tag{26}$$

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 latter on, Γ and Γ_1 can not 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 $\varphi_N = 1 - \frac{c}{\sqrt{N}}$, exploiting the above moment conditions.

Theorem 3 *Under Assumptions 1, 2 and the assumption that the order of serial correlation is at most p, we have*

$$\sqrt{N} V_{FDIV}^{-1/2} (\hat{\varphi}_{FDIV} - 1) \xrightarrow{d} N(-c k_{FDIV}, 1), \tag{27}$$

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

$$k_{FDIV} = \frac{tr(\Lambda^{*\prime} \Pi_p^* \Lambda^* \Theta)}{\sqrt{2 tr((A_{FDIV} \Theta)^2)}} \tag{28}$$

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

The results of Theorem 3 indicate that, 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. Corollary 3 derives the value of the slope parameter k_{FDIV}, if error terms u_{it} follow MA(1) process.

Corollary 3 *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*

$$k_{FDIV}(p,0) = \frac{T-p-3}{\sqrt{2(T-p-2)}} \tag{29}$$

$$\text{and } k_{FDIV}(1,\theta) = \frac{(T-4)\theta^2 - \theta + T - 4}{\sqrt{2(P_1\theta^4 + P_2\theta^3 + P_3\theta^2 + P_2\theta + P_1)}}, \tag{30}$$

where polynomials P_1, P_2, and P_3 are defined in the appendix.

Table 3 presents values of $k_{FDIV}(p,\theta)$, obtained through relationship (30), 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. However, if $T \to \infty$, it can be shown that $k^*_{FDIV} = \frac{T-p-3}{T\sqrt{2(T-p-2)}} \to 0$, which means that the incidental parameter problem remains. This is due to the normalization of the statistic with T. These results mean that the asymptotic power of the FDIV test comes from the assumption that T is fixed and 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 $M1$.

Table 3: Values of slope parameter p.

			T=7		
θ	-0.9	-0.5	0.0	0.5	0.9
k_{FDIV}	0.862	0.896	1.264	1.186	1.179
k_{WGT}	0.694	0.466	0.00	-0.212	-0.248
k_{FOD}	0.148	0.110	0.00	-0.062	-0.073
			T=10		
θ	-0.9	-0.5	0.0	0.5	0.9
k_{FDIV}	1.160	1.229	1.750	1.989	2.008
k_{WGT}	1.042	0.645	0.00	-0.216	-0.248
k_{FOD}	0.151	0.110	0.00	-0.047	-0.054

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

$$p\lim_{N\to\infty} \hat{\Gamma} = p\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\Delta y_i \Delta y_i' = \Gamma + E(\beta_i^2)ee'. \tag{31}$$

To remove the effects of β_i from the estimator of matrix Γ, the following selection matrix will be defined.[8] 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

$$p\lim_{N\to\infty} \frac{1}{tr(Mee')N}\sum_{i=1}^{N}\Delta y_i' M \Delta y_i = E(\beta_i^2). \tag{32}$$

The last relationship can be employed to substitute out individual effects $E(\beta_i^2)$ from (31), 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 (31) and (32), we can define selection matrix $\Phi_{p,WGT} = \Psi_{p,WGT} - \frac{tr(\Lambda' Q^* M)}{e'Me} M$, where $\Psi_{p,WGT}$ is a (TXT) matrix having in its 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 $tr(\Lambda' Q^* \Gamma)$:

$$p\lim_{N\to\infty} tr(\Phi_{p,WGT}\hat{\Gamma}) = tr(\Lambda' Q^* \Gamma). \tag{33}$$

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

Theorem 4 *Under Assumptions 1, 2 and the assumption that the order of serial correlation*

[8] Note that, as in case of model $M1$ (see fn 6), 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).

is at most p, we have

$$\sqrt{N}\hat{V}_{WGT}^{-\frac{1}{2}}\hat{\delta}_{WGT}\left(\hat{\varphi}_{WGT}-1-\frac{\hat{b}_{WGT}}{\hat{\delta}_{WGT}}\right)\xrightarrow{d}N(-ck_{WGT},1), \tag{34}$$

as $N\to+\infty$, where

$$k_{WGT}=\frac{tr(\Lambda'Q^{*}\Gamma)+tr(F'Q^{*}\Gamma)-tr(\Phi_{p,WGT}\Lambda\Gamma)-tr(\Lambda'\Phi_{p,WGT}\Gamma)}{2tr((A_{WGT}\Gamma)^{2})}, \tag{35}$$

$\hat{\varphi}_{WGT}=\left(\sum_{i=1}^{N}y_{i-1}'Q^{*}y_{i-1}\right)^{-1}\left(\sum_{i=1}^{N}y_{i-1}'Q^{*}y_{i}\right)$, $\frac{\hat{b}_{WGT}}{\hat{\delta}_{WGT}}=\frac{tr(\Phi_{p,WGT}\hat{\Gamma})}{\frac{1}{N}\sum_{i=1}^{N}y_{i,-1}'Q^{*}y_{i,-1}}$, *and* $V_{WGT}=2tr((A_{WGT}\Gamma)^{2})$, *with* $A_{WGT}=\frac{1}{2}(\Lambda'Q^{*}+Q^{*}\Lambda-\Phi_{p,WGT}-\Phi_{p,WGT}')$, *is the variance of the limiting distribution of the WGT test.*

The implementation of the WG test statistic is based on the estimator of Γ given by $\hat{\Gamma}$. As was made clear by our analysis above, Premultiplying $\hat{\Gamma}$ by selection matrix $\Phi_{p,WGT}$ renders this estimator net of the incidental trends nuisance parameters effects. The results of Theorem 4 imply that, if there is no serial correlation, test statistic WGT has trivial power. This is true for any order of serial correlation p. These results are established in next corollary, which derives values of the power slope parameter $k_{WGT}(p,\theta)$ under MA process (12) of u_{it}, for different values of p and θ.

Corollary 4 *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*

$$k_{WGT}(p,0) = 0, \text{ for } p=0,1,2,...,T-2, \tag{36}$$

$$\text{and} \quad k_{WGT}(1,\theta) \neq 0, \text{ for } \theta\neq 0. \tag{37}$$

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. This power can be attributed to the effects of quantities $tr(\Phi_{p,WGT}\Lambda\Gamma)$ and $tr(\Lambda'\Phi_{p,WGT}\Gamma)$ on slope parameter $k_{WGT}(p,\theta)$. As for the FDIV test, it can be shown that the large-T version of the WGT test has trivial power when $T\to\infty$.

FOD panel unit root test: This test is initially suggested by Breitung (2000) as a large-T panel unit root test. It is based on forward orthogonal deviations transformation of the individual series y_{it} of model $M2$, known as Helmert transformation, to avoid estimating incidental trend parameters β_i. As shown by Moon et al. (2006), the joint T,N asymptotic local power of the test is zero at the natural rate of $T^{-1}N^{-1/2}$. Below, we present a fixed-T version of the test and examine its asymptotic local power, as $N \to \infty$.

In a first step, the orthogonal transformation of series y_{it} requires subtracting initial observations y_{i0} from y_{it}, for all i, and taking the transformed series $z_{it} = y_{it} - y_{i0}$. Then, define the following $(T-1)XT$ matrices:

$$A = \begin{pmatrix} 0_{1XT} \\ GH \end{pmatrix} \text{ and } B = \begin{pmatrix} 0_{1X(T-2)} & 0 & 0 \\ I_{T-2} & 0_{(T-2)X1} & -\frac{1}{T}\tau_{T-2} \end{pmatrix},$$

where

$$G = \begin{pmatrix} \sqrt{\frac{T-2}{T-1}} & & & 0 \\ & \sqrt{\frac{T-3}{T-2}} & & \\ & & \ddots & \\ 0 & & & \sqrt{\frac{1}{2}} \end{pmatrix} \text{ and } H = \begin{pmatrix} 1 & -\frac{1}{T-1} & \cdots & \cdots & \cdots & \cdots & -\frac{1}{T-1} \\ 0 & 1 & -\frac{1}{T-2} & \cdots & \cdots & \cdots & -\frac{1}{T-2} \\ \vdots & & \ddots & & & & \vdots \\ \vdots & & & \ddots & & & \vdots \\ \vdots & & & & 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \cdots & \cdots & \cdots & 0 & 1 & -1 \end{pmatrix},$$

with dimensions $(T-2)X(T-1)$ and $(T-1)XT$ respectively, and vector $\tau_{T-2} = \begin{pmatrix} 1 \\ 2 \\ \vdots \\ T-2 \end{pmatrix}$.

In case of no serial correlation of error terms u_{it}, multiplying Δz_i with matrix A and z_i with matrix B implies the following orthogonal moment conditions under null hypothesis H_0: $c = 0$:

$$E(z_i' B' A \Delta z_i) = 0. \tag{38}$$

These conditions imply that $E(u_i u_i') = \sigma^2 I_T$. They can be tested based on the following LS estimator:

$$\hat{\varphi}_{FOD} = 1 + \frac{\sum_{i=1}^{N} z_i' B' A \Delta z_i}{\sum_{i=1}^{N} z_i' B' B z_i}, \tag{39}$$

which is equal to that of Breitung (2000) plus 1. To test conditions (38) in the case of serial correlation in u_{it}, we will first adjust estimator $\hat{\varphi}_{FOD}$ for its inconsistency, which arises from the presence of serial correlation. The next theorem derives the limiting distribution of estimator $\hat{\varphi}_{FOD}$ corrected for its inconsistency under $\varphi_N = 1 - \frac{c}{\sqrt{N}}$.

Theorem 5 *Under Assumptions 1, 2 and the assumption that the order of serial correlation* *is at most p, we have*

$$\sqrt{N}\hat{V}_{FOD}^{-1/2}\hat{\delta}_{FOD}\left(\hat{\varphi}_{FOD}-1-\frac{\hat{b}_{FOD}}{\hat{\delta}_{FOD}}\right)\xrightarrow{d}N(-ck_{FOD},1), \tag{40}$$

as $N\to\infty$, where

$$\begin{aligned} & k_{FOD} = \qquad\qquad (41)\\ = & \frac{tr(\Lambda'B'A\Lambda\Gamma)+tr(B'A\Lambda\Gamma)+tr(\Lambda'B'A\Gamma)+tr(F'B'A\Gamma)-tr(\Lambda'\Phi_{p,FOD}\Gamma)-tr(\Phi_{p,FOD}\Lambda\Gamma)}{2tr((A_{FOD}\Gamma)^2)}, \end{aligned}$$

$\frac{\hat{b}_{FOD}}{\hat{\delta}_{FOD}}=\frac{tr(\Phi_{p,FOD}\hat{\Gamma})}{\frac{1}{N}\sum_{i=1}^{N}z_i'B'Bz_i}$ *is a consistent estimator of the inconsistency of* $\hat{\varphi}_{FOD}$*,* $\Phi_{p,FOD}=\Psi_{p,FOD}-\frac{tr(\Xi M)}{e'Me}M$*, where* $\Psi_{p,FOD}$ *is a (TXT) matrix having in its diagonals* $\{-p,..,0,...p\}$ *the corresponding elements of matrix* Ξ *and zero everywhere else, where* $\Xi=\Lambda'B'A+B'A$*, and* $V_{FOD}=2tr((A_{FOD}\Gamma)^2)$*, with* $A_{FOD}=\frac{1}{2}(\Xi+\Xi'-\Phi_{p,FOD}-\Phi_{p,FOD}')$*, is the variance of the adjusted for its inconsistency estimator* $\hat{\varphi}_{FOD}$*.*

As with WG, the limiting distribution of the FOD test statistic depends on the estimator of Γ, $\hat{\Gamma}$. This estimator now becomes invariant of the incidental trends nuisance parameters effects by being premultiplied by selection matrix $\Phi_{p,FOD}$. Theorem 5 implies that, if there is no serial correlation in u_{it} (i.e., $p=0$), the asymptotic local power of the FOD test statistic is zero, since $k_{FOD}=0$.[9] As with WGT, the test has power only if there is serial correlation in u_{it}. These results are established in the next corollary, which gives values of the power slope parameter $k_{FOD}(p,\theta)$ in the case that u_{it} follows MA(1) process (12).

