On the Local Power of Fixed T Panel Unit Root Tests with Serially Correlated Errors

Karavias, Yiannis and Tzavalis, Elias

December 2012

On the local power of fixed T panel unit root tests with serially correlated errors.

Yiannis Karavias[a] and Elias Tzavalis[b]

a :School of Economics, University of Nottingham

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

Working Paper

December 2012

Abstract

Analytical asymptotic local power functions are employed to study the effects of general form short term serial correlation on fixed$-T$ panel data unit root tests. Two models are considered, one that has only individual intercepts and one that has both individual intercepts and individual trends. It is shown that tests based on IV estimators are more powerful in all cases examined. Evenmore, for the model with individual trends an IV based test is shown to have non-trivial local power at the natural $\sqrt{N}$ rate.

JEL classification: C22, C23

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

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, since they can be applied to short panels. Early contributions in this area include Sargan and Bhargava (1983), Breitung and Meyer (1994), Harris and Tzavalis (1999, 2004), Kruininger and Tzavalis (2002), Hahn and Kuersteiner (2002), Bond et al (2005), Kruiniger (2008), Hahn and Phillips (2010) and De Blander and Dhaene (2011). These papers derive the limiting distribution of the suggested tests under the null hypothesis of a unit root in all individual series of the panel. Despite the plethora of studies for the distribution of large-T panel unit root tests under local alternatives[1], only recently there has been some interest to do so in the literature for fixed-T tests.

Specifically, for the AR(1) model with individual intercepts, Bond et al. (2005) derive the limiting distributions, under local alternatives, of the Breitung and Meyer (1994) test, the Harris and Tzavalis (1999) test, the first differenced MLE test of Kruiniger (2002) and Hsiao et al. (2002) and the first differenced and system GMM of Arellano and Bond (1991) and Blundell and Bond (1998). Kruiniger (2008) also derives the distribution under local alternatives for the first differenced MLE in the same model. Madsen (2010) compares the local power functions of the pooled OLS test, the Breitung and Meyer (1994) test and the Harris and Tzavalis (1999) test under alternative assumptions about the initial conditions.

The main purpose of this paper is to derive analytically the limiting distribution of panel unit root tests allowing for serial correlation under local alternatives and, then, to study the asymptotic power properties of these tests. In this framework. This can shed light on how higher order dynamics, or serial correlation of the error term, can affect the local asymptotic power of panel unit root tests and, thus, to choose the best (in terms of power) testing procedure in practice. These effects are studied through Monte Carlo simulations, but not analytically (see De Blander and Dhaene (2011), for panels, and Schwert (1989), for single time series, respectively). The paper consider two first order autoregressive panel models: the first one has individual specific intercepts and the second one has individual specific intercepts and individual specific trends. Since the model with trends has not been considered in the fixed T literature, one of the aims of the paper is to discuss the "incidental trends" problem of Moon and Phillips (1999).

The present study makes several contributions. First, for the model with individual intercepts, the panel unit root tests of De Wachter et al. (2007) and Kruiniger and Tzavalis (2002) are reformulated and have their local power function analytically derived. De Wachter et al. (2007) propose a panel unit root test based on an IV estimator while Kruiniger and

[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).

Tzavalis (2002) propose a test based on the WG estimator. Analytical local power functions are derived for general forms of short term serial correlation up to order $T-2$. In the context of an MA(1) model for serial correlation it is shown that the IV test is always more powerful than the WG test although their behaviour is significantly different due to the the different ways that the two tests use the available moments. In both cases positive values of the moving average help power while negative values reduce it. In contrast to the IV test, the WG test becomes biased for large negative values of the parameter. The order of serial correlation has a significant impact on the IV test but not on the WG. Monte Carlo results show that the asymptotic theory provides very good small sample approximations.

Second, for the model with individual trends, the local power for the corresponding test of Kruiniger and Tzavalis (2002) is derived, denoted henceforth as WG*. When there is no serial correlation the test has trivial power as would be expected by the "incidental trends problem". When there is serial correlation, it is shown that the test has power in the natural root-N neighbourhood of unity. In an MA(1) context, positive serial correlation reduces the power of the test and negative increases it; this is the opposite of the case where the model contains only individual effects. An attempt is made to find more powerful tests for this case by proposing two new tests: a fixed-T version of the Breitung (2000) test, called henceforth FOD due to the forward orthogonal deviations matrix involved, and an extension of the De Wachter et al. (2007) test for the case of individual trends, called double difference IV test (DDIV) henceforth. It is shown that the fixed-T version of the Breitung (2000) test behaves as the WG* test having in general smaller power or bias, depending on the moving average parameter, and trivial power when there is no serial correlation. The DDIV test is shown to have non-trivial power even in the case when there is no serial correlation. In the MA(1) context it is shown to behave like the IV test of De Wachter et al. (2007). Monte Carlo experiments show that asymptotic theory provides moderate approximations for this case.

The paper is organized as follows: Section 2 introduces the models and the assumptions required for the derivation of asymptotic results. Section 3 derives the asymptotic local power functions and provides results on the behaviour of the tests. Section 4 compares the statistical properties of the two tests. Section 5 conducts a small Monte Carlo experiment confirming the analytical 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 models with individual effects:

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

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

where $y_i = (y_{i1}, ..., y_{iT})'$ and $y_i = (y_{i0}, ..., y_{iT-1})'$ are $(TX1)$ vectors, u_i is the error term and a_i and β_i are the individual specific coefficients of the deterministic components. The $(TX1)$ vector e has elements $e_t = 1$ for $t = 1...T$ and $\tau_t = t$ the time trend. The asymptotic distributions of the tests are derived under the assumptions:

Assumption 1: (1.1)$\{u_i\}$ is a sequence of independent normal random vectors with $E(u_i) = 0$, $E(u_i u_i') = \Gamma \equiv \gamma_{st}$ where Γ is of unknown form apart from $\gamma_{i,1T} = \gamma_{i,T1} = 0$ and that at least one $\gamma_{tt} \neq 0$ for $t = 1, ..., T$. (1.2) $\gamma_{tt} > 0$ for at least one $t = 1, ..., T$. (1.3) The $4+\delta-th$ population moments of Δy_i, $i = 1, ..., N$ are uniformly bounded i.e. 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. (1.4) $l'Var(vec(\Delta y_i \Delta y_i')l > 0$ for every $l \in R^{0.5T(T+1)}$ such that $l'l = 1$.(1.5) $Var(y_{i0}) < +\infty$.

Assumption 2: The following hold: $E(u_{it}a_i) = 0$, $E(u_{it}\beta_i) = 0$ and $E(u_{it}y_{i0}) = 0$ for $t = 1, ..., T$ and $i = 1, ..., N$.

Condition (1.1) restricts the order of serial correlation to be at most $T-2$. This condition can be strengthened to allow for smaller orders of serial correlation. Condition (1.2) imposes finite fourth moments on the initial conditions, the error terms and the individual effects. Along with conditions (1.3) and (1.4) they allow application of the Markov LLN and the Lindeberg -Levy CLT and ensure that all quantities in the denominators are non-zero.

The original WG and IV tests allow for heterogeneous disturbances across i but this assumption is trimmed so that tractable results may be obtained. For the same reasons, normal errors provide analytic formulas for the variances of the tests. Condition (1.1) ensures the existence of at least one moment condition free of correlation nuisance parameters but does not specify the true order of serial correlation. Define p the order of serial correlation assumed by the researcher and p^* the true order. As long as $p \geq p^*$ the limiting distribution of the test statistics is valid. Since inference in both tests is based on moments that are free of correlation parameters, choosing $p > p^*$ means selecting fewer than possible moments for inference. For a discussion on how to estimate the order of serial correlation see Hayakawa (2010).

Define $\varphi_N = 1 - \frac{c}{\sqrt{N}}$. Then the hypothesis of interest is:

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

$$H_1 \;:\; c > 0 \tag{4}$$

where c is the local to unity parameter. Assumption 2 is only required when $c > 0$.

The only assumption on the initial condition is (1.5). Bond et al. (2005), Kruiniger (2008) and Madsen (2010) assume covariance stationary initial conditions which means that $Var(y_{i0}) = \frac{\sigma^2}{1-\varphi_N^2}$. Then, as N diverges, $\varphi_N \to 1$ and $Var(y_{i0}) \to +\infty$. This assumption is not appropriate because it is not plausible that the variance of the initial condition increases with the number of cross section units, see also Moon et al. (2007). All tests in the paper are invariant to the initial condition. The IV, FOD and DDIV tests subtract the initial values of the individual series of the panel from their levels, across all units of the panel[2]. This is done for all time-series observations of the panel. The WG tests, for both models become invariant to the initial conditions of the panel by relying on the "within" transformation of its individual time series[3].

To study the asymptotic local power of the tests we employ a "slope" parameter, denoted as k, which is found in functions of the form

$$\Phi(z_a + ck)$$

where Φ is the standard normal cumulative distribution function and z_a the $\alpha - level$ percentile. Since Φ is strictly monotonic, for the same c, a larger k means greater power. Thus, it suffices to compare cases by comparing their k, instead of having a visual inspection of the power function. If k is positive then the test has non-trivial power, if it is zero it has power 0.05 and if it is negative the test is biased.

3 Asymptotic local power functions

This section presents all tests that are studied and derives their asymptotic local power functions. The first half is dedicated to model $M1$ and the second to model $M2$.

3.1 Individual intercepts

IV panel unit root test: De Wachter, Harris and Tzavalis (2007) propose a fixed-T panel unit root test based on the IV estimator where in a first step they subtract the initial observations from all series as in Breitung and Meyer (1994). The test exploits moments of

[2] This approach is suggested by Schmidt and Phillips (1992), for single time series, and Breitund and Meyer (1994).

[3] 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-Fuller (1979) 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).

the form:

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

and is based on the estimator:

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

where $z_{it} = y_{it} - y_{i0}$. These moments can be rewritten in matrix notation as:

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

where Π_p is a (TXT) matrix that selects the appropriate moments according to equation (5) and $z_{i-1} = y_{i-1} - y_{i0}e$. Π_p has ones in the *pth* diagonal and zeros everywhere else. The estimator in (6) 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 test is given in the following theorem:

Theorem 1 *Under Assumptions 1, 2, the assumption that the order of serial correlation is at most p and as* $N \to \infty$*:*

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

where

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

and $V_{IV} = \frac{2tr((A_{IV}\Gamma)^2)}{tr(\Lambda'\Pi_p\Lambda\Gamma)^2}$, $A_{IV} = \frac{1}{2}(\Lambda'\Pi_p + \Pi'_p\Lambda)$.

Theorem 1 nests both the null and the local alternative hypotheses. For $c = 0$ (9) presents the distribution of the test under the null as found by De Wachter, Harris and Tzavalis (2007). Result (10) shows explicitly how local power depends on a)the assumed order of serial correlation through matrix Π_p and b)on the form of serial correlation found in matrix Γ. The test of Breitung and Meyer (1994) can be seen as a special case of the IV test for $p = 0$. Also, as Bond et al. (2005) show, this can be also seen as a maximum likelihood estimator.