Corollary 5 *If error terms u_{it} follow MA(1) process (12), and Assumptions 1 and 2 hold, then slope parameter $k_{FOD}(p,\theta)$ is given as*

$$\begin{aligned} k_{FOD}(p,0) &= 0 \quad \text{for } p=0,1,2,...,T-2. & (42)\\ \text{and}\quad k_{FOD}(1,\theta) &\neq 0 \quad \text{for } \theta\neq 0. & (43) \end{aligned}$$

Values of $k_{FOD}(p,\theta)$, for $p=\{0,1\}$, $T\in\{7,10\}$ and $\theta\in\{-0.9,-0.5,0.0,0.5,0.9\}$, are

[9] In this case, it can be shown that

$$p\lim_{N\to\infty}(\hat{\phi}_{FOD}-1)=tr(\Xi\Gamma)=0,$$

since $tr(\Xi\Gamma)=0$ when $\Gamma=\sigma^2I_T$.In other words $\hat{\phi}_{FOD}$ is consistent. In this case a test that does not require a bias correction can be derived. In a previous version of this paper we showed that

$$\sqrt{N}V_{FOD}^{-1/2}(\hat{\phi}_{FOD}-1)\to N(-c0,1),$$

where $V_{FOD}=\frac{2tr((A_{\Xi})^2)}{tr((\Lambda'+I_T)B'B(\Lambda+I_T))^2}$ and $A_{\Xi}=\frac{1}{2}(\Xi+\Xi')$. The incidental trend problem remains. But if $\Gamma=\sigma^2I_T$, this version of the test has better finite sample properties.

given in Table 3. These indicate that the FOD test statistic has asymptotic local power only if $\theta < 0$, which increases slowly with T. The power of the test for $\theta < 0$ can be attributed to the effects of quantities $tr(\Lambda'\Phi_{p,FOD}\Gamma)$ and $tr(\Phi_{p,FOD}\Lambda\Gamma)$ on $k_{FOD}(p,\theta)$. These have the same qualitative effects on power slope parameter $k_{FOD}(p,\theta)$ to those of quantities $tr(\Phi_{p,WGT}\Lambda\Gamma)$ and $tr(\Lambda'\Phi_{p,WGT}\Gamma)$ on $k_{WGT}(p,\theta)$, for the WGT test statistic which has also no-trivial power when $\theta < 0$. The results of the table also indicate that the test has smaller local power than that of test statistics FDIV and WGT. Finally, if $T \to \infty$ the test has trivial power, as the WGT test statistic.

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 MA process (12). This is done for $N \in \{50, 100, 200, 300, 1000\}$, $T \in \{7, 10\}$, $\theta \in \{-0.9, -0.5, 0.0, 0.5, 0.9\}$ and $p \in \{0, 1\}$. The order of serial correlation p is assumed to be zero in the case of $\theta = 0.0$, otherwise it is set to $p = 1$. The nuisance parameters of models $M1$ and $M2$ which do not appear in the above local power functions are set to zero, i.e., $a_i = 0$, $\beta_i = 0$, $y_{i0} = 0$, for all i.

Tables 4 and 5 present the results 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 $M2$, allowing also for incidental trends. 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 that of the WG test independently of T, as is predicted by the theory. For $\theta \geq 0$, 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, and its small sample estimate deviates considerably from its theoretical value, TV. This can be obviously attributed to second, or higher order effects, which are not captured by the first order approximation of the local power function. 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. 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.

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. It is biased in small samples, and 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 the presence of more complicated deterministic terms (see also Moon et al. (2007) and Han and Phillips (2010)).

Table 4: Size and power of the IV and WG tests.

		T=7						T=10					
N		50	100	200	300	1000	TV	50	100	200	300	1000	TV
$\theta = -0.9$													
c=0	IV	0.068	0.065	0.057	0.048	0.054	0.050	0.064	0.054	0.053	0.051	0.051	0.050
	WG	0.053	0.054	0.053	0.053	0.053	0.050	0.050	0.050	0.048	0.047	0.047	0.050
c=1	IV	0.089	0.101	0.087	0.071	0.067	0.433	0.088	0.102	0.086	0.074	0.081	0.623
	WG	0.054	0.052	0.054	0.053	0.047	0.018	0.055	0.053	0.043	0.048	0.045	0.004
$\theta = -0.5$													
c=0	IV	0.053	0.053	0.053	0.051	0.051	0.050	0.050	0.054	0.053	0.047	0.050	0.050
	WG	0.045	0.050	0.047	0.049	0.043	0.050	0.049	0.053	0.052	0.052	0.049	0.050
c=1	IV	0.285	0.382	0.444	0.496	0.567	0.793	0.462	0.639	0.773	0.807	0.904	0.989
	WG	0.057	0.066	0.068	0.076	0.087	0.069	0.044	0.048	0.048	0.053	0.056	0.043
$\theta = 0$													
c=0	IV	0.087	0.073	0.070	0.066	0.066	0.050	0.069	0.065	0.062	0.065	0.060	0.050
	WG	0.058	0.058	0.053	0.053	0.049	0.050	0.051	0.052	0.047	0.051	0.049	0.050
c=1	IV	0.997	0.997	0.997	0.997	0.998	0.998	0.997	0.999	0.999	0.999	1.00	1.00
	WG	0.220	0.274	0.321	0.344	0.414	0.500	0.117	0.156	0.213	0.245	0.357	0.519
$\theta = 0.5$													
c=0	IV	0.072	0.062	0.058	0.055	0.054	0.050	0.071	0.075	0.068	0.056	0.061	0.050
	WG	0.047	0.049	0.049	0.048	0.052	0.050	0.049	0.052	0.053	0.049	0.055	0.050
c=1	IV	0.979	0.985	0.988	0.990	0.993	0.994	1.00	1.00	1.00	1.00	0.999	1.00
	WG	0.388	0.489	0.57	0.610	0.678	0.730	0.236	0.325	0.429	0.477	0.632	0.731
$\theta = 0.9$													
c=0	IV	0.063	0.067	0.061	0.063	0.061	0.050	0.077	0.068	0.063	0.067	0.057	0.050
	WG	0.041	0.045	0.048	0.049	0.054	0.050	0.053	0.050	0.047	0.057	0.053	0.050
c=1	IV	0.977	0.985	0.986	0.986	0.992	0.994	1.00	1.00	0.999	1.00	0.999	1.00
	WG	0.483	0.593	0.671	0.689	0.754	0.764	0.302	0.412	0.529	0.569	0.700	0.757

Table 5: Size and local power of FDIV, WGT and FOD tests.

		T=7						T=10					
N		50	100	200	300	1000	TV	50	100	200	300	1000	TV
$\theta = -0.9$													
c=0	FDIV	0.038	0.038	0.048	0.050	0.050	0.050	0.040	0.041	0.045	0.044	0.050	0.050
	WGT	0.044	0.049	0.045	0.047	0.051	0.050	0.049	0.046	0.048	0.054	0.049	0.050
	FOD	0.051	0.045	0.048	0.050	0.052	0.050	0.051	0.053	0.052	0.047	0.050	0.050
c=1	FDIV	0.031	0.044	0.039	0.047	0.041	0.216	0.036	0.045	0.043	0.047	0.046	0.313
	WGT	0.046	0.053	0.050	0.054	0.058	0.170	0.046	0.048	0.053	0.059	0.056	0.273
	FOD	0.089	0.070	0.078	0.068	0.067	0.067	0.129	0.113	0.102	0.091	0.078	0.067
$\theta = -0.5$													
c=0	FDIV	0.032	0.043	0.045	0.050	0.050	0.050	0.038	0.045	0.042	0.045	0.050	0.050
	WGT	0.043	0.049	0.047	0.051	0.047	0.050	0.048	0.055	0.052	0.050	0.047	0.050
	FOD	0.048	0.053	0.053	0.049	0.052	0.050	0.055	0.055	0.054	0.055	0.050	0.050
c=1	FDIV	0.039	0.039	0.045	0.048	0.055	0.226	0.043	0.046	0.050	0.047	0.047	0.338
	WGT	0.083	0.082	0.081	0.088	0.087	0.119	0.102	0.102	0.116	0.103	0.115	0.158
	FOD	0.100	0.089	0.0858	0.081	0.073	0.062	0.163	0.143	0.128	0.109	0.088	0.062
$\theta = 0$													
c=0	FDIV	0.054	0.055	0.056	0.055	0.056	0.050	0.052	0.055	0.052	0.049	0.053	0.050
	WGT	0.067	0.061	0.062	0.058	0.057	0.050	0.064	0.062	0.057	0.063	0.060	0.050
	FOD	0.053	0.053	0.049	0.056	0.051	0.050	0.061	0.059	0.056	0.057	0.052	0.050
c=1	FDIV	0.053	0.057	0.050	0.057	0.054	0.351	0.056	0.050	0.060	0.053	0.053	0.541
	WGT	0.137	0.105	0.094	0.084	0.070	0.050	0.155	0.127	0.106	0.094	0.074	0.050
	FOD	0.112	0.090	0.091	0.079	0.063	0.050	0.226	0.167	0.139	0.119	0.087	0.050
$\theta = 0.5$													
c=0	FDIV	0.042	0.048	0.048	0.046	0.045	0.050	0.048	0.053	0.050	0.049	0.050	0.050
	WGT	0.058	0.056	0.054	0.060	0.047	0.050	0.063	0.056	0.053	0.053	0.054	0.050
	FOD	0.057	0.059	0.057	0.057	0.054	0.050	0.069	0.065	0.061	0.063	0.055	0.050
c=1	FDIV	0.037	0.039	0.043	0.047	0.048	0.323	0.035	0.036	0.048	0.040	0.053	0.634
	WGT	0.083	0.068	0.057	0.053	0.038	0.031	0.152	0.122	0.085	0.069	0.051	0.031
	FOD	0.100	0.086	0.077	0.067	0.052	0.043	0.235	0.158	0.132	0.108	0.070	0.045
$\theta = 0.9$													
c=0	FDIV	0.038	0.036	0.048	0.047	0.048	0.050	0.045	0.050	0.050	0.045	0.043	0.050
	WGT	0.060	0.059	0.056	0.053	0.050	0.050	0.061	0.056	0.059	0.053	0.055	0.050
	FOD	0.059	0.065	0.056	0.047	0.047	0.050	0.072	0.067	0.052	0.052	0.055	0.050
c=1	FDIV	0.030	0.037	0.033	0.041	0.044	0.320	0.029	0.034	0.035	0.036	0.047	0.641
	WGT	0.072	0.053	0.040	0.033	0.023	0.029	0.135	0.099	0.071	0.054	0.035	0.029
	FOD	0.098	0.082	0.058	0.050	0.043	0.042	0.206	0.165	0.118	0.105	0.066	0.044

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

5 Conclusions

This paper examines the power properties of fixed-T panel data unit root tests under serial correlation, assuming that only the cross-section dimension of the panel (N) grows large. To this end, the paper provides an extension of the IV based test statistic of De Wachter et al. (2007), which exploits orthogonal moment conditions of the data under serial correlation, to allow for incidental trends. It also gives a fixed-T version of Breitung's (2000) test statistic, based on forward orthogonal deviations transformation of the data to avoid estimating incidental trends parameters, which allows for serial correlation in the error terms of the individual series of the panel. The paper derives the asymptotic local power functions of the above tests and LS based panel unit root statistics relying on the "within group" transformations of the data to wipe off individual effects or incidental trends. Analytic forms of these power functions are also derived for the case that the error terms of the panel follow a moving average procedure of lag-order one, often assumed in practice for many economic series.

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 "within group" LS based test. This can be attributed to the fact that the last test requires an adjustment of the LS 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 bigger 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, only the LS based tests relying on the "within group" and forward orthogonal deviations transformation of the individual series of the panel are found to have non-trivial power, as is predicted by the theory. These tests have always power when the serial correlation in the error term is negative. They also retain their power even for small N. This non-trivial power can be attributed to the impact of the inconsistency correction, required by the LS estimator, for the serial correlation nuisance parameters. For panel data models with incidental trends, the IV based test is found to be biased 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 LS based fixed-T panel unit root tests in mitigating the incidental trends problem in short panels with serially correlated error terms.

References

[1] 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.

[2] 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.

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

[4] 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.

[5] 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.

[6] Dickey D.A., and Fuller W.A., 1979. Distribution of estimators of autoregressive time series with a unit root. Journal of the American Statistical Association 74, 427-431.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] Hayakawa K., 2010. A unit root test for micro panels with serially correlated errors. Working Paper, Hiroshima University.

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

[14] Karavias, Y., and Tzavalis, E., 2012. Testing for unit roots in short panels allowing for structural breaks. Computational Statistics and Data Analysis. (In press). http://dx.doi.org/10.1016/j.csda.2012.10.014

[15] 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.

[16] 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.

[17] 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.

[18] 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.

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

[20] 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.

[21] Moon, H.R., Perron, B., Phillips, P.C.B., 2006. A note on "The local power of some unit root tests for panel data" by J. Breitung. Econometric Theory 22, 1177–1188.

[22] 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.

[23] Sargan, J. D. and A. Bhargava (1983). Testing residuals from least squares regression for being generated by the Gaussian random walk. Econometrica 51, 153–74.

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

[25] 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.

6 Appendix

Proof of Theorem 1 Under the sequence of local alternatives $\varphi_N = 1 - \frac{c}{\sqrt{N}}$, the IV test statistic can be written as follows:

$$\sqrt{N}(\hat{\varphi}_{IV} - \varphi_N) = \sqrt{N}\left(\frac{\frac{1}{N}\sum_{i=1}^{N} z'_{i-1}\Pi_p u_i + \frac{1}{N}\sum_{i=1}^{N}(1-\varphi_N)a_i z'_{i-1}\Pi_p e}{\frac{1}{N}\sum_{i=1}^{N} z'_{i-1}\Pi_p z_{i-1}} + \varphi_N - \varphi_N\right)$$

$$= \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N} z'_{i-1}\Pi_p u_i + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(1-\varphi_N)a_i z'_{i-1}\Pi_p e}{\frac{1}{N}\sum_{i=1}^{N} z'_{i-1}\Pi_p z_{i-1}} = \frac{(\alpha)+(b)}{(g)}, \tag{44}$$

where $(\alpha) \equiv \frac{1}{\sqrt{N}}\sum_{i=1}^{N} z'_{i-1}\Pi_p u_i$, $(b) \equiv \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(1-\varphi_N)a_i z'_{i-1}\Pi_p e$ and $(g) \equiv \frac{1}{N}\sum_{i=1}^{N} z'_{i-1}\Pi_p z_{i-1}$. Under H_1: $c > 0$, vector y_{i-1} can be expanded as

$$y_{i-1} = wy_{i0} + \Omega e(1-\varphi_N)a_i + \Omega u_i,\ i = 1, 2, ..., N, \tag{45}$$

where

$$\Omega = \begin{pmatrix} 0 & . & . & . & . & . & 0 \\ 1 & 0 & & & & & . \\ \varphi_N & 1 & . & & & & . \\ \varphi_N^2 & \varphi_N & . & . & & & . \\ . & & . & . & . & & . \\ . & & & . & 1 & 0 & . \\ \varphi_N^{T-2} & \varphi_N^{T-3} & . & . & \varphi_N & 1 & 0 \end{pmatrix} \tag{46}$$

and $w = (1, \varphi_N, \varphi_N^2, ..., \varphi_N^{T-1})'$. Note that, for $\varphi_N = 1$, we have $\Omega \equiv \Lambda$.

The first order Taylor expansions of Ω and w yields

$$\Omega = \Lambda + F(\varphi_N - 1) + o_p(1) \text{ and} \tag{47}$$

$$w = e + f(\varphi_N - 1) + o_P(1), \tag{48}$$

respectively, where $F = \frac{d\Omega}{d\varphi_N} \mid_{\varphi_N=1}$ and $f = \frac{dw}{d\varphi_N} \mid_{\varphi_N=1}$. Using these expansions, vector $z_{i-1} = y_{i-1} - y_{i0}e$ can be written as

$$z_{i-1} = y_{i-1} - ey_{i0} = (w-e)y_{i0} + \Omega e(1-\varphi_N)a_i + \Omega u_i. \tag{49}$$

Substituting (49) into quantity (a), defined by (44), yields

$$\begin{aligned}(a) &\equiv \frac{1}{\sqrt{N}}\sum_{i=1}^{N} z'_{i-1}\Pi_p u_i = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}((w-e)y_{i0} + \Omega e(1-\varphi_N)a_i + \Omega u_i)'\Pi_p u_i \\ &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i0}(w-e)'\Pi_p u_i + (1-\varphi_N)a_i e'\Omega'\Pi_p u_i + u_i\Omega'\Pi_p u_i.\end{aligned}$$

This quantity has a limiting distribution $N(0, 2tr((A_{IV}\Gamma)^2))$, since the following results hold:

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i0}(w-e)'\Pi_p u_i \xrightarrow{p} 0 \tag{50}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(1-\varphi_N)a_i e'\Omega'\Pi_p u_i \xrightarrow{p} 0 \tag{51}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} u_i\Omega'\Pi_p u_i \xrightarrow{d} N(0, 2tr((A_{IV}\Gamma)^2)) \tag{52}$$

The results given by (50) and (51) can be derived by using (47)-(48) and standard results on quadratic forms (see Schott (1997)), while (52) can be proved using $tr(\Lambda'\Pi_p\Gamma) = tr(F'\Pi_p\Gamma) = 0$ and Lindeberg-Levy's CLT. For quantities (b) and (g), the following results can be easily derived:

$$(b) \equiv \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(1-\varphi_N)a_i z'_{i-1}\Pi_p e \xrightarrow{p} 0 \tag{53}$$

$$\text{and} \qquad (g) \equiv \frac{1}{N}\sum_{i=1}^{N} z'_{i-1}\Pi_p z_{i-1} \xrightarrow{p} tr(\Lambda'\Pi_p\Lambda\Gamma). \tag{54}$$

Using (50)-(54) yields

$$\begin{aligned}&\sqrt{N}(\hat{\varphi}_{IV} - \varphi_N) \xrightarrow{d} N\left(0, \frac{2tr((A_{IV}\Gamma)^2)}{tr(\Lambda'\Pi_p\Lambda\Gamma)^2}\right) \\ &\sqrt{N}(\hat{\varphi}_{IV} - 1 - \frac{c}{\sqrt{N}}) \xrightarrow{d} N\left(0, \frac{2tr((A_{IV}\Gamma)^2)}{tr(\Lambda'\Pi_p\Lambda\Gamma)^2}\right) \\ &\sqrt{N}(\hat{\varphi}_{IV} - 1) \xrightarrow{d} N(-c, V_{IV}) \\ &\sqrt{N}V_{IV}^{-1/2}(\hat{\varphi}_{IV} - 1) \xrightarrow{d} N\left(-cV_{IV}^{-1/2}, 1\right),\end{aligned} \tag{55}$$

where $V_{IV} = \frac{2tr((A_{IV}\Gamma)^2)}{tr(\Lambda'\Pi_p\Lambda\Gamma)^2}$, which proves Theorem 1.