WG panel unit root test: The WG test transforms model (1) by removing the individual effects with the annihilator matrix Q and then relies on the inconsistent WG estimator $\hat{\varphi}_{WG}$

for inference. Define the annihilator matrix $Q = I_T - e(e'e)^{-1}e'$ where I_T is the (TXT) identity matrix. Then $\hat{\varphi}_{WG}$ is defined as:

$$\hat{\varphi}_{WG} = (\sum_{i=1}^{N} y'_{i-1}Qy_{i-1})^{-1}(\sum_{i=1}^{N} y'_{i-1}Qy_i) \tag{11}$$

The test and its asymptotic distribution under Assumption 1 and as $N \to \infty$:

$$\sqrt{N}V_O^{-\frac{1}{2}}\hat{\delta}(\hat{\varphi}_{WG} - 1 - \frac{\hat{b}_O}{\hat{\delta}_O}) \xrightarrow{d} N(0,1) \tag{12}$$

where:

$$\frac{\hat{b}_O}{\hat{\delta}_O} = \frac{vec(Q\Lambda)S(\frac{1}{N}\sum_{i=1}^{N} vec(\Delta y_i \Delta y_i'))}{\frac{1}{N}\sum_{i=1}^{N} y'_{i-1}Qy_{i-1}} \text{ and} \tag{13}$$

$$V_O = vec(Q\Lambda)'(I_{T^2} - S)Var(vec(\Delta y_i \Delta y_i'))(I_{T^2} - S)vec(Q\Lambda). \tag{14}$$

$\frac{\hat{b}_O}{\hat{\delta}_O}$ is an estimator of the asymptotic bias of $\hat{\varphi}_{WG}$ given as $\frac{tr(\Lambda'Q\Gamma)}{tr(\Lambda'Q\Lambda\Gamma)}$ since

$$\hat{b}_O = vec(Q\Lambda)S\left(\frac{1}{N}\sum_{i=1}^{N} vec(\Delta y_i \Delta y_i')\right) \xrightarrow{p} tr(\Lambda'Q\Gamma) \text{ and} \tag{15}$$

$$\hat{\delta}_O = \frac{1}{N}\sum_{i=1}^{N} y'_{i-1}Qy_{i-1} \xrightarrow{p} tr(\Lambda'Q\Lambda\Gamma) \tag{16}$$

where I_{T^2} is the (T^2XT^2) identity matrix and S is a (T^2XT^2) diagonal selection matrix with elements s_{ij} defined as $s_{(i-1)T+j,(i-1)T+j} = 1 - d(\gamma_{ji} = 0)$ with $i, j = 1, 2, ..., T$ and $d(.)$ is the Dirac function. In the above, $\frac{\hat{b}_O}{\hat{\delta}_O}$ is an estimator of the asymptotic bias of $\hat{\varphi}_{WG}$. Λ is a deterministic matrix given in the appendix. S is an interlayer matrix in $\hat{b}_O$ which selects only the elements of $\frac{1}{N}\sum_{i=1}^{N} vec(\Delta y_i \Delta y_i')$ which are nonzero, thus maintaining the consistency of $\hat{b}_O$ and on the same time avoiding the equality between $\frac{\hat{b}_O}{\hat{\delta}_O}$ and $\hat{\varphi}_{WG} - 1$ which would deprive the test of any variability. Notice in equations (13) and (14) that the bias correction terms affect both the mean and the variance of the test.

The reformulation avoids the (T^2XT^2) matrix S employed in the original test because (T^2XT^2) matrices are far more demanding in computational power even for moderate values of T. A new selection matrix $\Psi_{p,WG}$ is employed which is (TXT)-dimensional having in diagonals $\{-p, .., 0, ...p\}$ the corresponding elements of matrix $\Lambda'Q$ and zero everywhere else[4]. Then

$$tr(\Psi_{p,WG}\hat{\Gamma}) = vec(Q\Lambda)S\left(\frac{1}{N}\sum_{i=1}^{N} vec(\Delta y_i \Delta y_i')\right) \tag{17}$$

$$2tr((A_{WG}\Gamma)^2) = vec(Q\Lambda)'(I_{T^2} - S)Var(vec(\Delta y_i \Delta y_i'))(I_{T^2} - S)vec(Q\Lambda) \tag{18}$$

where

$$\hat{\Gamma} = \frac{1}{N}\sum_{i=1}^{N} \Delta y_i \Delta y_i', \tag{19}$$

a consistent estimator of Γ under the null because $\Delta y_i = u_i$ and $A_{WG} = \frac{1}{2}(\Lambda' Q + Q\Lambda - \Psi_{p,WG} - \Psi_{p,WG}')$. The following theorem provides the limiting distribution of the reformulated statistic for $c \geq 0$:

Theorem 2 *Under Assumptions 1, 2, the assumption that the order of serial correlation is at most p and as* $N \to \infty$*:*

$$\sqrt{N}\hat{V}_{WG}^{-\frac{1}{2}}\hat{\delta}(\hat{\varphi}_{WG} - 1 - \frac{\hat{b}}{\hat{\delta}}) \xrightarrow{d} N(-ck_{WG}, 1) \text{ as } N \to +\infty. \tag{20}$$

where

$$k_{WG} = \frac{tr(\Lambda' Q\Lambda\Gamma) + tr(F'Q\Gamma) - tr(\Psi_{p,WG}\Lambda\Gamma) - tr(\Lambda'\Psi_{p,WG}\Gamma)}{\sqrt{2tr((A_{WG}\Gamma)^2)}}. \tag{21}$$

and $\hat{\varphi}_{WG} = (\sum_{i=1}^{N} y_{i-1}' Q y_{i-1})^{-1}(\sum_{i=1}^{N} y_{i-1}' Q y_i)$, $\frac{\hat{b}}{\hat{\delta}} = \frac{tr(\Psi_{p,WG}\hat{\Gamma})}{\frac{1}{N}\sum_{i=1}^{N} y_{i,-1}' Q y_{i,-1}}$, $F = \frac{d\Omega}{d\varphi}\ |_{\varphi=1}$ *where* Ω *is given in the appendix. The variance is given by* $V_{WG} = 2tr((A_{WG}\Gamma)^2)$ *where* $A_{WG} = \frac{1}{2}(\Lambda' Q + Q\Lambda - \Psi_{p,WG} - \Psi_{p,WG}')$*. The proof is given in the appendix.*

The annihilator matrix Q and the inconsistency correction estimator based on $\Psi_{p,WG}$ complicate the local power function. Equation (21) shows that the k_{WG} depends on the quantities $tr(\Lambda' Q\Lambda\Gamma)$, $tr(F'Q\Gamma)$, $tr(\Psi_{p,WG}\Lambda\Gamma), tr(\Lambda'\Psi_{p,WG}\Gamma)$. The first two quantities come

[4] Assume $T = 3$ and serial correlation of the form $MA(1)$. Then $\Gamma = \begin{pmatrix} \sigma_u^2(1+\theta^2) & \sigma_u^2\theta & 0 \\ \sigma_u^2\theta & \sigma_u^2(1+\theta^2) & \sigma_u^2\theta \\ 0 & \sigma_u^2\theta & \sigma_u^2(1+\theta^2) \end{pmatrix}$,

$\Psi_p = \begin{pmatrix} -\frac{2}{3} & -\frac{1}{3} & 0 \\ \frac{1}{3} & -\frac{1}{3} & 0 \\ 0 & \frac{2}{3} & 0 \end{pmatrix}$ and $S = \begin{pmatrix} 1&0&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ 0&0&0&1&0&0&0&0&0 \\ 0&0&0&0&1&0&0&0&0 \\ 0&0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&1&0 \\ 0&0&0&0&0&0&0&0&1 \end{pmatrix}$

from the annihilator matrix Q and the last two come from the selection matrix $\Psi_{p,WG}$. For $p=0$ the selection matrix mean effect disappears as $tr(\Psi_{p,WG}\Lambda\Gamma) = tr(\Lambda'\Psi_{p,WG}\Gamma) = 0$.

Connection to the large T literature As both panel dimensions increase, to derive an asymptotic distribution define

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

Corollary 1 *Under Assumptions 1,2 and as $T, N \to \infty$ jointly/$(\sqrt{N}/T) \to 0$:*

$$a)\;\; T\sqrt{N}(\hat{\varphi}_{IV} - 1)(\sqrt{2})^{-1} \xrightarrow{d} N(-c\frac{1}{\sqrt{2}}, 1), \tag{22}$$

$$b)\;\; T\hat{\delta}\sqrt{N}(\hat{\varphi}_{WG} - 1 - \frac{\hat{b}}{\hat{\delta}})(3)^{-1} \xrightarrow{d} N(0,1).$$

The proof is given in the appendix.

Corollary 1 shows the asymptotic distributions under the null and the local alternatives for the appropriately scaled IV and WG test statistics. Under the null hypothesis convergence is joint for both cases, see e.g. (Hahn and Kuersteiner (2002), De Wachter et al. (2007) and Harris and Tzavalis (1999). Under the local alternatives, joint convergence requires the additional assumption that $\frac{\sqrt{N}}{T} \to 0$ (see Levin et al. (2002) and Moon and Perron (2004)). This corollary applies for every fixed p and any form of short term serial correlation, which means that in an asymptotic framework, short term serial correlation does not affect the limiting distribution. This was already expected, see e.g. Moon and Perron (2008). For $c = 0$ result a) coincides with that found by De Wachter et al. (2007). The numerator only bias corrected WG group test has also been proposed by Moon and Perron (2004). To see the correspondence, it can be rewritten under the null as

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

where, assuming no serial correlation, $V_{WG} = 2\sigma^4 tr((A_{WG})^2)$ and $\delta = \sigma^2 tr(\Lambda' Q \Lambda)$. Accordingly scaling by T :

$$T\sqrt{N}(\hat{\varphi}_{WG} - 1 - \frac{\hat{b}}{\hat{\delta}}) \xrightarrow{d} N(0, T^2 V_{WG}\delta^2).$$

Using similar arguments as above,it can be shown that as $N, T \to \infty$ jointly

$$T\sqrt{N}(\hat{\varphi}_{WG} - 1 - \frac{\hat{b}}{\hat{\delta}}) \xrightarrow{d} N(0, 3).$$

Moon and Perron (2008) derive the last result.

Table 1 compares the long T version of the IV test with the tests found in Moon et al. (2007), Moon and Perron (2008) and Harris et al. (2010), assuming homogeneous alternatives i.e. $c_i = c$ and thus $E(c_i) = c$. The IV test has the maximum possible power which is equal to that of the common point-optimal test of Moon et al. (2007). The large T version of the Harris and Tzavalis (1999) test coincides with the LLC test as can be seen from Madsen (2010). The WG test has trivial power in this case.

Table 1: Slopes for large T tests.