Proof of Corollary 1 The results of the corollary can be proved based on the formula of the variance of the limiting distribution of the IV test $V_{IV} = \frac{2tr((A_{IV}\Gamma)^2)}{tr(\Lambda'\Pi_p\Lambda\Gamma)^2}$, $A_{IV} = \frac{1}{2}(\Lambda'\Pi_p + \Pi_p'\Lambda)$, given by Theorem 1. Under no serial correlation (i.e., $\theta = 0$ and $\Gamma_N = \sigma_u^2 I_T$), the following relationship holds:

$$2tr({A_{IV}}^2) = tr(\Lambda'\Pi_p\Lambda), \text{ for all } p. \tag{56}$$

This yields

$$\begin{aligned}
tr(\Lambda'\Pi_p\Lambda) &= \frac{1}{2}(T^2 - T), \text{ for } p = 0 \\
tr(\Lambda'\Pi_p\Lambda) &= \frac{[\frac{1}{2}(T-2)(T-1)]^2}{\frac{1}{2}T(T-3)+1}, \text{ for } p = 1 \\
tr(\Lambda'\Pi_p\Lambda) &= \frac{T^2}{2} - \frac{5T}{2} + 3, \text{ for } p = 2 \\
\text{and} \quad tr(\Lambda'\Pi_p\Lambda) &= \frac{T^2}{2} - \frac{7T}{2} + 6, \text{ for } p = 3
\end{aligned}$$

Using the last relationships, we can derive the following values of the slope parameter of the power function $k_{IV}(p, \theta)$, for $\theta = 0$:

$$\begin{aligned}
k_{IV}(0,0) &= \sqrt{\frac{1}{2}(T^2 - T)}, \\
k_{IV}(1,0) &= \sqrt{\frac{T^2}{2} - \frac{3T}{2} + 1}, \\
k_{IV}(2,0) &= \sqrt{\frac{T^2}{2} - \frac{5T}{2} + 3}, \\
\text{and} \quad k_{IV}(3,0) &= \sqrt{\frac{T^2}{2} - \frac{7T}{2} + 6}.
\end{aligned}$$

For the case of $\theta \neq 0$ and $p = 1$, the formula of $k_{IV}(p, \theta)$ is derived by De Wachter et al. (2007). The coefficients of this formula are analytically given as

$$\begin{aligned}
D_{1,IV} &= \frac{T^2}{2} - \frac{3T}{2} + 1 \\
D_{2,IV} &= T^2 - 4T + 4 \\
R_{1,IV} &= \frac{1}{2}T(T-3) + 1 \\
R_{2,IV} &= 2T(T-5) + 12 \\
R_{3,IV} &= 3T(T-5) + 20.
\end{aligned}$$

Proof of Theorem 2 The proof of the theorem for the case $c = 0$ is given, separately, in Part I, for a direct comparison to that of Kruiniger and Tzavalis (2002). For case $c > 0$, it is given in Part II.

I) To derive the limiting distribution of test statistic WG under null hypothesis H_0: $c = 0$, we will proceed into stages. First, we will show that the LS estimator $\hat{\varphi}_{WG}$ is inconsistent, as $N \to \infty$. Then, we will construct a normalized statistic based on $\hat{\varphi}_{WG}$ corrected for its inconsistency and we will derive its limiting distribution under H_0: $c = 0$, as $N \to \infty$.

Decompose vector y_{i-1} for model (1) under H_0: $c = 0$ as

$$y_{i-1} = ey_{i0} + \Lambda u_i, \tag{57}$$

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

Premultiplying (57) with matrix Q yields

$$Qy_{i,-1} = Q\Lambda u_i, \tag{58}$$

since $Qe = (0, 0, ..., 0)'$. Substituting (58) into $\hat{\varphi}_{WG}$ yields

$$\hat{\varphi}_{WG} - 1 = \frac{\frac{1}{N}\sum_{i=1}^{N} y_{i,-1}' Q u_i}{\frac{1}{N}\sum_{i=1}^{N} y_{i,-1}' Q y_{i,-1}} = \frac{\frac{1}{N}\sum_{i=1}^{N} u_i' \Lambda' Q u_i}{\frac{1}{N}\sum_{i=1}^{N} u_i' \Lambda' Q \Lambda u_i}. \tag{59}$$

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

$$\frac{1}{N}\sum_{i=1}^{N} u_i' \Lambda' Q u_i \xrightarrow{p} tr(\Lambda' Q \Gamma) \text{ and } \frac{1}{N}\sum_{i=1}^{N} u_i' \Lambda' Q \Lambda u_i \xrightarrow{p} tr(\Lambda' Q \Lambda \Gamma), \tag{60}$$

where "$\xrightarrow{p}$" signifies convergence in probability. Based on the last two results, the yet non-standardized test statistic WG can be written as

$$\begin{aligned} \sqrt{N}\hat{\delta}_{WG}\left(\hat{\varphi}_{WG} - 1 - \frac{\hat{b}_{WG}}{\hat{\delta}_{WG}}\right) &= \sqrt{N}\hat{\delta}_{WG}\left(\frac{\frac{1}{N}\sum_{i=1}^{N} y_{i,-1}' Q u_i}{\hat{\delta}_{WG}} - \frac{tr(\Psi_{p,WG}\hat{\Gamma})}{\hat{\delta}_{WG}}\right) \\ &= \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} y_{i,-1}' Q u_i - \frac{1}{N}\sum_{i=1}^{N} \Delta y_i' \Psi_{p,WG} \Delta y_i\right), \end{aligned} \tag{61}$$

where

$$\hat{\Gamma} = \frac{1}{N}\sum_{i=1}^{N} \Delta y_i' \Psi_{p,WG} \Delta y_i. \tag{62}$$

Since, under H_0: $c = 0$ we have $u_i = \Delta y_i$, the last relationship can be written as follows:

$$\begin{aligned}
& \sqrt{N}\delta\left(\hat{\varphi}_{WG} - 1 - \frac{\hat{b}_{WG}}{\hat{\delta}_{WG}}\right) \\
= & \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} u_i'\Lambda' Q u_i - \frac{1}{N}\sum_{i=1}^{N} u_i'\Psi_{p,WG} u_i\right) \\
= & \frac{1}{\sqrt{N}}\sum_{i=1}^{N} u_i'(\Lambda' Q - \Psi_{p,WG})u_i = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} tr\left[(\Lambda' Q - \Psi_{p,WG})u_i u_i'\right] \qquad (63) \\
= & \frac{1}{\sqrt{N}}\sum_{i=1}^{N} W_i,
\end{aligned}$$

where W_i constitutes a random variable with zero mean, i.e.,

$$\begin{aligned}
E(W_i) &= E[u_i'(\Lambda' Q - \Psi_{p,WG})u_i] = tr[(\Lambda' Q - \Psi_{p,WG})E(u_i u_i')] \\
&= tr(\Lambda' Q - \Psi_{p,WG}) = 0
\end{aligned}$$

since $tr(\Lambda' Q) = tr(\Psi_{p,WG})$ (or $tr(\Lambda' Q - \Psi_{p,WG}) = 0$), and variance

$$Var(W_i) = Var(u_i'(\Lambda' Q - \Psi_{p,WG})u_i) = 2tr((A_{WG}\Gamma)^2). \qquad (64)$$

The last relationship follows from standard linear algebra results (see e.g. Schott(1997). The results of Theorem 2 follow by applying Lindeberg-Levy's CLT to the sequence of IID random variables W_i.

II) To derive the limiting distribution of the WG test under H: $c > 0$, subtract y_{i-1} from both sides of model $M1$:

$$\Delta y_i = u_i + (\varphi_N - 1)y_{i-1} + (1 - \varphi_N)a_i e. \qquad (65)$$

The limiting distribution of the yet unstandardized WG test statistic around φ_N can be obtained by writing

$$\hat{\delta}_{WG}\sqrt{N}\left(\hat{\varphi}_{WG}-\frac{\hat{b}_{WG}}{\hat{\delta}_{WG}}-\varphi_N\right) = \hat{\delta}_{WG}\sqrt{N}\left(\varphi_N+\frac{\frac{1}{N}\sum_{i=1}^{N}y'_{i-1}Qu_i}{\frac{1}{N}\sum_{i=1}^{N}y'_{i-1}Qy'_{i-1}}-\frac{\hat{b}_{WG}}{\hat{\delta}_{WG}}-\varphi_N\right)$$

$$= \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}y'_{i-1}Qu_i - tr(\Psi_{p,WG}\hat{\Gamma})\right)$$

$$= \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}y'_{i-1}Qu_i - \frac{1}{N}\sum_{i=1}^{N}\Delta y'_i\Psi_{p,WG}\Delta y_i\right)$$

$$= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}y'_{i-1}Qu_i - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Delta y'_i\Psi_{p,WG}\Delta y_i = (d)-(h), \quad (66)$$

where $(d) \equiv \frac{1}{\sqrt{N}}\sum_{i=1}^{N}y'_{i-1}Qu_i$ and $(h) \equiv \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Delta y'_i\Psi_{p,WG}\Delta y_i$. By applying Lindeberg-Levy's CLT, we can find the limiting distributions of quantities (d) and (h). To this end, write (d) as

$$\begin{aligned}(d) &\equiv \frac{1}{\sqrt{N}}\sum_{i=1}^{N}y'_{i-1}Qu_i = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(wy_{i0}+\Omega e(1-\varphi_N)a_i+\Omega u_i)'Qu_i \\ &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(y_{i0}w'Qu_i + a_i(1-\varphi_N)e'\Omega'Qu_i + u'_i\Omega'Qu_i\right) \\ &= (d_1)+(d_2)+(d_3),\end{aligned} \quad (67)$$

where $(d_1) \equiv \frac{1}{\sqrt{N}}\sum_{i=1}^{N}y_{i0}w'Qu_i$, $(d_2) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}a_i(1-\varphi_N)e'\Omega'Qu_i$ and $(d_3) \equiv \frac{1}{\sqrt{N}}\sum_{i=1}^{N}u'_i\Omega'Qu_i$. The limits of (d_1), (d_2) and (d_3) can be obtained after substituting Taylor's series expansions for w and Ω given in (47) and (48), respectively, into them and $\varphi_N = 1-\frac{c}{\sqrt{N}}$. For (d_1), we have

$$\begin{aligned}(d_1) &\equiv \frac{1}{\sqrt{N}}\sum_{i=1}^{N}y_{i0}w'Qu_i = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}y_{i0}(e+f(\varphi_N-1)+o_P(1))'Qu_i \\ &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}y_{i0}e'Qu_i + \frac{1}{N}\sum_{i=1}^{N}f'Qu_iy_{i0} + o_p(1) \xrightarrow{p} 0,\end{aligned} \quad (68)$$

since $\frac{1}{N}\sum_{i=1}^{N}f'Qu_iy_{i0} \to f'QE(u_iy_{i0}) = 0$ by Assumption 2 and $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}y_{i0}e'Qu_i = 0$, as

$e'Q = 0$. For (d_2), we have

$$(d_2) \equiv \frac{1}{\sqrt{N}}\sum_{i=1}^{N} a_i(1-\varphi_N)e'\Omega'Qu_i = \frac{c}{N}\sum_{i=1}^{N} a_i e'(\Lambda' + F'\left(\frac{-c}{\sqrt{N}}\right) + o_p(1))Qu_i \quad (69)$$

$$= \frac{c}{N}\sum_{i=1}^{N} a_i e'\Lambda' \; Qu_i - \frac{c^2}{N^{3/2}}\sum_{i=1}^{N} a_i e'F'Qu_i + o_p(1) \xrightarrow{p} 0,$$

since $\frac{c}{N}\sum_{i=1}^{N} a_i e'\Lambda' \; Qu_i \xrightarrow{p} c_i e'\Lambda' \; QE(a_i u_i) = 0$ by Assumption 2 and $\frac{c^2}{N^{3/2}}\sum_{i=1}^{N} a_i e'F'Qu_i \xrightarrow{p} 0$, as $\frac{1}{N^{3/2}}$ goes to zero. For (d_3), we have

$$(d_3) \equiv \frac{1}{\sqrt{N}}\sum_{i=1}^{N} u_i'\Omega'Qu_i = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} u_i'(\Lambda' + F'(\frac{-c}{\sqrt{N}}) + o_p(1))Qu_i \quad (70)$$

$$= \frac{1}{\sqrt{N}}\sum_{i=1}^{N} u_i'\Lambda'Qu_i - \frac{c}{N}\sum_{i=1}^{N} u_i'F'Qu_i + o_p(1),$$

where

$$\frac{c}{N}\sum_{i=1}^{N} u_i'F'Qu_i \xrightarrow{p} ctr(F'Q\Gamma), \quad (71)$$

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} u_i'\Lambda'Qu_i - tr(\Lambda'Q\Gamma)\right) \xrightarrow{d} N(0, V_{WG,d}). \quad (72)$$

$V_{WG,d}$ is the variance of the limiting distribution of (d_3) and (d), since $(d_1) \xrightarrow{p} 0$ and $(d_1) \xrightarrow{p} 0$. It can be assumed as a known quantity, under the normality assumption (see Assumption 1). Term $tr(\Lambda'Q\Gamma)$, which is added in (72), does not have to be subtracted from the WG test statistic, as it cancels out with a similarly added term, given as $tr(\Psi_{p,WG}\Gamma)$, in (74) below, since by construction $tr(\Lambda'Q\Gamma) = tr(\Psi_{p,WG}\Gamma)$.