Test	k
IV	$\frac{1}{\sqrt{2}}$
MPP	$\frac{1}{\sqrt{2}}$
LLC/HT	$\frac{3}{2}\sqrt{\frac{5}{51}}$
SGLS	$\frac{1}{\sqrt{3}}$
IPS	0.282
WG	0

3.2 Incidental trends

WG* panel unit root test: Kruiniger and Tzavalis (2002) propose a version of the WG test that allows for incidental trends. Consider an augmented annihilator matrix $Q^* = I_T - X(X'X)^{-1}X'$ where $X = [e, \tau]$. Multiplying $M2$ with Q^* wipes off both individual effects and incidental trends. A new problem in this case is that under the null, the covariance matrix estimator $\hat{\Gamma}$ is no longer consistent:

$$\Delta y_i = \beta_i e + u_i, \; i = 1, ..., N.$$

leading to

$$\frac{1}{N}\sum_{i=1}^{N} \Delta y_i \Delta y_i' \rightarrow \Gamma + E(\beta_i^2)ee'.$$

To remove the nuisance parameters a selection matrix (algebraically equivalent to the original selection matrix of Kruiniger and Tzavalis (2002)) is applied. Define the matrix M with elements $m_{ts} = 0$ if $\gamma_{ts} \neq 0$ and $m_{ts} = 1$ if $\gamma_{ts} = 0$. Then, $tr(M\Gamma) = 0$ and thus

$$\frac{1}{tr(Mee')N}\sum_{i=1}^{N} \Delta y_i' M \Delta y_i \rightarrow E(\beta_i^2). \tag{23}$$

Result (23) means that the selection matrix $\Phi_p^{WG} = \Psi_{p,WG}^* - \frac{tr(\Lambda' Q^* M)}{e'Me} M$ where $\Psi_{p,WG}^*$ is a (TXT) matrix having in diagonals $\{-p, .., 0, ...p\}$ the corresponding elements of matrix $\Lambda' Q^*$ and zero everywhere else, has the property $tr(\Phi_p^{WG} ee') = 0$ and leads to the consistent estimator

$$tr(\Phi_p^{WG}\hat{\Gamma}) \rightarrow tr(\Lambda' Q^* \Gamma) \tag{24}$$

Theorem 3 *For model $M2$ under Condition 1, the assumption that the order of serial correlation is at most p and as $N \rightarrow \infty$:*

$$\sqrt{N}\hat{V}_{WG}^{*-\frac{1}{2}}\hat{\delta}^*(\hat{\varphi}_{WG}^* - 1 - \frac{\hat{b}^*}{\hat{\delta}^*}) \rightarrow N(-ck_{WG}^*, 1) \text{ as } N \rightarrow +\infty. \tag{25}$$

where

$$k_{WG}^* = \frac{tr(\Lambda' Q^* \Gamma) + tr(F' Q^* \Gamma) - tr(\Phi_p^{WG} \Lambda \Gamma) - tr(\Lambda' \Phi_p^{WG} \Gamma)}{2tr((A_{WG}^* \Gamma)^2)}, \tag{26}$$

$\hat{\varphi}_{WG}^* = (\sum_{i=1}^{N} y_{i-1}' Q^* y_{i-1})^{-1}(\sum_{i=1}^{N} y_{i-1}' Q^* y_i)$, $\frac{\hat{b}^*}{\hat{\delta}^*} = \frac{tr(\Phi_p^{WG}\hat{\Gamma})}{\frac{1}{N}\sum_{i=1}^{N} y_{i,-1}' Q^* y_{i,-1}}$ *and* $A_{WG}^* = \frac{1}{2}(\Lambda' Q^* + Q^* \Lambda - \Phi_p^{WG} - \Phi_p^{WG\prime})$. *The proof is given in the appendix.*

FOD panel unit root test: Breitung (2000) proposed an unbiased panel unit root test for $M2$ based on an appropriate transformation of the dependent and the independent variables. Moon et al. (2006) show analytically that its local power is zero at the natural rate of $T^{-1}N^{-1/2}$ and thus, that the incidental trends problem applies in this case as well. A fixed-T version of the test is proposed on the assumption that unbiasedness provides better power performance. The estimator $\hat{\varphi}_{FOD}$ equals that of Breitung (2000) plus 1. The dependent variable is transformed with the Helmert or forward orthogonal deviation transformation and thus the name of the test. The method is as follows: In a first step subtract the initial observations from all series as in the panel IV test of DWHT. Then, by multiplying Δz_i with matrix A and z_i with matrix B:

$$E(z_i' B' A \Delta z_i) = 0, \tag{27}$$

where A and B are $(T-1)XT$ defined as

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

where

$$E = \begin{pmatrix} \sqrt{\frac{T-2}{T-1}} & & & 0 \\ & \sqrt{\frac{T-3}{T-2}} & & \\ & & \ddots & \\ 0 & & & \sqrt{\frac{1}{2}} \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & -\frac{1}{T-1} & \cdots & \cdots & \cdots & \cdots & -\frac{1}{T-1} \\ 0 & 1 & -\frac{1}{T-2} & \cdots & \cdots & \cdots & -\frac{1}{T-2} \\ \vdots & & \ddots & & & & \vdots \\ \vdots & & & \ddots & & & \vdots \\ \vdots & & & & 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \cdots & \cdots & \cdots & 0 & 1 & -1 \end{pmatrix}$$

where E is $(T-2)X(T-1)$, Σ is $(T-1)XT$ and $\tau_{T-2} = \begin{pmatrix} 1 \\ 2 \\ \vdots \\ T-2 \end{pmatrix}$. Based on the above, define the FOD estimator as

$$\hat{\phi}_{FOD} = 1 + \frac{\sum_{i=1}^{N} z_i' B' A \Delta z_i}{\sum_{i=1}^{N} z_i' B' B z_i}. \tag{28}$$

The following theorem derives its asymptotic distribution.

Theorem 4 *For model M2 under Condition 1, the assumption that the order of serial correlation is at most p and as $N \to \infty$:*

$$\sqrt{N}\hat{V}_{FOD}^{-\frac{1}{2}}\hat{\delta}_{FOD}(\hat{\varphi}_{FOD} - 1 - \frac{\hat{b}_{FOD}}{\hat{\delta}_{FOD}}) \to N(-ck_{FOD}, 1)\ , \tag{29}$$

where

$$k_{FOD} = \frac{tr(\Lambda' B' A \Lambda \Gamma) + tr(B' A \Lambda \Gamma) + tr(\Lambda' B' A \Gamma) + tr(F' B' A \Gamma) - tr(\Lambda' \Phi_p^{FOD} \Gamma) - tr(\Phi_p^{FOD} \Lambda \Gamma)}{2tr((A_{FOD}\Gamma)^2)}. \tag{30}$$

and $\frac{\hat{b}^*}{\hat{\delta}^*} = \frac{tr(\Phi_p^{FOD}\hat{\Gamma})}{\frac{1}{N}\sum_{i=1}^{N} z_i' B' B z_i}$, $\Phi_p^{FOD} = \Psi_{p,FOD} - \frac{tr(\Xi M)}{e' M e} M$ *where* $\Psi_{p,FOD}$ *is a* (TXT) *matrix having in diagonals* $\{-p, .., 0, ...p\}$ *the corresponding elements of matrix* Ξ *and zero everywhere else.* $\Xi = \Lambda' B' A + B' A$*. The variance is given by* $V_{FOD} = 2tr((A_{FOD}\Gamma)^2)$ *where* $A_{FOD} = \frac{1}{2}(\Xi + \Xi' - \Phi_p^{FOD} - \Phi_p^{FOD\prime})$*. The proof is given in the appendix.*

The FOD test is unbiased when there is no serial correlation, as in the large T case, but is no longer unbiased when there is. It is then corrected for its bias in a similar way with the WG* test. The mean value in (27) applies only if $Var(u_i) = \sigma^2 I_T$ and thus, so does consistency. This can be seen in

$$p\lim_{N \to \infty}(\hat{\phi}_{FOD} - 1) = tr(\Xi\Gamma) \tag{31}$$

where $tr(\Xi\Gamma) = 0$ if and only if $\Gamma = \sigma^2 I_T$.

DDIV panel unit root test Han and Phillips (2010) have the intuition that there might be a maximum likelihood test based on second differences that evades the "incidental trends" problem. This motivates the proposition of the DDIV test which is a generalization of the

IV test to allow for incidental trends. Take first differences in model $M2$:

$$\Delta y_i = \varphi \Delta y_{i-1} + (1-\varphi)\beta_i e^* + \Delta u_i, \quad i = 1, ..., N. \tag{32}$$

where $y_i = (y_{i2}, ..., y_{iT})'$, $y_{i-1} = (y_{i1}, ..., y_{iT-1})'$, $y_{i-2} = (y_{i0}, ..., y_{iT-2})'$,$\mathrm{u}_i = (u_{i2}, ..., u_{iT})'$, $u_{i-1} = (u_{i1}, ..., u_{iT-1})'$ and $e^* = (1, 1, ..., 1)$ are $(T-1)X1$ vectors. From these series subtract the initial observation Δy_{i1} to find

$$y_i^* = \varphi y_{i-1}^* + (1-\varphi)a_i^* + u_i^*, \quad i = 1, ..., N. \tag{33}$$

where $y_i^* = \Delta y_i - \Delta y_{i1} e$, $y_{i-1}^* = \Delta y_{i-1} - \Delta y_{i1} e$ and $a_i^* = (\beta_i - \Delta y_{i1})$.Model (33) clearly shows that moments similar to (7) can be exploited to test the null hypothesis of a unit root.Specifically:

$$E(y_{i-1}^{*\prime} \Pi_p^* u_i^*) = 0 \tag{34}$$

where Π_p^* is a $(T-1)X(T-1)$ matrix with unities in its $p+1$ diagonal and zeros everywhere else.

Theorem 5 *For model $M2$ under Condition 1, the assumption that the order of serial correlation is at most p and as $N \to \infty$:*

$$\sqrt{N}(\hat{\varphi}_{IV}^* - 1)\hat{V}_{IV}^* \to N(-ck_{IV}^*, 1) \tag{35}$$

where

$$k_{IV}^* = \frac{tr(\Lambda^{*\prime} \Pi_p^* \Lambda^* \Theta)}{\sqrt{2tr((A_{IV}^* \Theta)^2)}} \tag{36}$$

and $\hat{\varphi}_{IV}^* = (\sum_{i=1}^{N} y_{i-1}^{*\prime} \Pi_p^* y_{i-1}^*)^{-1} (\sum_{i=1}^{N} y_{i-1}^{*\prime} \Pi_p^* y_i^*)$, $V^* = \frac{2tr((A_{IV}^* \Theta)^2)}{tr(\Lambda^{*\prime} \Pi_p^* \Lambda^* \Theta)^2}$, $A_{IV}^* = \frac{1}{2}(\Lambda^{*\prime} \Pi_p^* + \Pi_p^{*\prime} \Lambda^*)$. *$\Lambda^*$ is a $(T-1)X(T-1)$ version of Λ and $\Theta = 2\Gamma_1 - \Gamma_2 - \Gamma_2'$ where $\Gamma_1 = E(u_i u_i')$ and $\Gamma_2 = E(u_i u_{i-1}')$. The proof is given in the appendix.*

The incidental trends problem appears to be relevant in the fixed-T literature as the WG* and FOD tests have trivial local power when there is no serial correlation. However, serial correlation, depending on its form, provides evidence in favour of the null or the alternative hypothesis which results in tests with non-trivial power or a bias. The DDIV test is superior to the WG* and FOD tests since it has power even when there is no serial correlation.

A discussion on the bias correction: Breitung (2000) and the previous theorems show that tests based on consistent estimators are more powerful than tests based on inconsistent estimators. Furthermore, Moon and Perron (2008) find that a test based on an inconsistent estimator which is bias corrected only for its numerator is less powerful than when the estimator is bias corrected for both its numerator and its denominator. The WG* and FOD tests are bias corrected only for their numerator therefore other versions of these tests that correct for both the numerator and the denominator might be more powerful. This motivates the following proposition:

Proposition 1 *For model M2 under Assumptions 1, 2, the assumption that the order of serial correlation is zero and as $N \to \infty$:*

$$a) \quad \sqrt{N}\hat{V}_{HT}^{-\frac{1}{2}}\left(\hat{\varphi}^*_{WG} - 1 - \frac{tr(\Lambda' Q^*)}{tr(\Lambda' Q^* \Lambda)}\right) \quad \to \quad N(0,1) \ ,$$
$$b) \quad \sqrt{N}V_{FOD}^{*-\frac{1}{2}}(\hat{\phi}_{FOD} - 1) \quad \to \quad N(0,1) \ ,$$

where $V_{HT} = \frac{15(193T^2-728T+1147)}{112(T+2)^3(T-2)}$ *and* $V^*_{FOD} = \frac{2tr((A_{\Xi})^2)}{tr((\Lambda'+I_T)B'B(\Lambda+I_T))^2}$. *A proof is given in the appendix.*

The previous theorem derives the asymptotic local power of the WG* and FOD based tests having both their numerators and denominators bias corrected. This case cannot accommodate serial correlation in the way that the WG* and the FOD tests did because the method would result in an identity. Both tests have trivial local power and can be thought of as part of the incidental parameters problem. Result a) is first found by Harris and Tzavalis (1999). For an intercept only case see e.g. Madsen (2010).

4 Assuming MA(1) serial correlation

The results of the above section are very general to provide some intuition about the behaviour of tests. This section studies focuses on the simple and representative case of MA(1) errors.

Assumption 3: $\{u_{it}\}$ is generated as $u_{it} = v_{it} + \theta v_{it-1}$ with $\theta \neq -1$ and $v_{it} \sim NIID(0, \sigma_u^2)$.

4.1 Individual intercepts

The following two corollaries provide simplified results for the IV and WG tests.

Corollary 2 *Under assumptions 1,2 and 3, k_{IV} depends only on T, p and θ. For selected cases of p and θ, $k_{IV}(p, \theta)$ is given by:*

$$k_{IV}(0,0) = \sqrt{\frac{1}{2}(T^2 - T)}, \tag{37}$$

$$k_{IV}(1,0)=\sqrt{\frac{T^2}{2}-\frac{3T}{2}+1}, \tag{38}$$

$$k_{IV}(2,0)=\sqrt{\frac{T^2}{2}-\frac{5T}{2}+3}, \tag{39}$$

$$k_{IV}(3,0)=\sqrt{\frac{T^2}{2}-\frac{7T}{2}+6}, \tag{40}$$

$$k_{IV}(1,\theta)=\frac{D_{1,IV}\theta^2+D_{2,IV}\theta+D_{1,IV}}{\sqrt{R_{1,IV}\theta^4+R_{2,IV}\theta^3+R_{3,IV}\theta^2+R_{2,IV}\theta+R_{1,IV}}}, \tag{41}$$

where $D_{i,IV}$ and $R_{j,IV}$ for $i = 1,2$ $j = 1,2,3$, are functions of T and are given in the appendix. The dependence of k_{IV} on T is evident and thus supressed.

Relation (37) is the slope parameter of the original Breitung and Meyer (1994) test, found also by Bond et al. (2005) and Madsen (2010). Relations (41) and (38) coincide with those found by De Wachter, Harris and Tzavalis (2007).

Corollary 3 *Under assumptions 1,2 and 3, k_{WG} depends only on T, p and θ so for selected cases of p and θ, $k_{WG}(p,\theta)$ is given by:*

$$k_{WG}(0,0)=\frac{\sqrt{3}(T-1)}{\sqrt{T^2-2T-\frac{4}{T}+5}}, \tag{42}$$

$$k_{WG}(1,0)=\frac{\sqrt{3}(T^2-3T+2)}{T\sqrt{T^2-6T-\frac{24}{T}+\frac{12}{T^2}+17}}, \tag{43}$$

$$k_{WG}(2,0)=\frac{\sqrt{3}(T^2-5T+6)}{T\sqrt{T^2-10T-\frac{80}{T}+\frac{60}{T^2}+41}}, \tag{44}$$

$$k_{WG}(3,0)=\frac{\sqrt{3}(T^2-7T+12)}{T\sqrt{T^2-14T-\frac{196}{T}+\frac{192}{T^2}+77}}, \tag{45}$$

$$k_{WG}(1,\theta)=\frac{(T-2)(T\theta^2-\theta^2+3T\theta-7\theta+T-1)}{2T\sqrt{R_{1,WG}\theta^4+R_{2,WG}\theta^3+R_{3,WG}\theta^2+R_{2,WG}\theta+R_{1,WG}}}, \tag{46}$$

where $R_{1,WG}, R_{2,WG}$ and $R_{3,WG}$ are functions of T defined in the appendix.

The following two graphs show how serial correlation affects the slope parameters of the IV and WG tests. Assuming MA(1) type of serial correlation, figure 1 shows the slope parameter k_{IV} for $\theta \in \{-0.9,-0.5,0,0.5,0.9\}$.When considering $\theta = 0$ then also $p = 0$ which means that for $\theta = 0$ there is one more moment available.

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

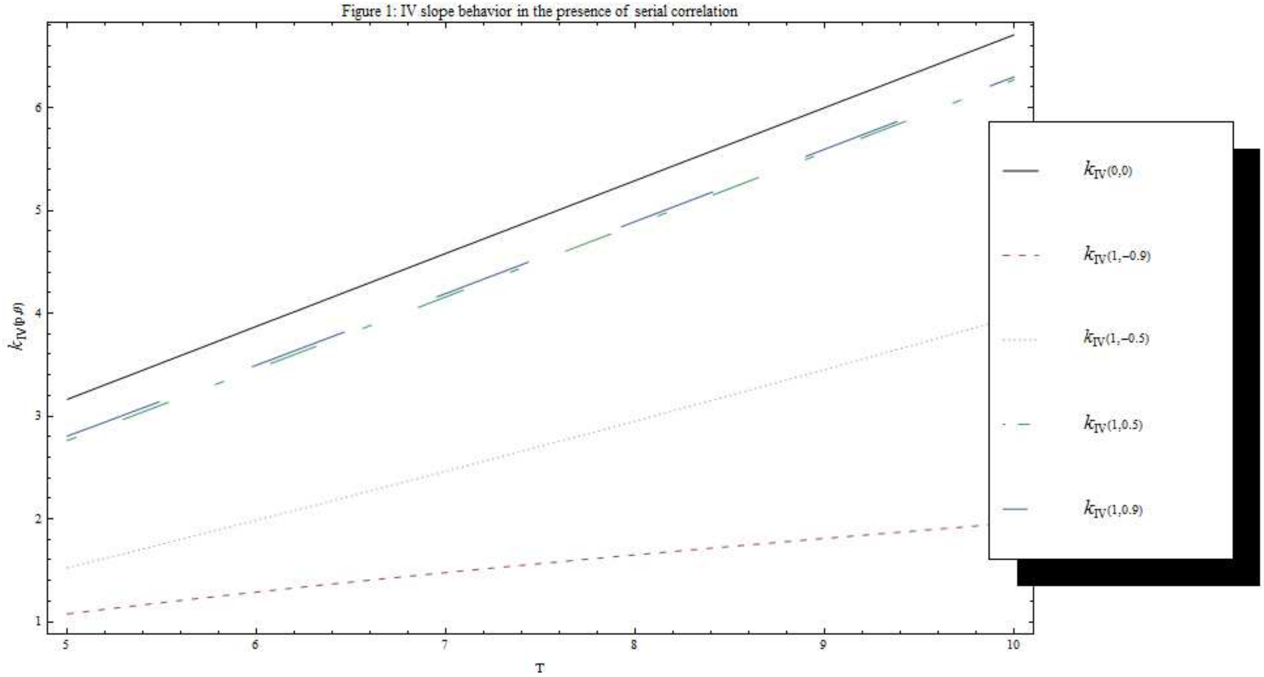
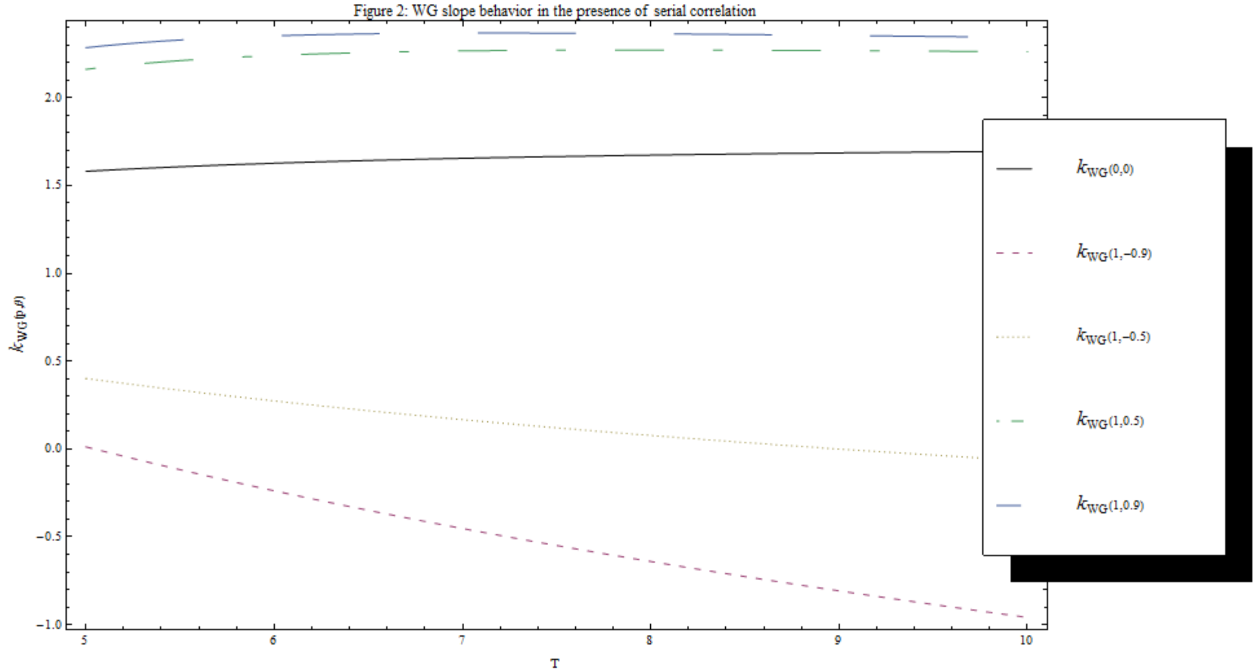


Figure 2: WG slope behavior in the presence of serial correlation



The IV test is most powerful when $\theta = 0$; if $\theta \neq 0$ one moment is lost so is power. Positive values of θ result in more power than negative values and T increases power in all cases. The effect of θ on the WG test slope is qualitatively the same as in the IV test but the effects are more intense. For positive θ the test has more power than for $\theta = 0$ and for θ negative it even becomes biased, something that never happens in the IV test. This happens because the bias correction affects the slope through $tr(\Psi_{p,WG} t\Lambda\Gamma) + tr(\Lambda'\Psi_{p,WG}\Gamma)$. Since it is subtracted, large negative values increase the slope and thus the power of the test. For $\theta < 0$ it takes positive values and thus reduces the power of the test. As T increases this effect becomes stronger. For $\theta > 0$, $tr(\Psi_{p,WG} t\Lambda\Gamma) + tr(\Lambda'\Psi_{p,WG}\Gamma)$ it is negative and thus moves the limiting distribution towards the critical region.

4.2 Individual trends

The following corolaries correspond to the tests WG*, FOD and DDIV. Complexity of the slope parameteres makes analytical formulas unavailable for the cases of WG* and FOD*.

Corollary 4 *Under assumptions 1,2 and 3 the following result holds:*

$$k^*_{WG}(p,0) = 0 \quad \textit{for } p = 0,1,2,...,T-2. \tag{47}$$
$$k^*_{WG}(1,\theta) \neq 0 \ \textit{for } \theta \neq 0 \tag{48}$$

A proof is given in the appendix.