Using the results of equations (68), (69) and (70), we can obtain the limiting distribution of (d) as

$$(d) \equiv \frac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i-1}'Qu_i \xrightarrow{d} N(-ctr(F'Q\Gamma), V_{WG,d}). \quad (73)$$

The proof for the limiting distribution of quantity (h) follows analogous steps to those of

(d), but it is more tedious. Substituting (65) in (h) yields:

$$
\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Delta y_i'\Psi_{p,WG}\Delta y_i = \\
= &\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(u_i+(\varphi_N-1)y_{i-1}+(1-\varphi_N)a_i e)'\Psi_{p,WG}(u_i+(\varphi_N-1)y_{i-1}+(1-\varphi_N)a_i e) \\
= &\frac{1}{\sqrt{N}}\sum_{i=1}^{N} u_i'\Psi_{p,WG}u_i + u_i'\Psi_{p,WG}y_{i-1}(\varphi_N-1) + u_i'\Psi_{p,WG}e(1-\varphi_N)a_i + (\varphi_N-1)y_{i-1}'\Psi_{p,WG}u_i \\
&+(\varphi_N-1)^2 y_{i-1}'\Psi_{p,WG}y_{i-1} + (\varphi_N-1)y_{i-1}'\Psi_{p,WG}e(1-\varphi_N)a_i + (1-\varphi_N)a_i e'\Psi_{p,WG}u_i \\
&+(1-\varphi_N)a_i e'\Psi_{p,WG}y_{i-1}(\varphi_N-1) + (1-\varphi_N)^2 a_i^2 e'\Psi_{p,WG}e
\end{aligned}
$$

Then, we can derive the following results:

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}u_i'\Psi_{p,WG}u_i - tr(\Psi_{p,WG}\Gamma)\right) \xrightarrow{d} N(0, V_{WG,h}) \tag{74}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}u_i'\Psi_{p,WG}y_{i-1}(\varphi_N-1) \xrightarrow{p} -c\,tr(\Psi_{p,WG}\Lambda\Gamma) \tag{75}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}u_i'\Psi_{p,WG}e(1-\varphi_N)a_i \xrightarrow{p} 0 \tag{76}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\varphi_N-1)y_{i-1}'\Psi_{p,WG}u_i \xrightarrow{p} -c\,tr(\Lambda'\Psi_{p,WG}\Gamma) \tag{77}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\varphi_N-1)^2 y_{i-1}'\Psi_{p,WG}y_{i-1} \xrightarrow{p} 0 \tag{78}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\varphi_N-1)y_{i-1}'\Psi_{p,WG}e(1-\varphi_N)a_i \xrightarrow{p} 0 \tag{79}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(1-\varphi_N)a_i e'\Psi_{p,WG}u_i \xrightarrow{p} 0 \tag{80}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(1-\varphi_N)a_i e'\Psi_{p,WG}y_{i-1}(\varphi_N-1) \xrightarrow{p} 0 \tag{81}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(1-\varphi_N)^2 a_i^2 e'\Psi_{p,WG}e \xrightarrow{p} 0 \tag{82}$$

The above results, given by equations (74)-(82), imply that the limiting distribution of (h)

is given as

$$(h) \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Delta y_i' \Psi_{p,WG} \Delta y_i \to N(-ctr(\Lambda' \Psi_{p,WG} \Gamma) - ctr(\Psi_{p,WG} \Lambda \Gamma), V_{WG,h}), \tag{83}$$

where $V_{WG,h}$ is the variance of the distribution, defined by (74). Based on these results and (73), we can derive the limiting distribution of WG around φ_N as

$$\hat{\delta}_{WG} \sqrt{N} \left(\hat{\varphi}_{WG} - \frac{\hat{b}_{WG}}{\hat{\delta}_{WG}} - \varphi_N \right) \xrightarrow{d} N \left(-c(tr(F'Q\Gamma) - tr(\Lambda' \Psi_{p,WG} \Gamma) - tr(\Psi_{p,WG} \Lambda \Gamma)), V_{WG} \right). \tag{84}$$

The variances of (d) and (h) and their covariance add up to the variance of the of statistic WG under H_0: $c = 0$, given by Theorem 2 (see also (64)). It is not necessary to show this algebraically because, as can be seen from (72) and (74), the variance of the estimator under H_0: $c > 0$ does not depend on the local parameter c. It is constant independently on whether $c > 0$ or $c = 0$. Given the above result and

$$\frac{1}{N} \sum_{i=1}^{N} y_{i-1}' Q y_{i-1} \xrightarrow{p} tr(\Lambda' Q \Lambda \Gamma), \tag{85}$$

we can obtain the limiting distribution of the WG test statistic, given Theorem 2, as follows:

$$\sqrt{N} \hat{\delta}_{WG} \left(\hat{\varphi}_{WG} - \frac{\hat{b}_{WG}}{\hat{\delta}_{WG}} - 1 + \frac{c}{\sqrt{N}} \right) \xrightarrow{d} N \left(-c(tr(F'Q\Gamma) - tr(\Lambda \Psi_{p,WG} \Gamma) - tr(\Lambda \Psi_{p,WG} \Gamma)), V_{WG} \right) \tag{86}$$

$$\sqrt{N} \hat{\delta}_{WG} \left(\hat{\varphi}_{WG} - \frac{\hat{b}_{WG}}{\hat{\delta}_{WG}} - 1 \right) \xrightarrow{d} N \left(-c(tr(\Lambda' Q \Lambda \Gamma) + tr(F'Q\Gamma) - tr(\Lambda \Psi_{p,WG} \Gamma) - tr(\Lambda \Psi_{p,WG} \Gamma)), V_{WG} \right) \tag{87}$$

$$\sqrt{N} \hat{\delta}_{WG} V_{WG}^{-1/2} \left(\hat{\varphi}_{WG} - \frac{\hat{b}_{WG}}{\hat{\delta}_{WG}} - 1 \right) \xrightarrow{d} N \left(-c \frac{(tr(\Lambda' Q \Lambda \Gamma) + tr(F'Q\Gamma) - tr(\Lambda \Psi_{p,WG} \Gamma) - tr(\Lambda \Psi_{p,WG} \Gamma))}{V_{WG}^{1/2}}, 1 \right). \tag{88}$$

Proof of Corollary 2 Equation (13) of the corollary, implying $\Gamma = \sigma_u^2 I$, can be proved by substituting into (18) the following relationships:

$$\begin{aligned} tr(\Lambda'\Psi_{p,WG}) &= 0, \\ tr(\Psi_{p,WG}\Lambda) &= 0, \\ tr(F'Q) &= -\frac{T^2}{6} + \frac{T}{2} - \frac{1}{3}, \\ tr(\Lambda'Q\Lambda) &= \frac{T^2-1}{6} \end{aligned}$$

and writing

$$\begin{aligned} tr(A_{WG}^2) &= tr\left[\frac{1}{4}\left(\Lambda'Q + Q\Lambda - \Psi_{p,WG} - \Psi'_{p,WG}\right)^2\right] \\ &= \frac{1}{2}tr((\Lambda'Q)^2) + \frac{1}{2}tr(\Lambda'Q\Lambda) - tr(\Psi_{p,WG}^2) \end{aligned}$$

where $tr((\Lambda'Q)^2) = -\frac{T^2}{12} + \frac{T}{2} - \frac{5}{12}$ and $tr(\Psi_{p,WG}^2) = \frac{T}{3} + \frac{1}{6T} - \frac{1}{2}$, since the following relationships hold:

$$\begin{aligned} tr(\Lambda'Q\Psi_{p,WG}) &= tr(\Lambda'Q\Psi'_{p,WG}) = tr(Q\Lambda\Psi_{p,WG}) = tr(Q\Lambda\Psi'_{p,WG}) \\ tr(\Psi_{p,WG}\Lambda'Q) &= tr(\Psi_{p,WG}Q\Lambda) = tr(\Psi'_{p,WG}\Lambda'Q) = tr(\Psi'_{p,WG}Q\Lambda) \\ tr(\Psi'_{p,WG}\Psi_{p,WG}) &= tr(\Psi_{p,WG}^2) = tr(\Lambda'Q\Psi_{p,WG}) = tr(\Psi_{p,WG}\Lambda'Q). \end{aligned}$$

For $p > 0$, the last relationships become:

$$\begin{aligned} tr(\Psi_{p,WG}\Psi'_{p,WG}) &= tr(\Lambda'Q\Psi_{p,WG}) = tr(Q\Lambda\Psi'_{p,WG}) = tr(\Psi_{p,WG}\Lambda'Q) = tr(\Psi'_{p,WG}Q\Lambda) \\ tr(\Psi_{p,WG}^2) &= tr(Q\Lambda\Psi_{p,WG}) = tr(\Lambda'Q\Psi'_{p,WG}) = tr(\Psi_{p,WG}Q\Lambda) = tr(\Psi'_{p,WG}\Lambda'Q) \end{aligned}$$

and, thus,

$$tr(A_{WG}^2) = \frac{1}{2}tr((\Lambda'Q)^2) + \frac{1}{2}tr(\Lambda'Q\Lambda) - tr(\Psi_{p,WG}^2) - \frac{1}{2}tr(\Psi_{p,WG}\Psi'_{p,WG}).$$