Corollary 5 *Under assumptions 1,2 and 3 the following result holds:*

$$k_{FOD}(p,0) = 0 \quad \textit{for } p = 0,1,2,...,T-2. \tag{49}$$
$$k_{FOD}(1,\theta) \neq 0 \ \textit{for } \theta \neq 0 \tag{50}$$

A proof is given in the appendix.

Corollary 6 *Under assumptions 1,2 and 3, k^*_{IV} depends only on T, p and θ so for selected cases of p and θ, $k_{IV}(p,\theta)$ is given by:*

$$k_{DDIV}(p,0) = \frac{T-p-3}{\sqrt{2(T-p-2)}} \tag{51}$$

$$k_{DDIV}(1,\theta) = \frac{(T-4)\theta^2 - \theta + T - 4}{\sqrt{2(P_1\theta^4 + P_2\theta^3 + P_3\theta^2 + P_2\theta + P_1)}} \tag{52}$$

where polynomials $P_1, P_2,$ and P_3 are given in the appendix. A proof is given also in the appendix.

Results (47) and (49) signify the incidental trends problem in the fixed T literature. But from relation (48) it is clear that the nature of serial correlation gives power or bias to the test. In this case, specific functions for the trace quantities are extremely difficult to derive except for the DDIV test.

Table 2: Slope parameter values.

T=7					
$p\backslash\theta$	-0.5	-0.9	0	0.9	0.5
k^*_{WG}	0.466	0.694	0	-0.248	-0.212
k_{FOD}	0.110	0.148	0	-0.073	-0.062
k_{DDIV}	0.896	0.862	1.264	1.179	1.186
T=10					
$p\backslash\theta$	-0.9	-0.5	0	0.5	0.9
k^*_{WG}	1.042	0.645	0	-0.216	-0.248
k_{FOD}	0.151	0.110	0	-0.047	-0.054
k_{DDIV}	1.160	1.229	1.750	1.989	2.008

Table 2 contains values of the slope parameters for the WG*, FOD and DDIV tests. The "incidental trends" problem is evident for $\theta = 0$. Negative values of the moving average parameter result in non-trivial power while positive result in bias for tests WG* and FOD and the opposite for the DDIV test. This means that serial correlation affects power in the opposite way than it did for tests that had only individual intercepts, except for the DDIV test.

It is easy to see that for $T \to \infty$, $k^*_{IV} = \frac{T-p-3}{T\sqrt{2(T-p-2)}} \to 0$, thus, in a large T, the incidental parameter problem remains. This is already known as Moon et al. (2007) derive the local power envelope for this case.

5 Simulation Results

This section presents some Monte Carlo results whose purpose is to show how good the asymptotic theory approximates the small sample results. Every experiment is conducted 5000 times. T=7 for $M1$ and 15 for $M2$. All nuisance parameters that do not appear in the above local power functions are a priori set to zero, such as: $a_i = 0$, $\beta_i = 0$, $y_{i0} = 0$. The local alternatives are set as $\varphi = 1 - c/\sqrt{N}$ for $N \in \{50, 100, 200, 300, 1000\}$ and $c \in \{0, 1\}$. The errors are generated according to assumption 3 with $\theta \in \{-0.5, 0, 0.5\}$ and $v_{it} \sim NIID(0, 1)$. The size is selected to be 0.05.

Table 3 shows that for $M1$ the approximation is very good, especially for the IV test. Both size and power are close to their theoretical values. Table 4 shows that for model $M2$ a larger N is required for the tests to have local power close to the predicted. This is also found by Moon et al. (2007) and can be attributed to the presence of more complicated deterministic elements. The size is always close to the nominal level. The local power of the WG* test converges to the predicted value from below while the local power for the FOD test converges from above. The DDIV has size close to the nominal but also local power close to the size of the test, the approximation in this case is very poor.

Table 3: Size and local power for the IV and WG tests.

N		50	100	200	300	1000	*Theory*
$\theta = -0.5$							
c=0	IV	0.053	0.053	0.053	0.051	0.051	0.050
	WG	0.045	0.050	0.047	0.049	0.043	0.050
c=1	IV	0.285	0.382	0.444	0.496	0.567	0.793
	WG	0.057	0.066	0.068	0.076	0.087	0.069
$\theta = 0$							
c=0	IV	0.087	0.073	0.070	0.066	0.066	0.050
	WG	0.058	0.058	0.053	0.053	0.049	0.050
c=1	IV	0.997	0.997	0.997	0.997	0.998	0.998
	WG	0.220	0.274	0.321	0.344	0.414	0.500
$\theta = 0.5$							
c=0	IV	0.072	0.062	0.058	0.055	0.054	0.050
	WG	0.047	0.049	0.049	0.048	0.052	0.050
c=1	IV	0.979	0.985	0.988	0.990	0.993	0.994
	WG	0.388	0.489	0.57	0.610	0.678	0.730

6 Conclusions

This paper examines the power properties of fixed-T tests under serial correlation. Two models are considered, one that has individual intercepts and one that has both individual intercepts and incidental trends. For the first model, asymptotic local power functions of the WG and the IV test show that the two tests behave differently but the IV test is superior to the WG test in many aspects. In the context of MA(1) errors the IV test is always more powerful than the WG. Furthermore, its power increases with T irrespective of the moving average parameter and it is never biased. The WG test shows great gains or losses of power that depend on the sign of the moving average parameter. For both tests positive values support greater power while negative values lead to power loss. This loss can even result in bias for the WG test.

For the model with individual effects and individual trends the WG* and FOD tests behave similarly. They have trivial power when no serial correlation is present and in general lower power than the IV and WG tests. When the errors follow an MA(1) process both tests have power for negative values of the parameter and bias for positive values. This behaviour is the opposite from the IV and WG tests. On the contrary, the DDIV test behave qualitatively like the IV test but has less power.

Table 4:Size and local power for tests WG*, FOD and DDIV.

N		50	100	200	300	1000	*Theory*
$\theta = -0.5$							
c=0	DDIV	0.040	0.045	0.045	0.049	0.047	0.050
	WG*	0.051	0.051	0.048	0.051	0.049	0.050
	FOD	0.057	0.054	0.052	0.049	0.055	0.050
c=1	DDIV	0.039	0.045	0.050	0.049	0.050	0.498
	WG*	0.093	0.105	0.108	0.105	0.107	0.158
	FOD	0.256	0.239	0.224	0.194	0.130	0.060
$\theta = 0$							
c=0	DDIV	0.059	0.052	0.048	0.053	0.049	0.050
	WG*	0.062	0.059	0.056	0.057	0.055	0.050
	FOD	0.084	0.080	0.067	0.062	0.056	0.050
c=1	DDIV	0.050	0.051	0.051	0.052	0.048	0.760
	WG*	0.169	0.148	0.122	0.112	0.089	0.050
	FOD	0.575	0.435	0.309	0.252	0.146	0.050
$\theta = 0.5$							
c=0	DDIV	0.055	0.054	0.051	0.055	0.050	0.050
	WG*	0.063	0.056	0.055	0.056	0.055	0.050
	FOD	0.077	0.062	0.065	0.062	0.061	0.050
c=1	DDIV	0.031	0.038	0.038	0.040	0.045	0.900
	WG*	0.217	0.179	0.135	0.121	0.070	0.033
	FOD	0.533	0.402	0.292	0.231	0.130	0.046

As a by-product of the above analysis, the problem first encountered by Moon and Phillips (1999) and named "incidental trends" is discussed. Asymptotic local power is found only when there is serial correlation for the WG* and FOD tests and always for the DDIV test.

Monte Carlo experiments are conducted to examine the usefulness of the theory. For the first model asymptotic approximations are satisfactory. For the second model the approximation is deemed moderate, a greater N is required. An exception is the double differenced test for which the approximation is very bad. This can be attributed to the presence of individual trends that complicate the detrending of models.

References

[1] Bond, S., Nauges, C., and Windmeijer, F., 2002. 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., 2011. Unit Root Tests for Panel Data with AR(1) Errors and Small T. Econometrics Journal. 15(1), 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] 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.

[7] 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.

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

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

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

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] Moon, H.R. & B. Perron (2004) Asymptotic Local Power of Pooled t-Ratio Tests for Unit Roots in Panels with Fixed Effects. Mimeo, University of Southern California.

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

[17] Moon, H. R., B. Perron and P. C. B. Phillips (2006). A note on "The local power of some unit root tests for panel data" by Breitung. Econometric Theory 22, 1177–88.

[18] 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.

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

[20] Moon H.R., and P. C. B. Phillips (1999). Maximum Likelihood Estimation in Panels with Incidental Trends. Oxford Bulletin of Economics and Statistics, Special Issue (1999), 0305-9049.

[21] 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.

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

[23] Wansbeek T.J. and Knaap T. 1999. Estimating a Dynamic Panel Data Model with Heterogeneous Trends. Annales d'Economie et de Statistique, 55-56, pp. 331-349.

7 Appendix

Proof of Theorem 1 Under local alternatives the statistic is written as:

$$\begin{aligned}
\sqrt{N}(\hat{\varphi}_{IV}-\varphi_N) &= \sqrt{N}\left(\frac{\frac{1}{N}\sum_{i=1}^{N} z'_{i-1}\Pi_p u_i + \frac{1}{N}\sum_{i=1}^{N}(1-\varphi_N)a_i z'_{i-1}\Pi_p e}{\frac{1}{N}\sum_{i=1}^{N} z'_{i-1}\Pi_p z_{i-1}} + \varphi_N - \varphi_N\right) = \\
&= \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N} z'_{i-1}\Pi_p u_i + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(1-\varphi_N)a_i z'_{i-1}\Pi_p e}{\frac{1}{N}\sum_{i=1}^{N} z'_{i-1}\Pi_p z_{i-1}} = \frac{(A)+(B)}{(C)} \qquad (53)
\end{aligned}$$

Under the alternative:

$$y_{-1} = wy_{i0} + \Omega e(1-\varphi_N)a_i + \Omega u_i,\ \ i = 1, ..., N. \qquad (54)$$

where

$$\Omega = \begin{pmatrix} 0 & . & . & . & . & . & 0 \\ 1 & 0 & & & & & . \\ \varphi_N & 1 & . & & & & . \\ \varphi_N^2 & \varphi_N & . & . & & & . \\ . & & . & . & . & & . \\ . & & & . & 1 & 0 & . \\ \varphi_N^{T-2} & \varphi_N^{T-3} & . & . & \varphi_N & 1 & 0 \end{pmatrix} \qquad (55)$$

and $w = (1, \varphi_N, \varphi_N^2, ..., \varphi_N^{T-1})'$. For $\varphi_N = 1$: $\Omega \equiv \Lambda$. The first order Taylor expansions of Ω and w are:

$$\begin{aligned}
\Omega &= \Lambda + F(\varphi_N - 1) + o_p(1) \ \text{ and} \qquad (56)\\
w &= e + f(\varphi_N - 1) + o_P(1) \qquad (57)
\end{aligned}$$

where $F = \frac{d\Omega}{d\varphi_N}\mid_{\varphi_N=1}$ and $f = \frac{dw}{d\varphi_N}\mid_{\varphi_N=1}$. Then,

$$z_{i-1} = y_{i-1} - ey_{i0} = (w-e)y_{i0} + \Omega e(1-\varphi_N)a_i + \Omega u_i \qquad (58)$$

Substituting (58) in (A) we obtain:

$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} z_{i-1}'\Pi_p u_i = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}((w-e)y_{i0} + \Omega e(1-\varphi_N)a_i + \Omega u_i)'\Pi_p u_i = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i0}(w-e)'\Pi_p u_i + (1-\varphi_N)a_i e'\Omega'\Pi_p u_i + u_i\Omega'\Pi_p u_i.$

But

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i0}(w-e)'\Pi_p u_i \xrightarrow{p} 0 \tag{59}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(1-\varphi_N)a_i e'\Omega'\Pi_p u_i \xrightarrow{p} 0 \tag{60}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} u_i\Omega'\Pi_p u_i \xrightarrow{d} N(0, V_{IV,1}) \tag{61}$$

because by construction of $\Pi_p : tr(\Lambda'\Pi_p\Theta) = tr(F'\Pi_p\Theta) = 0$. The previous limits occur after substituting (56) and (57) and by using standard results on quadratic forms found in Schott (1997). Also (B) and (C) :

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(1-\varphi_N)a_i z_{i-1}'\Pi_p e \xrightarrow{p} 0 \tag{62}$$

$$\frac{1}{N}\sum_{i=1}^{N} z_{i-1}'\Pi_p z_{i-1} \xrightarrow{p} tr(\Lambda'\Pi_p\Lambda\Theta) \tag{63}$$

Combining (53)-(63)

$$\sqrt{N}(\hat{\varphi}_{IV} - \varphi_N) \xrightarrow{d} N(0, \frac{V_{IV,1}}{tr(\Lambda'\Pi_p\Lambda\Theta)^2}) \tag{64}$$

$$\sqrt{N}(\hat{\varphi}_{IV} - 1) \xrightarrow{d} N(-c, V_{IV})$$

$$\sqrt{N}(\hat{\varphi}_{IV} - 1)V_{IV}^{-1/2} \xrightarrow{d} N(-\frac{c}{\sqrt{V_{IV}}}, 1) \tag{65}$$

Proof of Theorem 2 The proof is segmented in two parts. Part A contains proof for $c = 0$ that can be directly compared to that of Kruiniger and Tzavalis (2002), Part B contains the proof for $c > 0$.

A) To derive the limiting distribution of the test statistic of the theorem, we will proceed into stages. We first show that the LSDV estimator$\hat{\varphi}_{WG}$ is inconsistent, as $N \to \infty$. Then, will construct a normalized statistic based on $\hat{\varphi}_{WG}$ corrected for its inconsistency (bias) and derive its limiting distribution under the null hypothesis of $\varphi = 1$, as $N \to \infty$.

Decompose the vector $y_{i,-1}$ for model (1) under hypothesis $\varphi = 1$ as

$$y_{i,-1} = e y_{i0} + \Lambda u_i, \tag{66}$$

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

Premultiplying (66) with matrix Q yields

$$Qy_{i,-1} = Q\Lambda u_i, \tag{67}$$

since $Qe = (0, 0, ..., 0)'$. Substituting (67) into $\hat{\varphi}_{WG}$:

$$\hat{\varphi}_{WG} - 1 = \frac{\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1}Qu_i}{\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1}Qy_{i,-1}} = \frac{\frac{1}{N}\sum_{i=1}^{N} u'_i\Lambda' Qu_i}{\frac{1}{N}\sum_{i=1}^{N} u'_i\Lambda' Q\Lambda u_i}. \tag{68}$$

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

$$\frac{1}{N}\sum_{i=1}^{N} u'_i\Lambda' Qu_i \xrightarrow{p} tr(\Lambda' Q\Gamma) \text{ and } \frac{1}{N}\sum_{i=1}^{N} u'_i\Lambda' Q\Lambda u_i \xrightarrow{p} tr(\Lambda' Q\Lambda\Gamma), \tag{69}$$

where "$\xrightarrow{p}$" signifies convergence in probability. Using the last results, the yet non standardized statistic can be written by (68) as

$$\begin{aligned}\sqrt{N}\hat{\delta}\left(\hat{\varphi}_{WG} - 1 - \frac{\hat{b}}{\hat{\delta}}\right) &= \sqrt{N}\hat{\delta}\left(\frac{\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1}Qu_i}{\hat{\delta}} - \frac{tr(\Psi_{p,WG}\hat{\Gamma})}{\hat{\delta}}\right) \\ &= \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} y'_{i,-1}Qu_i - \frac{1}{N}\sum_{i=1}^{N}\Delta y'_i\Psi_{p,WG}\Delta y_i\right).\end{aligned} \tag{70}$$

where

$$\hat{\Gamma} = \frac{1}{N}\sum_{i=1}^{N}\Delta y'_i\Psi_{p,WG}\Delta y_i \tag{71}$$

Since, under the null hypothesis $\phi = 1$, we have $u_i = \Delta y_i$, the last relationship can be written as follows:

$$\begin{aligned}&\sqrt{N}\delta\left(\hat{\varphi}_{WG} - 1 - \frac{\hat{b}}{\hat{\delta}}\right) \\ =\ &\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} u'_i\Lambda' Qu_i - \frac{1}{N}\sum_{i=1}^{N} u'_i\Psi_{p,WG}u_i\right) \\ =\ &\frac{1}{\sqrt{N}}\sum_{i=1}^{N} u'_i(\Lambda' Q - \Psi_{p,WG})u_i = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} tr\left[(\Lambda' Q - \Psi_{p,WG})u_i u'_i\right] \\ =\ &\frac{1}{\sqrt{N}}\sum_{i=1}^{N} W_i,\end{aligned} \tag{72}$$

where W_i constitute random variables with mean

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

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

$$Var(W_i) = Var(u_i'(\Lambda'Q - \Psi_{p,WG})u_i) = 2tr((A_{WG}\Gamma)^2). \tag{73}$$

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

B) The following proof applies for $c > 0$. Also after subtracting y_{i-1} from both sides of (1):

$$\Delta y_i = u_i + (\varphi_N - 1)y_{i-1} + (1 - \varphi_N)a_ie \tag{74}$$

In a first step we find the distribution of the unstandardized statistic around φ_N :

$$\hat{\delta}\sqrt{N}(\hat{\varphi}_{WG} - \tfrac{\hat{b}}{\hat{\delta}} - \varphi_N) = \hat{\delta}\sqrt{N}(\varphi_N + \frac{\frac{1}{N}\sum_{i=1}^{N} y_{i-1}'Qu_i}{\frac{1}{N}\sum_{i=1}^{N} y_{i-1}'Qy_{i-1}'} - \tfrac{\hat{b}}{\hat{\delta}} - \varphi_N) = \sqrt{N}(\tfrac{1}{N}\sum_{i=1}^{N} y_{i-1}'Qu_i -$$

$$tr(\Psi_{p,WG}\hat{\Gamma}) = \sqrt{N}(\tfrac{1}{N}\sum_{i=1}^{N} y_{i-1}'Qu_i - \tfrac{1}{N}\sum_{i=1}^{N} \Delta y_i'\Psi_{p,WG}\Delta y_i) =$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i-1}'Qu_i - \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \Delta y_i'\Psi_{p,WG}\Delta y_i = (C) - (D). \tag{75}$$

We then apply the Lindeberg-Levy CLT to find the limiting distributions of (C) and (D). For (C) :

$$\tfrac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i-1}'Qu_i = \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}(wy_{i0} + \Omega e(1 - \varphi_N)a_i + \Omega u_i)'Qu_i =$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i0}w'Qu_i + a_i(1 - \varphi_N)e'\Omega'Qu_i + u_i'\Omega'Qu_i \tag{76}$$

To find the limit of (76) we first find every limit separately: In all quantities we substitute w and Ω with their Taylor expansions given in (56) and (57) respectively and we substitute φ_N with its local alternatives representation given in (4).

$$\tfrac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i0}w'Qu_i = \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i0}(e + f(\varphi_N - 1) + o_P(1))'Qu_i =$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i0}e'Qu_i + \frac{1}{N}\sum_{i=1}^{N} f'Qu_i y_{i0} + o_p(1) \xrightarrow{p} 0, \text{because} \tag{77}$$

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} f'Qu_i y_{i0} &\rightarrow f'QE(u_i y_{i0}) = 0 \text{ by Assumption 2 and} \\
\frac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i0}e'Qu_i &= 0 \text{ because } e'Q = 0.
\end{aligned}$$

The second summand: $\tfrac{1}{\sqrt{N}}\sum_{i=1}^{N} a_i(1-\varphi_N)e'\Omega'Qu_i = \tfrac{c}{N}\sum_{i=1}^{N} a_i e'(\Lambda' + F'(\tfrac{-c}{\sqrt{N}}) + o_p(1))Qu_i =$

$$\frac{c}{N}\sum_{i=1}^{N} a_i e'\Lambda'\ Qu_i - \frac{c^2}{N^{3/2}}\sum_{i=1}^{N} a_i e'F'Qu_i + o_p(1) \xrightarrow{p} 0 \text{ because} \tag{78}$$

$$\frac{c}{N}\sum_{i=1}^{N} a_i e'\Lambda'\ Qu_i \rightarrow c_i e'\Lambda'\ QE(a_i u_i) = 0, \quad \text{by Assumption 2 and}$$

$$\frac{c^2}{N^{3/2}}\sum_{i=1}^{N} a_i e'F'Qu_i \xrightarrow{p} 0 \text{ due to the fast rate that } \frac{1}{N^{3/2}} \text{ goes to zero.}$$

Then the third summand: $\tfrac{1}{\sqrt{N}}\sum_{i=1}^{N} u_i'\Omega'Qu_i = \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N} u_i'(\Lambda' + F'(\tfrac{-c}{\sqrt{N}}) + o_p(1))Qu_i =$

$$= \frac{1}{\sqrt{N}}\sum_{i=1}^{N} u_i'\Lambda'Qu_i - \frac{c}{N}\sum_{i=1}^{N} u_i'F'Qu_i + o_p(1) \tag{79}$$

$$\frac{c}{N}\sum_{i=1}^{N} u_i'F'Qu_i \rightarrow ctr(F'Q\Gamma) \tag{80}$$

$$\sqrt{N}(\frac{1}{N}\sum_{i=1}^{N} u_i'\Lambda'Qu_i - tr(\Lambda'Q\Gamma)) \xrightarrow{d} N(0, V_{WG,1}) \tag{81}$$

The variance $V_{WG,1}$ is known due to the normality assumption in Assumption 1 but we will deal with the variances towards he end of the proof. Relation (81) requires $\sqrt{N}tr(\Lambda'Q\Gamma)$

to be subtracted, so that the limiting distribution is centered around 0. There is no reason to do so because later, relation (83) requires $\sqrt{N}tr(\Psi_{p,WG}\Gamma)$ to be subtracted. But by construction, $tr(\Lambda' Q\Gamma) = tr(\Psi_{p,WG}\Gamma)$ and thus, cancel out. By adding the results of (77), (78) and (79):

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i-1}' Q u_i \xrightarrow{d} N(-ctr(F' Q\Gamma), V_{WG,1}). \tag{82}$$

The proof for the limiting distribution of (D) is more tedious but it follows the same steps: After substituting (74) in (D)

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

Then by the same arguments that gave (77), (78) and (79) we have:

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} u_i'\Psi_{p,WG}u_i - tr(\overset{d}{\Psi_{p,WG}}\Gamma)\right) \xrightarrow{d} N(0, V_{WG,2}) \tag{83}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} u_i'\Psi_{p,WG}y_{i-1}(\varphi_N-1) \xrightarrow{p} -ctr(\Psi_{p,WG}\Lambda\Gamma) \tag{84}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} u_i'\Psi_{p,WG}e(1-\varphi_N)a_i \xrightarrow{p} 0 \tag{85}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\varphi_N-1)y_{i-1}'\Psi_{p,WG}u_i \xrightarrow{p} -ctr(\Lambda'\Psi_{p,WG}\Gamma) \tag{86}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\varphi_N-1)^2 y_{i-1}'\Psi_{p,WG}y_{i-1} \xrightarrow{p} 0 \tag{87}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\varphi_N-1)y_{i-1}'\Psi_{p,WG}e(1-\varphi_N)a_i \xrightarrow{p} 0 \tag{88}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(1-\varphi_N)a_i e'\Psi_{p,WG}u_i \xrightarrow{p} 0 \tag{89}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(1-\varphi_N)a_i e'\Psi_{p,WG}y_{i-1}(\varphi_N-1) \xrightarrow{p} 0 \tag{90}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(1-\varphi_N)^2 a_i^2 e'\Psi_{p,WG}e \xrightarrow{p} 0 \tag{91}$$

Combining the results (83)-(91) we find that:

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Delta y_i'\Psi_{p,WG}\Delta y_i \to N(-ctr(\Lambda'\Psi_{p,WG}\Gamma) - ctr(\Psi_{p,WG}\Lambda\Gamma), V_{WG,2}) \tag{92}$$

Thus from relations (82) and (92):

$$\hat{\delta}\sqrt{N}(\hat{\varphi}_{WG} - \frac{\hat{b}}{\hat{\delta}} - \varphi_N) \xrightarrow{d} N(-c(tr(F'Q\Gamma) - tr(\Lambda'\Psi_{p,WG}\Gamma) - tr(\Psi_{p,WG}\Lambda\Gamma)), V_{WG}) \tag{93}$$

Notice that $V_{WG} \neq V_{WG,1} + V_{WG,2}$ because (C) and (D) are correlated. V_{WG} can be easily calculated from the variances of (C) and (D) and the covariance between them but the result is straightforward when noticing that the quantities to which the CLT applies do not depend on c. Finally as before

$$\frac{1}{N}\sum_{i=1}^{N} y_{i-1}'Qy_{i-1} \xrightarrow{p} tr(\Lambda'Q\Lambda\Gamma) \tag{94}$$

Therefore the variance of the test is the same under the null and under local alternative hypotheses. Substituting (4) to (93):

$$\begin{aligned}
&\hat{\delta}\sqrt{N}(\hat{\varphi}_{WG} - \frac{\hat{b}}{\hat{\delta}} - 1 + \frac{c}{\sqrt{N}}) \xrightarrow{d} N(-c(tr(F'Q\Gamma) - tr(\Lambda\Psi_{p,WG}\Gamma) - tr(\Lambda\Psi_{p,WG}\Gamma)), V_{WG})\\
&\hat{\delta}\sqrt{N}(\hat{\varphi}_{WG} - \frac{\hat{b}}{\hat{\delta}} - 1) \xrightarrow{d} N(-c(tr(\Lambda'Q\Lambda\Gamma) + tr(F'Q\Gamma) - tr(\Lambda\Psi_{p,WG}\Gamma) - tr(\Lambda\Psi_{p,WG}\Gamma)), V_{WG})\\
&\hat{\delta}\sqrt{N}(\hat{\varphi}_{WG} - \frac{\hat{b}}{\hat{\delta}} - 1)V_{WG}^{-1/2} \xrightarrow{d} N(-c\frac{tr(\Lambda'Q\Lambda\Gamma) + tr(F'Q\Gamma) - tr(\Lambda\Psi_{p,WG}\Gamma) - tr(\Lambda\Psi_{p,WG}\Gamma)}{\sqrt{V_{WG}}}
\end{aligned} \tag{95}$$

Proof of Theorem 3 The main point of departure from the proof of theorem 2 is:

$$y_{i-1} = wy_{i0} + \Omega X\zeta_i + \Omega u_i,\ i = 1, ..., N. \tag{96}$$

where $\zeta_i = \begin{pmatrix} (1-\varphi_N)a_i + \varphi\beta_i \\ (1-\varphi_N)\beta_i \end{pmatrix} = (1-\varphi_N)\begin{pmatrix} a_i \\ \beta_i \end{pmatrix} + \varphi_N\begin{pmatrix} \beta_i \\ 0 \end{pmatrix} = \frac{c}{\sqrt{N}}\begin{pmatrix} a_i - \beta_i \\ \beta_i \end{pmatrix} + \beta_i\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{c}{\sqrt{N}}\mu_i + \beta_i e^*$ where $e^* = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Thus

$$\zeta_i = \frac{c}{\sqrt{N}}\mu_i + \beta_i e^* \tag{97}$$

and

$$\Delta y_i = u_i + (\varphi_N - 1)y_{i-1} + X\zeta_i,\ i = 1, ..., N \tag{98}$$

Following the same steps with the proof of theorem 2:

$\hat{\delta}^* \sqrt{N}(\hat{\varphi}^*_{KT} - \frac{\hat{b}^*}{\hat{\delta}^*} - \varphi_N) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} y'_{i-1} Q^* u_i - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Delta y'_i \Phi_p^{WG} \Delta y_i = (A^*) - (B^*)$. Then

$(A^*) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} y'_{i-1} Q^* u_i = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (w y_{i0} + \Omega X \zeta_i + \Omega u_i)' Q^* u_i.$

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} y_{i0} w' Q^* u_i \ \rightarrow \ 0 \tag{99}$$

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \zeta'_i X \Omega' Q^* u_i \ \rightarrow \ 0 \tag{100}$$

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} u'_i \Omega Q^* u_i \ \rightarrow \ N(tr(\Lambda' Q^* \Gamma) - c tr(F' Q^* \Gamma), V^*_{KT,1}) \tag{101}$$

Limits in (99) and (101) are derived as before. To see why (100) stands:

$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \zeta'_i X \Omega' Q^* u_i = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\frac{c}{\sqrt{N}} \mu_i + \beta_i e^*)' X \Omega' Q^* u_i = \frac{c}{N} \sum_{i=1}^{N} \mu'_i X \Omega' Q^* u_i + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \beta_i e^{*\prime} X \Omega' Q^* u_i.$
But

$$\frac{c}{N} \sum_{i=1}^{N} \mu'_i X \Omega' Q^* u_i \ \rightarrow \ 0 \tag{102}$$

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \beta_i e^{*\prime} X \Omega' Q^* u_i \ = \ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \beta_i e^{*\prime} X \Lambda' Q^* u_i = 0, \tag{103}$$

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

$(B^*) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Delta y'_i \Phi_p^{WG} \Delta y_i = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (u_i + (\varphi_N - 1) y_{i-1} + X \zeta_i)' \Phi_p^{WG} (u_i + (\varphi_N - 1) y_{i-1} + X \zeta_i) =$

$= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} u'_i \Phi_p^{WG} u_i + u'_i \Phi_p^{WG} y_{i-1} (\varphi_N - 1) + u'_i \Phi_p^{WG} X \zeta_i + (\varphi_N - 1) y'_{i-1} \Phi_p^{WG} u_i + (\varphi_N - 1)^2 y'_{i-1} \Phi_p^{WG} y_{i-1} + (\varphi_N - 1) y'_{i-1} \Phi_p^{WG} X \zeta_i +$
$\zeta'_i X' \Phi_p^{WG} u_i + \zeta'_i X' \Phi_p^{WG} y_{i-1} (\varphi_N - 1) + \zeta'_i X' \Phi_p^{WG} X \zeta_i.$

Then

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}u_i'\Phi_p^{WG}u_i \rightarrow N(tr(\Phi_p^{WG}\Gamma),V_{KT,1}^*) \quad (104)$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}u_i'\Phi_p^{WG}y_{i-1}(\varphi_N-1) \rightarrow -ctr(\Phi_p^{WG}\Lambda\Gamma) \quad (105)$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}u_i'\Phi_p^{WG}X\zeta_i \rightarrow 0 \quad (106)$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\varphi_N-1)y_{i-1}'\Phi_p^{WG}u_i \rightarrow -ctr(\Lambda'\Phi_p^{WG}\Gamma) \quad (107)$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\varphi_N-1)^2y_{i-1}'\Phi_p^{WG}y_{i-1} \rightarrow 0 \quad (108)$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\varphi_N-1)y_{i-1}'\Phi_p^{WG}X\zeta_i \rightarrow -cE(\beta_i^2)e^{*\prime}X'\Lambda'\Phi_p^{WG}Xe^* \quad (109)$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\zeta_i'X'\Phi_p^{WG}u_i \rightarrow 0 \quad (110)$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\zeta_i'X\Phi_p^{WG}y_{i-1}(\varphi_N-1) \rightarrow -cE(\beta_i^2)e^{*\prime}X'\Phi_p^{WG}\Lambda Xe^* \quad (111)$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\zeta_i'X\Phi_p^{WG}X\zeta_i \rightarrow ctr(X'\Phi_p^{WG}Xe^*E(\mu_i'\beta_i))+ctr(e^{*\prime}X'\Phi_p^{WG}XE(\mu_i\beta_i)) \quad (112)$$

Define $E(\mu_i\beta_i)=E(\beta_i^2)\tilde{e}$ where $\tilde{e}=\begin{pmatrix}-1\\1\end{pmatrix}$Combining (104)-(112)

$$\hat{\delta}^*\sqrt{N}(\hat{\varphi}_{KT}^*-\frac{\hat{b}^*}{\hat{\delta}^*}-1)\rightarrow N(-c(C+E(\beta_i^2)D),V_{KT}^*) \quad (113)$$

wher

$$C = tr(\Lambda'Q^*\Lambda\Gamma)+tr(F'Q^*\Gamma)-tr(\Phi_p^{WG}\Lambda\Gamma)-tr(\Lambda'\Phi_p^{WG}\Gamma) \text{ and} \quad (114)$$

$$D = tr(X'\Phi_p^{WG}Xe^*\tilde{e}')+tr(e^{*\prime}X'\Phi_p^{WG}X\tilde{e})-e^{*\prime}X'\Lambda'\Phi_p^{WG}Xe^*-e^{*\prime}X'\Phi_p^{WG}\Lambda Xe^* \quad (115)$$

To reach the final result the following identities hold:

$$tr(X'\Phi_p^{WG}Xe^*\tilde{e}') = e^{*\prime}X'\Lambda'\Phi_p^{WG}Xe^* \quad (116)$$

$$tr(e^{*\prime}X'\Phi_p^{WG}X\tilde{e}) = e^{*\prime}X'\Phi_p^{WG}\Lambda Xe^* \quad (117)$$

Thus

$$\hat{\delta}^* \sqrt{N}(\hat{\varphi}^*_{KT} - \frac{\hat{b}^*}{\hat{\delta}^*} - 1)V_{KT}^{*-1/2} \rightarrow N(-c\frac{tr(\Lambda' Q^* \Lambda \Gamma) + tr(F' Q^* \Gamma) - tr(\Phi_p^{WG} \Lambda \Gamma) - tr(\Lambda' \Phi_p^{WG} \Gamma)}{\sqrt{V^*_{KT}}}, 1) \tag{118}$$

As can be seen the nuisance parameters do not affect the distribution.