Thus, we have the following results: for $p = 1$

$$\begin{aligned} tr(\Lambda'\Psi_{p,WG}) &= \frac{T-1}{2}, \\ tr(\Psi\Lambda_p) &= -\frac{T}{2} - \frac{1}{T} + \frac{3}{2}, \\ tr(\Psi_{p,WG}^2) &= -\frac{1}{2T} + \frac{1}{2}, \\ tr(\Psi_{p,WG}\Psi'_{p,WG}) &= T + \frac{5}{2T} - \frac{1}{T^2} - \frac{5}{2}, \end{aligned}$$

for $p = 2$

$$
\begin{aligned}
tr(\Psi_{p,WG}\Lambda) &= -\frac{(T-2)^2}{T}, \\
tr(\Lambda'\Psi_{p,WG}) &= \frac{(T-1)^2}{T}, \\
tr(\Psi_{p,WG}^2) &= -\frac{1}{3}T - \frac{25}{6T} + \frac{2}{T^2} + \frac{5}{2}, \\
tr(\Psi_{p,WG}\Psi'_{p,WG}) &= \frac{5}{3}T + \frac{65}{6T} - \frac{7}{T^2} - \frac{13}{2},
\end{aligned}
$$

and for $p = 3$

$$
\begin{aligned}
tr(\Psi_{p,WG}\Lambda) &= -\frac{3T}{2} - \frac{10}{T} + \frac{15}{2}, \\
tr(\Lambda'\Psi_{p,WG}) &= \frac{3T}{2} + \frac{4}{T} - \frac{9}{2}, \\
tr(\Psi_{p,WG}^2) &= -\frac{2}{3}T - \frac{77}{6T} + \frac{10}{T^2} + 11, \\
tr(\Psi_{p,WG}\Psi'_{p,WG}) &= \frac{7}{3}T + \frac{175}{6T} - \frac{26}{T^2} - \frac{25}{2}.
\end{aligned}
$$

Substituting the above relationships into (18), we can obtain the following analytic forms of $k_{WG}(p,\theta)$, for $\theta = 0$:

$$
\begin{aligned}
k_{WG}(1,0) &= \frac{\sqrt{3}(T^2 - 3T + 2)}{T\sqrt{T^2 - 6T - \frac{24}{T} + \frac{12}{T^2} + 17}}, \\
k_{WG}(2,0) &= \frac{\sqrt{3}(T^2 - 5T + 6)}{T\sqrt{T^2 - 10T - \frac{80}{T} + \frac{60}{T^2} + 41}}, \\
\text{and} \quad k_{WG}(3,0) &= \frac{\sqrt{3}(T^2 - 7T + 12)}{T\sqrt{T^2 - 14T - \frac{196}{T} + \frac{192}{T^2} + 77}}.
\end{aligned}
\tag{89}
$$

To prove equation (20) of the corollary, for $\theta \neq 0$, substitute into (18) the following relationships:

$$\begin{aligned}
tr(\Lambda' Q \Lambda \Gamma) &= (\frac{T^2-1}{6})\theta^2 + (\frac{1}{3}T^2 - T + \frac{2}{3})\theta + (\frac{T^2-1}{6}), \\
tr(F' Q \Gamma) &= (-\frac{T^2}{6} + \frac{T}{2} - \frac{1}{3})\theta^2 + (-\frac{T^2}{3} + \frac{3T}{2} + \frac{1}{T} - \frac{13}{6})\theta + (-\frac{T^2}{6} + \frac{T}{2} - \frac{1}{3}), \\
tr(\Lambda' \Psi \Gamma) &= \frac{1}{2}(T-1)\theta^2 + (-\frac{2}{T} + 1)\theta + \frac{1}{2}(T-1), \\
tr(\Psi \Lambda \Gamma) &= (-\frac{T}{2} - \frac{1}{T} + \frac{3}{2})\theta^2 + (-T - \frac{4}{T} + 4)\theta + (-\frac{T}{2} - \frac{1}{T} + \frac{3}{2}), \\
\text{and } \; 2tr((A_{WG}\Gamma)^2) &= R_{1,WG}\theta^4 + R_{2,WG}\theta^3 + R_{3,WG}\theta^2 + R_{2,WG}\theta + R_{1,WG},
\end{aligned}$$

where $R_{1,WG} = \frac{T^2}{12} - \frac{T}{2} - \frac{2}{T} + \frac{1}{T^2} + \frac{17}{12}$, $R_{2,WG} = \frac{T^2}{3} - \frac{10T}{3} - \frac{80}{3T} + \frac{20}{T^2} + \frac{41}{3}$ and $R_{3,WG} = \frac{T^2}{2} - 5T - \frac{45}{T} + \frac{38}{T^2} + \frac{43}{2}$.

Proof of Proposition 1 The proof of (22), for $\theta = 0$, is given by Moon and Peron (2008). The results of both equations (21) and (22) of the corollary can be seen based on the results of Corollaries 1 and 2, after scaling the IV and WG statistics appropriately with T and applying the continuous mapping theorem. The joint convergence of the scaled statistics is guaranteed by the results of Hahn and Kuersteiner (2002). Since this proof is not intuitive/clear under the sequence of local alternatives considered by the above corollaries, we give a more rigorous proof of (21) under the sequence of local alternatives $\varphi_{NT} = 1 - \frac{c}{T\sqrt{N}}$, considered in the large-T panel data literature.

Rewrite the IV estimator as $\hat{\varphi}_{IV}$

$$\hat{\varphi}_{IV} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T-p-1} z_{it} z_{it+p+1}}{\sum_{i=1}^{N} \sum_{t=1}^{T-p-1} z_{it} z_{it+p}}.$$

Then, the yet not standardized IV test statistic under the sequence of local alternatives

$\varphi_{NT} = 1 - \frac{c}{T\sqrt{N}}$can be written as follows:

$$
\begin{aligned}
T\sqrt{N}(\hat{\varphi}_{IV} - \varphi_{NT}) &= T\sqrt{N}\left(\frac{\sum_{i=1}^{N}\sum_{t=1}^{T-p-1} z_{it}(\varphi_{NT}z_{it+p} + u_{it+p+1} + (1-\varphi_{NT})a_i)}{\sum_{i=1}^{N}\sum_{t=1}^{T-p-1} z_{it}z_{it+p}} - \varphi_{NT} \right) \\
&= \frac{\frac{1}{\sqrt{N}}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T-p-1}(z_{it}u_{it+p+1} + z_{it}(1-\varphi_{NT})a_i))}{\frac{1}{N}\frac{1}{T^2}\sum_{i=1}^{N}\sum_{t=1}^{T-p-1} z_{it}z_{it+p}} = \frac{(m)+(k)}{(l)}, \qquad (90)
\end{aligned}
$$

where $(m) \equiv \frac{1}{\sqrt{N}}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T-p-1}(z_{it}u_{it+p+1})$, $(k) \equiv \frac{1}{\sqrt{N}}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T-p-1} z_{it}(1-\varphi_{NT})a_i)$ and $(l) \equiv \frac{1}{N}\frac{1}{T^2}\sum_{i=1}^{N}\sum_{t=1}^{T-p-1} z_{it}z_{it+p}$. Under the sequence of $\varphi_{NT} = 1 - \frac{c}{T\sqrt{N}}$, z_{it} can be written as

$$
\begin{aligned}
z_{it} &= \varphi_{NT}^{t}z_{i0} + \varphi_{NT}^{t-1}u_{it-1} + ... + u_{it} \qquad (91) \\
&= \varphi_{NT}^{t-1}u_{i1} + \varphi_{NT}^{t-2}u_{i2} + ... + u_{it}.
\end{aligned}
$$

Substituting

$$\varphi_{NT}^{t} = 1 + o(T),$$

which holds by the binomial theorem, into (91) yields

$$z_{it}u_{it+p+1} = u_{i1} + ... + u_{it} + o(T). \qquad (92)$$

The last relationship enables to apply standard asymptotic results about AR(1) processes (see also Hamilton (1994)).

To obtain the limiting distribution of (90), next we derive asymptotic results for quantities (m), (k) and (l), defined above. First, note that the probability limit of (k) is zero, as $T, N \to \infty$. This can be seen by writing (k) as

$$
\begin{aligned}
(k) &\equiv \frac{1}{\sqrt{N}}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T-p-1} z_{it}(1-\varphi_{NT})a_i = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} a_i \frac{1}{T}\sum_{t=1}^{T-p-1} z_{it}\frac{c}{T\sqrt{N}} \\
&= \frac{1}{N}\sum_{i=1}^{N} a_i \frac{1}{T^2}\sum_{t=1}^{T-p-1} z_{it}.
\end{aligned}
$$

Taking first the limit $T \to \infty$ gives

$$\frac{1}{T^2} \sum_{t=1}^{T-p-1} z_{it} \xrightarrow{p} 0,$$

and, thus,

$$\frac{1}{\sqrt{N}} \frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T-p-1} z_{it}(1 - \varphi_{NT})a_i \xrightarrow{p} 0. \tag{93}$$

As $T, N \to \infty$, (m) converges to a limiting distribution. This can be seen by writing

$$(m) \equiv \frac{1}{\sqrt{N}} \frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T-p-1} z_{it} u_{it+p+1} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T-p-1} z_{it} u_{it+p+1}. \tag{94}$$

As $T \to \infty$, we have

$$\frac{1}{T} \sum_{t=1}^{T-p-1} z_{it} u_{it+p+1} \xrightarrow{d} \frac{1}{2} \sigma^2 \left\{ [W_i(1)]^2 - 1 \right\}, \tag{95}$$

where $W(r)$ denotes the standard Wiener process at time r. $[W_i(1)]^2$ follows a chi-squared distribution with one degree of freedom, which means that $E\{[W_i(1)]^2\} = 1$ and $Var\{[W_i(1)]^2\} = 2$. Next, by taking the limit of (94) for $N \to \infty$ gives

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{2} \sigma^2 \left\{ [W_i(1)]^2 - 1 \right\} \xrightarrow{d} N(0, \frac{\sigma^4}{2}). \tag{96}$$

Finally, to find the limit of (l) write it, first, as

$$(l) \equiv \frac{1}{N} \frac{1}{T^2} \sum_{i=1}^{N} \sum_{t=1}^{T-p-1} z_{it} z_{it+p} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T^2} \sum_{t=1}^{T-p-1} z_{it} z_{it+p}. \tag{97}$$

Substituting the following representation of z_{it+p}:

$$z_{it+p} = \varphi_{NT}^{p} z_{it} + \sum_{k=1}^{p} \varphi_{NT}^{k-1} (1 - \varphi_{NT}) a_i + \sum_{k=1}^{p} \varphi_{NT}^{k-1} u_{it+(p-k-1)},$$

obtained under H_1: $\varphi > 0$, into into (97) and using $\varphi_{NT}^{t} = 1 + o(T)$ yields the following results:

$$
\begin{aligned}
\frac{1}{T^2}\sum_{t=1}^{T-p-1}\varphi_{NT}^{p} z_{it}^2 &= \frac{1}{T^2}\sum_{t=1}^{T-p-1} z_{it}^2 + o(T) \xrightarrow{d} \sigma^2\int_0^1 [W_i(r)]^2 dr, && (99)\\
&\frac{1}{T^2}\sum_{t=1}^{T-p-1} z_{it}\sum_{k=1}^{p}\varphi_{NT}^{k-1}(1-\varphi_{NT})a_i \xrightarrow{p} 0, && (100)\\
\frac{1}{T^2}\sum_{t=1}^{T-p-1} z_{it}\sum_{k=1}^{p}\varphi_{NT}^{k-1}u_{it+(p-k-1)} &= \frac{1}{T^2}\sum_{t=1}^{T-p-1} z_{it}\sum_{k=1}^{p} u_{it+(p-k-1)} + o(T) \xrightarrow{p} 0. && (101)
\end{aligned}
$$

Based on the results of equations (99), (100) and (101), we can show that (m) converges to the following quantity:

$$
\frac{1}{N}\sum_{i=1}^{N}\sigma^2\int_0^1 [W_i(r)]^2 dr \xrightarrow{p} -\frac{\sigma^2}{2}, \qquad (102)
$$

as $N \to \infty$ (see also Levin et al. (2002)).

The proof of the proposition follows immediately by using the results of equations (93), (96) and (102).

Proof of Theorem 3 The theorem can be proved following analogous steps to those of the proof of Theorem 1. First, write $\sqrt{N}(\hat{\varphi}_{FDIV} - \varphi_N)$ as

$$
\begin{aligned}
\sqrt{N}(\hat{\varphi}_{FDIV} - \varphi_N) &= \sqrt{N}\left(\frac{\sum_{i=1}^{N} y_{i-1}^{*\prime}\Pi_p^* y_i^*}{\sum_{i=1}^{N} y_{i-1}^{*\prime}\Pi_p^* y_{i-1}^*} - \varphi_N\right)\\
&= \sqrt{N}\left(\frac{\sum_{i=1}^{N} y_{i-1}^{*\prime}\Pi_p^*(\varphi_N y_{i-1} + (1-\varphi_N)\beta_i^* e^* + u_i^*)}{\sum_{i=1}^{N} y_{i-1}^{*\prime}\Pi_p^* y_{i-1}^*} - \varphi_N\right)\\
&= \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(1-\varphi_N)\beta_i^* y_{i-1}^{*\prime}\Pi_p^* e^* + \frac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i-1}^{*\prime}\Pi_p^* u_i^*}{\frac{1}{N}\sum_{i=1}^{N} y_{i-1}^{*\prime}\Pi_p^* y_{i-1}^*}.
\end{aligned}
$$

The proof of the theorem follows from the last relationship after using the following results:

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(1-\varphi_N)\beta_i^* y_{i-1}^{*\prime}\Pi_p^* e^* \xrightarrow{p} 0,$$
$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i-1}^{*\prime}\Pi_p^* u_i^* \xrightarrow{d} N(0, 2tr((A_{FDIV}\Theta)^2)),$$
$$\frac{1}{N}\sum_{i=1}^{N} y_{i-1}^{*\prime}\Pi_p^* y \xrightarrow{p} tr(\Lambda^{*\prime}\Pi_p^*\Lambda^*),$$

since $tr(\Lambda^{*\prime}\Pi_p^*\Theta) = 0$.

Proof of Corollary 3 Note that under the assumptions of Corollary 3 the variance-covariance matrix of the transformed error terms u_{it}^* is given as

$$E(u_i^* u_i^{*\prime}) = \begin{pmatrix} q & r & s & & & 0 \\ r & q & r & s & & \\ s & r & q & r & s & \\ & . & . & . & . & \\ & & . & . & . & . \\ 0 & & & s & r & q \end{pmatrix}, \tag{103}$$

where $q = (2+\theta^2) - 2\theta$, $r = 2\theta - (1+\theta^2)$ and $s = -\theta$. Substituting the following results

$$tr(\Lambda^{*\prime}\Pi_p^*\Lambda^*) = T - p - 3 \text{ and } tr((A_{FDIV})^2) = T - p - 2,$$

which holds for $\theta = 0$, into (28) gives (29). Equation (30) of the corollary, for $\theta \neq 0$, can be proved by substituting the following results into (28):

$$tr(\Lambda^{*\prime}\Pi_p^*\Lambda^*\Theta) = (T-4)\theta^2 - \theta + T - 4 \text{ and } tr((A_{FDIV}\Theta)^2) = P_1\theta^4 + P_2\theta^3 + P_3\theta^2 + P_2\theta + P_1,$$

where $P_1 = 2(T-3), P_2 = -2(2T-8)$ and $P_3 = 2(4T-15)$.

Proof of Theorem 4 To prove Theorem 4, first write vector y_{i-1} and its first difference Δy_i as

$$y_{i-1} = wy_{i0} + \Omega X\zeta_i + \Omega u_i,\ i = 1, 2, ..., N, \tag{104}$$

and

$$\Delta y_i = u_i + (\varphi_N - 1)y_{i-1} + X\zeta_i, \tag{105}$$

respectively, where Ω is defined by (45) and $\zeta_i = \begin{pmatrix} (1-\varphi_N)a_i + \varphi\beta_i \\ (1-\varphi_N)\beta_i \end{pmatrix}$. ζ_i can be written in more compact form as

$$\zeta_i = \frac{c}{\sqrt{N}}\mu_i + \beta_i e^*, \tag{106}$$

where $\frac{c}{\sqrt{N}} = (1-\varphi_N)$, $\mu_i = \begin{pmatrix} a_i - \beta_i \\ \beta_i \end{pmatrix}$ and $e^* = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Then, the theorem can be proved by following analogous steps to those of Theorem's 2 proof. That is, first write

$$\sqrt{N}\hat{\delta}_{WGT}\left(\hat{\varphi}_{WGT} - \frac{\hat{b}_{WGT}}{\hat{\delta}_{WGT}} - \varphi_N\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i-1}'Q^*u_i - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Delta y_i'\Phi_{p,WGT}\Delta y_i = (a^*)-(b^*), \tag{107}$$

where $(a^*) \equiv \frac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i-1}'Q^*u_i$ and $(b^*) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Delta y_i'\Phi_{p,WGT}\Delta y_i$. Writing$(a^*)$ as

$$(a^*) \equiv \frac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i-1}'Q^*u_i = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(wy_{i0} + \Omega X\zeta_i + \Omega u_i)'Q^*u_i,$$

it can be shown that the limiting distribution of (a^*) is given as

$$(a^*) \equiv \frac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i-1}'Q^*u_i \xrightarrow{d} N(tr(\Lambda'Q^*\Gamma) - ctr(F'Q^*\Gamma), V_{WGT,a^*}),$$

where V_{WGT,a^*} is the variance of the distribution. This result holds, since we have

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i0}w'Q^*u_i \xrightarrow{p} 0 \tag{108}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \zeta_i'X\Omega'Q^*u_i \xrightarrow{p} 0 \tag{109}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} u_i'\Omega Q^*u_i \xrightarrow{d} N(tr(\Lambda'Q^*\Gamma) - ctr(F'Q^*\Gamma), V_{WGT,a^*}) \tag{110}$$

The results given by equations (108) and (110) can be derived as before, for Theorem 2. To see why (109) holds, write

$$\begin{aligned}\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \zeta_i'X\Omega'Q^*u_i &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\frac{c}{\sqrt{N}}\mu_i + \beta_i e^*)'X\Omega'Q^*u_i \\ &= \frac{c}{N}\sum_{i=1}^{N}\mu_i'X\Omega'Q^*u_i + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\beta_i e^{*\prime}X\Omega'Q^*u_i\end{aligned}$$

and note the following results:

$$\frac{c}{N}\sum_{i=1}^{N}\mu_i' X\Omega' Q^* u_i \xrightarrow{p} 0 \tag{111}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\beta_i e^{*\prime} X\Omega' Q^* u_i = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\beta_i e^{*\prime} X\Lambda' Q^* u_i = 0, \tag{112}$$

since $e^{*\prime} X\Lambda' Q^* = (0, ..., 0)$.

To derive the limiting distribution of quantity (b^*), we can write

$$\begin{aligned}
(b^*) &\equiv \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Delta y_i'\Phi_{p,WGT}\Delta y_i \\
&= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(u_i + (\varphi_N - 1)y_{i-1} + X\zeta_i)'\Phi_{p,WGT}(u_i + (\varphi_N - 1)y_{i-1} + X\zeta_i) \\
&= \frac{1}{\sqrt{N}}\sum_{i=1}^{N} u_i'\Phi_{p,WGT}u_i + u_i'\Phi_{p,WGT}y_{i-1}(\varphi_N - 1) + u_i'\Phi_{p,WGT}X\zeta_i \\
&\quad + (\varphi_N - 1)y_{i-1}'\Phi_{p,WGT}u_i + (\varphi_N - 1)^2 y_{i-1}'\Phi_{p,WGT}y_{i-1} + (\varphi_N - 1)y_{i-1}'\Phi_{p,WGT}X\zeta_i \\
&\quad + \zeta_i' X'\Phi_{p,WGT}u_i + \zeta_i' X'\Phi_{p,WGT}y_{i-1}(\varphi_N - 1) + \zeta_i' X'\Phi_{p,WGT}X\zeta_i.
\end{aligned}$$

and use the following results:

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}u_i'\Phi_{p,WGT}u_i \xrightarrow{d} N(tr(\Phi_{p,WGT}\Gamma), V_{WGT,1}) \tag{113}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}u_i'\Phi_{p,WGT}y_{i-1}(\varphi_N - 1) \xrightarrow{p} -ctr(\Phi_{p,WGT}\Lambda\Gamma) \tag{114}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}u_i'\Phi_{p,WGT}X\zeta_i \xrightarrow{p} 0 \tag{115}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\varphi_N - 1)y_{i-1}'\Phi_{p,WGT}u_i \xrightarrow{p} -ctr(\Lambda'\Phi_{p,WGT}\Gamma) \tag{116}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\varphi_N - 1)^2 y_{i-1}'\Phi_{p,WGT}y_{i-1} \xrightarrow{p} 0 \tag{117}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\varphi_N - 1)y_{i-1}'\Phi_{p,WGT}X\zeta_i \xrightarrow{p} -cE(\beta_i^2)e^{*\prime}X'\Lambda'\Phi_{p,WGT}Xe^* \tag{118}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\zeta_i' X'\Phi_{p,WGT}u_i \xrightarrow{p} 0 \tag{119}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\zeta_i' X\Phi_{p,WGT}y_{i-1}(\varphi_N - 1) \xrightarrow{p} -cE(\beta_i^2)e^{*\prime}X'\Phi_{p,WGT}\Lambda X e^* \tag{120}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\zeta_i' X\Phi_{p,WGT}X\zeta_i \xrightarrow{p} ctr(X'\Phi_{p,WGT}Xe^*E(\mu_i'\beta_i)) + ctr(e^{*\prime}X'\Phi_{p,WGT}XE(\mu_i\beta_i))$$

Using the results of the above relationships gives

$$\sqrt{N}\hat{\delta}_{WGT}\left(\hat{\varphi}_{WGT} - \frac{\hat{b}_{WGT}}{\hat{\delta}_{WGT}} - 1\right) \xrightarrow{d} N(-c(d^* + E(\beta_i^2)g^*), V_{WGT}), \tag{121}$$

where

$$\begin{aligned} d^* &= tr(\Lambda' Q^*\Lambda\Gamma) + tr(F'Q^*\Gamma) - tr(\Phi_{p,WGT}\Lambda\Gamma) - tr(\Lambda'\Phi_{p,WGT}\Gamma), \\ g^* &= tr(X'\Phi_{p,WGT}Xe^*\tilde{e}') + tr(e^{*\prime}X'\Phi_{p,WGT}X\tilde{e}) - e^{*\prime}X'\Lambda'\Phi_{p,WGT}Xe^* - e^{*\prime}X'\Phi_{p,WGT}\Lambda X e^*, \end{aligned}$$

and $E(\mu_i\beta_i) = E(\beta_i^2)\tilde{e}$, where $\tilde{e} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$. As before, note that the variance of the above limiting distribution, denoted as V_{WGT}, is the same with that under hypothesis H_0: $c = 0$, and it is given as $V_{WGT} = 2tr((A_{WGT}\Gamma)^2)$ (see 34).

Relationship (34) of Theorem 5 can be proved by substituting

$$tr(X'\Phi_{p,WGT}Xe^*\tilde{e}') = e^{*\prime}X'\Lambda'\Phi_{p,WGT}Xe^* \text{ and } tr(e^{*\prime}X'\Phi_{p,WGT}X\tilde{e}) = e^{*\prime}X'\Phi_{p,WGT}\Lambda X e^*$$

into quantities (d^*) and (g^*). Then, (121) implies

$$\begin{aligned} &\sqrt{N}\hat{\delta}_{WGT}V_{WGT}^{-1/2}\left(\hat{\varphi}_{WGT} - \frac{\hat{b}_{WGT}}{\hat{\delta}_{WGT}} - 1\right) \xrightarrow{d} \\ &N(-c\left(tr(\Lambda'Q^*\Lambda\Gamma) + tr(F'Q^*\Gamma) - tr(\Phi_{p,WGT}\Lambda\Gamma) - tr(\Lambda'\Phi_{p,WGT}\Gamma)\right)V_{WGT}^{-1/2}, 1). \end{aligned}$$

Proof of Corollary 4 The proof of the corollary follows by following analogous steps to those of the proof of Corollary 3 and using the following relationship:

$$tr(\Lambda'Q^*\Lambda\Gamma) + tr(F'Q^*\Gamma) = tr(\Phi_{p,WGT}\Lambda\Gamma) + tr(\Lambda'\Phi_{p,WGT}\Gamma).$$

Proof of Theorem 5 To prove the theorem, we will employ following relationships:

$$z_i = \varphi_N z_{i-1} + X\zeta_i + u_i, \quad i = 1, 2, ..., N \tag{122}$$

$$z_{i-1} = \Omega X\zeta_i + \Omega u_i + (w - e)y_{i0}, \tag{123}$$

$$\text{and} \quad \Delta z_i = (\varphi_N - 1)z_{i-1} + X\zeta_i + u_i, \tag{124}$$

which hold under both the null H_0: $c = 0$ and alternative H_0: $c > 0$ hypotheses. Next, we derive the results of the theorem for $c = 0$ and, then, for $c > 0$.

To derive the limiting distribution of the FOD test statistic under the null hypothesis H_0: $c = 0$, we first need to derive the inconsistency of LS estimator $\hat{\varphi}_{FOD}$ To this end, write

$$\hat{\varphi}_{FOD} - 1 = \frac{\frac{1}{N}\sum_{i=1}^{N} z_i' B' A \Delta z_i}{\frac{1}{N}\sum_{i=1}^{N} z_i' B' B z_i} = \frac{(h^*)}{(g^*)},$$

where $(h^*) \equiv \frac{1}{N}\sum_{i=1}^{N} z_i' B' A \Delta z_i$ and $(g^*) \equiv \frac{1}{N}\sum_{i=1}^{N} z_i' B' B z_i$. (h^*) and (g^*) can be written as

$$\begin{aligned}
(h^*) &\equiv \frac{1}{N}\sum_{i=1}^{N} z_i' B' A \Delta z_i = \frac{1}{N}\sum_{i=1}^{N}(z_{i-1}' + \beta_i e' + u_{i'})B'A(\beta_i e + u_i) \\
&= \frac{1}{N}\sum_{i=1}^{N}(u_i'\Lambda' + \beta_i e'\Lambda' + \beta_i e' + u_{i'})B'A(\beta_i e + u_i) \\
&= \frac{1}{N}\sum_{i=1}^{N}(u_i'(\Lambda' + I_T) + \beta_i \tau')B'A(\beta_i e + u_i) \\
&= \frac{1}{N}\sum_{i=1}^{N} u_i'(\Lambda' + I_T)B'Au_i
\end{aligned}$$

and

$$\begin{aligned}
(g^*) &\equiv \frac{1}{N}\sum_{i=1}^{N} z_i' B' B z_i = \frac{1}{N}\sum_{i=1}^{N}(z_{i-1}' + \beta_i e' + u_{i'})B'B(z_{i-1} + \beta_i e + u_i) \\
&= \frac{1}{N}\sum_{i=1}^{N}(u_i'(\Lambda' + I_T) + \beta_i \tau')B'B((\Lambda + I_T)u_i + \beta_i \tau) \\
&= \frac{1}{N}\sum_{i=1}^{N} u_i'(\Lambda' + I_T)B'B(\Lambda + I_T)u_i,
\end{aligned}$$

respectively, since $(\Lambda + I_T)e = \tau$, $\tau' B' = 0_{1XT}$ and $B'Ae = 0_{TX1}$. By the KWLLN, the following results hold:

$$\frac{1}{N}\sum_{i=1}^{N} u_i'(\Lambda' + I_T)B'Au_i = \frac{1}{N}\sum_{i=1}^{N} u_i'\Xi u_i \xrightarrow{p} tr(\Xi\Gamma). \tag{125}$$

since $(\Lambda' + I_T)B'A = \Xi$, and

$$\frac{1}{N}\sum_{i=1}^{N} u_i'(\Lambda' + I_T)B'B(\Lambda + I_T)u_i \xrightarrow{p} tr((\Lambda' + I_T)B'B(\Lambda + I_T)\Gamma). \tag{126}$$

The last two relationships imply that the inconsistency of $\hat{\varphi}_{FOD}$ is given as

$$p\lim_{N\longrightarrow\infty}(\hat{\varphi}_{FOD} - 1) = \frac{tr(\Xi\Gamma)}{tr((\Lambda' + I_T)B'B(\Lambda + I_T)\Gamma)}.$$

From this relationship, it can be easily seen that $\hat{\varphi}_{FOD}$ becomes unbiased, if $tr(\Xi\Gamma) = 0$. This happens when $\Gamma = \sigma^2 I_T$, i.e., error terms u_{it} are both homoscedastic and serially uncorrelated (see also fn 9).

The limiting distribution of the corrected for its inconsistency estimator $\hat{\varphi}_{FOD}$ under null hypothesis H_0: $c = 0$ can be written by writing

$$\begin{aligned}
&\sqrt{N}\hat{\delta}_{FOD}\left(\hat{\varphi}_{FOD} - 1 - \frac{\hat{b}_{FOD}}{\hat{\delta}_{FOD}}\right) \\
= \; &\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} z_i'B'Bz_i\right)\left(1 + \frac{\frac{1}{N}\sum_{i=1}^{N} z_i'B'A\Delta z_i}{\frac{1}{N}\sum_{i=1}^{N} z_i'B'Bz_i} - 1 - \frac{\frac{1}{N}\sum_{i=1}^{N} \Delta z_i'\Phi_{p,FOD}\Delta z_i}{\frac{1}{N}\sum_{i=1}^{N} z_i'B'Bz_i}\right) \\
= \; &\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} u_i'(\Lambda' + I_T)B'Au_i - \frac{1}{N}\sum_{i=1}^{N} \Delta z_i'\Phi_{p,FOD}\Delta z_i\right) \\
= \; &\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} \Delta z_i'\left((\Lambda' + I_T)B'A - \Phi_{p,FOD}\right)\Delta z_i\right) \\
= \; &\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \Delta z_i'\left(\Xi - \Phi_{p,FOD}\right)\Delta z_i,
\end{aligned} \tag{127}$$

since

$$\Delta z_i'(\Lambda' + I_T)B'A\Delta z_i = (\beta_i e' + u_i')(\Lambda' + I_T)B'A(\beta_i e + u_i) = u_i'(\Lambda' + I_T)B'Au_i.$$

The result of the theorem for the case $c = 0$ can be obtained using

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Delta z_i' (\Xi - \Phi_{p,FOD}) \Delta z_i \xrightarrow{d} N(0, 2tr(A_{FOD}^2)). \tag{128}$$

For the case $c > 0$, the proof of the theorem follows analogous steps to those of the proof of Theorem 4. This can be easily seen by applying the arguments of the proof of Theorem 4 to the following quantities:

$$\begin{aligned} \sqrt{N}\hat{\delta}_{FOD}\left(\hat{\varphi}_{FOD} - \varphi_N - \frac{\hat{b}_{FOD}}{\hat{\delta}_{FOD}}\right) &= \\ &= \sqrt{N}\hat{\delta}_{FOD}\left(1 + \frac{\frac{1}{N}\sum_{i=1}^{N} z_i' B' A \Delta z_i}{\frac{1}{N}\sum_{i=1}^{N} z_i' B' B z_i} - \varphi_N - \frac{\frac{1}{N}\sum_{i=1}^{N} \Delta z_i' \Phi_{p,FOD} \Delta z_i}{\frac{1}{N}\sum_{i=1}^{N} z_i' B' B z_i}\right) \\ &= \frac{c}{N}\sum_{i=1}^{N} z_i' B' B z_i + \frac{1}{\sqrt{N}}\sum_{i=1}^{N} z_i' B' A \Delta z_i - \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \Delta z_i' \Phi_{p,FOD} \Delta z_i. \end{aligned} \tag{129}$$

The complete the proof, we need the following results:

$$\begin{aligned} tr(\Xi) &= 0 \;\; \text{and} \; tr(\Lambda' B' A) = -tr(B' A) \\ e'\Xi &= 0_{1XT} \;\; \text{and} \;\; \Xi e = 0_{TX1} \\ B' A X \tilde{e} &= 0_{TX1} \\ e^{*\prime} X' \Lambda' B' A \Lambda X e^* &= e^{*\prime} X' \Lambda' B' A X \tilde{e} \\ e^{*\prime} X' B' A \Lambda X e^* &= e^{*\prime} X' B' A X \tilde{e} \\ e^{*\prime} X' \Phi_{p,FOD} \Lambda X e^* &= e^{*\prime} X' \Phi_{p,FOD} X \tilde{e} \\ e^{*\prime} X' \Lambda' \Phi_{p,FOD} X e^* &= \tilde{e}' X' \Phi_{p,FOD} X e^* \end{aligned}$$

Proof of Corollary 5 It follows immediately from the proof of Theorem 5.