Proof of Theorem 4 In the following derivations we use the following equations, under the null and the alternative:

$$z_i = \varphi_N z_{i-1} + X\zeta_i + u_i, \tag{119}$$

$$z_{i-1} = \Omega X \zeta_i + \Omega u_i + (w - e) y_{i0}, \tag{120}$$

$$\Delta z_i = (\varphi_N - 1) z_{i-1} + X\zeta_i + u_i. \tag{121}$$

We start by showing its asymptotic distribution under the null.

$$\hat{\varphi}_B - 1 = \frac{\frac{1}{N}\sum_{i=1}^{N} z_i' B' A \Delta z_i}{\frac{1}{N}\sum_{i=1}^{N} z_i' B' B z_i} = \frac{(A)}{(B)}$$

Then $(A) : \frac{1}{N}\sum_{i=1}^{N} z_i' B' A \Delta z_i = \frac{1}{N}\sum_{i=1}^{N} (z_{i-1}' + \beta_i e' + u_{i'}) B' A (\beta_i e + u_i) = \frac{1}{N}\sum_{i=1}^{N} (u_i' \Lambda' + \beta_i e' \Lambda' + \beta_i e' + u_{i'}) B' A (\beta_i e + u_i) = \frac{1}{N}\sum_{i=1}^{N} (u_i'(\Lambda' + I_T) + \beta_i \tau') B' A (\beta_i e + u_i) = \frac{1}{N}\sum_{i=1}^{N} u_i'(\Lambda' + I_T) B' A u_i$ because $(\Lambda + I_T)e = \tau$, $\tau' B' = 0_{1XT}$ and $B' A e = 0_{TX1}$.

$$\frac{1}{N}\sum_{i=1}^{N} u_i'(\Lambda' + I_T) B' A u_i = \frac{1}{N}\sum_{i=1}^{N} u_i' \Xi u_i \rightarrow tr(\Xi \Gamma). \tag{122}$$

since $(\Lambda' + I_T) B' A = \Xi$ by definition. Thus if $tr(\Xi\Gamma) = 0$ the test is unbiased. This happens only if $\Gamma = \sigma^2 I_T$, both homoscedasticity and non-autocorrelation are essential.

$(B) : \frac{1}{N}\sum_{i=1}^{N} z_i' B' B z_i = \frac{1}{N}\sum_{i=1}^{N} (z_{i-1}' + \beta_i e' + u_{i'}) B' B (z_{i-1} + \beta_i e + u_i) = \frac{1}{N}\sum_{i=1}^{N} (u_i'(\Lambda' + I_T) + \beta_i \tau') B' B ((\Lambda + I_T) u_i + \beta_i \tau) = \frac{1}{N}\sum_{i=1}^{N} u_i'(\Lambda' + I_T) B' B (\Lambda + I_T) u_i$ and thus

$$\frac{1}{N}\sum_{i=1}^{N} u_i'(\Lambda' + I_T) B' B (\Lambda + I_T) u_i \rightarrow tr((\Lambda' + I_T) B' B (\Lambda + I_T) \Gamma) \tag{123}$$

Result (123) is only used in the no serial correlation version of the FOD test. The proof of this version is straightforward along the lines of the proof of theorem 2. The rest of this proof is focused on the general form of the FOD test. The test statistic

$$\begin{aligned}
\sqrt{N}\hat{\delta}_{FOD}(\hat{\varphi}_{FOD}-1-\frac{\hat{b}_{FOD}}{\hat{\delta}_{FOD}}) &= \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}z_i'B'Bz_i\right)\left(1+\frac{\frac{1}{N}\sum_{i=1}^{N}z_i'B'A\Delta z_i}{\frac{1}{N}\sum_{i=1}^{N}z_i'B'Bz_i}-1-\frac{\frac{1}{N}\sum_{i=1}^{N}\Delta z_i'\Phi_p^{FOD}\Delta z_i}{\frac{1}{N}\sum_{i=1}^{N}z_i'B'Bz_i}\right) \\
&= \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}u_i'(\Lambda'+I_T)B'Au_i-\frac{1}{N}\sum_{i=1}^{N}\Delta z_i'\Phi_p^{FOD}\Delta z_i\right) \\
&= \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}\Delta z_i'\left[(\Lambda'+I_T)B'A-\Phi_p^{FOD}\right]\Delta z_i\right) \\
&= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Delta z_i'\left[\Xi-\Phi_p^{FOD}\right]\Delta z_i \qquad (12
\end{aligned}$$

because $\Delta z_i'(\Lambda'+I_T)B'A\Delta z_i=(\beta_i e'+u_i')(\Lambda'+I_T)B'A(\beta_i e+u_i)=u_i'(\Lambda'+I_T)B'Au_i$ from the identities above. Asymptotically, based on standard matrix algebra results:

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Delta z_i'\left[\Xi-\Phi_p^{FOD}\right]\Delta z_i \to N(0,2tr(A_{FOD}^2)). \tag{125}$$

The above concludes the proof for the distribution under the null. Under local alternatives, the proof follows the steps of the proof of theorem 3, notting that the following identities apply:

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

Proof of Theorem 5 Like the proof of Theorem 1. Under local alternatives

$$
\begin{aligned}
\sqrt{N}(\hat{\varphi}^*_{IV} - \varphi_N) &= \sqrt{N}\left(\frac{\sum_{i=1}^{N} y^{*\prime}_{i-1}\Pi^*_p y^*_i}{\sum_{i=1}^{N} y^{*\prime}_{i-1}\Pi^*_p y^*_{i-1}} - \varphi_N\right) \\
&= \sqrt{N}\left(\frac{\sum_{i=1}^{N} y^{*\prime}_{i-1}\Pi^*_p(\varphi_N y_{i-1} + (1-\varphi_N)\beta^*_i e^* + u^*_i)}{\sum_{i=1}^{N} y^{*\prime}_{i-1}\Pi^*_p y^*_{i-1}} - \varphi_N\right) \\
&= \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(1-\varphi_N)\beta^*_i y^{*\prime}_{i-1}\Pi^*_p e^* + \frac{1}{\sqrt{N}}\sum_{i=1}^{N} y^{*\prime}_{i-1}\Pi^*_p u^*_i}{\frac{1}{N}\sum_{i=1}^{N} y^{*\prime}_{i-1}\Pi^*_p y^*_{i-1}}.
\end{aligned}
$$

Where

$$
\begin{aligned}
\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(1-\varphi_N)\beta^*_i y^{*\prime}_{i-1}\Pi^*_p e^* &\rightarrow 0, \\
\frac{1}{\sqrt{N}}\sum_{i=1}^{N} y^{*\prime}_{i-1}\Pi^*_p u^*_i &\rightarrow N(0, 2tr((A^*_{IV}\Theta)^2), \\
\frac{1}{N}\sum_{i=1}^{N} y^{*\prime}_{i-1}\Pi^*_p y^*_{i-1} &\rightarrow tr(\Lambda^{*\prime}\Pi^*_p\Lambda^*),
\end{aligned}
$$

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

Proof of Proposition 1 This proof is more general than that of Madsen (2010). The latter can be seen as a special case of this by substituting Q^* with Q. Under local alternatives the Harris and Tzavalis 1999 statistic can be written as

$$\sqrt{N}\left(\hat{\varphi}^*_{WG} - \varphi_N - \frac{tr(\Lambda' Q^*)}{tr(\Lambda' Q^* \Lambda)}\right)$$

because, under the null, $plim(\hat{\varphi}^*_{WG} - 1) = \frac{tr(\Lambda' Q^*)}{tr(\Lambda' Q^* \Lambda)}$. Then

$$\begin{aligned}
\sqrt{N}\left(\hat{\varphi}^*_{WG} - \varphi_N - \frac{tr(\Lambda' Q^*)}{tr(\Lambda' Q^* \Lambda)}\right) &= \sqrt{N}\left(\frac{\frac{1}{N}\sum_{i=1}^{N} y'_{i-1}Q^*(\varphi_N y_{i-1} + X\zeta_i + u_i)}{\frac{1}{N}\sum_{i=1}^{N} y'_{i-1}Q^* y_{i-1}} - \varphi_N - \frac{\frac{tr(\Lambda' Q^*)}{tr(\Lambda' Q^* \Lambda)}\frac{1}{N}\sum_{i=1}^{N} y'_{i-1}Q}{\frac{1}{N}\sum_{i=1}^{N} y'_{i-1}Q^* y_{i-}}\right. \\
&= \frac{\frac{1}{\sqrt{N}}\left(\sum_{i=1}^{N} y'_{i-1}Qu_i - \frac{tr(\Lambda' Q^*)}{tr(\Lambda' Q^* \Lambda)} y'_{i-1}Q^* y_{i-1}\right)}{\frac{1}{N}\sum_{i=1}^{N} y'_{i-1}Q^* y_{i-1}} \\
&= \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N} y'_{i-1}Qu_i - \frac{tr(\Lambda' Q^*)}{tr(\Lambda' Q^* \Lambda)}\frac{1}{\sqrt{N}}\sum_{i=1}^{N} y'_{i-1}Q^* y_{i-1}}{\frac{1}{N}\sum_{i=1}^{N} y'_{i-1}Q^* y_{i-1}}
\end{aligned}$$

Using similar arguments with the previous proofs and taking into account the fact that

$$tr(\Lambda' Q^* F)\frac{tr(\Lambda' Q^*)}{tr(\Lambda' Q^* \Lambda)} + tr(F' Q^* \Lambda)\frac{tr(\Lambda' Q^*)}{tr(\Lambda' Q^* \Lambda)} - tr(tr(\Lambda' Q^* \Lambda) - tr(F' Q^*) = 0 \qquad (127)$$

leads to the result.

Proof of Corollary 1 The proof is straightforward from (9) and the results of Corollaries 3 and 4. As in De Wachter at al. (2007), scale the statistic with T and apply the continuous mapping theorem. The joint convergence is guaranteed by Hahn and Kuersteiner (2002) and De Wachter at al. (2007). However this proof, is not intuitive with regard to the local alternatives in this case. To clearly show that the large T version of the test has a limiting distribution on the local alternatives $\varphi_{NT} = 1 - \frac{c}{T\sqrt{N}}$ we provide the following proof only for the IV test. Consider,

$$\hat{\varphi}_{IV} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T-p-1} z_{it}z_{it+p+1}}{\sum_{i=1}^{N}\sum_{t=1}^{T-p-1} z_{it}z_{it+p}}. \qquad (128)$$

Then, under the above local alternatives

$$
\begin{aligned}
T\sqrt{N}(\hat{\varphi}_{IV}-\varphi_{NT}) &= T\sqrt{N}\left(\frac{\sum_{i=1}^{N}\sum_{t=1}^{T-p-1} z_{it}(\varphi_{NT}z_{it+p}+u_{it+p+1}+(1-\varphi_{NT})a_i)}{\sum_{i=1}^{N}\sum_{t=1}^{T-p-1} z_{it}z_{it+p}}-\varphi_{NT}\right) \\
&= \frac{\frac{1}{\sqrt{N}}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T-p-1}(z_{it}u_{it+p+1}+z_{it}(1-\varphi_{NT})a_i))}{\frac{1}{N}\frac{1}{T^2}\sum_{i=1}^{N}\sum_{t=1}^{T-p-1} z_{it}z_{it+p}}. \qquad (129)
\end{aligned}
$$

Notice that under local alternatives

$$
\begin{aligned}
z_{it} &= \varphi_{NT}^{t}z_{i0}+\varphi_{NT}^{t-1}u_{it-1}+...+u_{it} \qquad (130)\\
&= \varphi_{NT}^{t-1}u_{i1}+\varphi_{NT}^{t-2}u_{i2}+...+u_{it}. \qquad (131)
\end{aligned}
$$

Also, from the binomial theorem

$$
\varphi_{NT}^{t}=1+o(T). \qquad (132)
$$

Inserting (132) in (130) we take

$$
z_{it}u_{it+p+1}=u_{i1}+...+u_{it}+o(T). \qquad (133)
$$

The last equality allows the use of standard asymptotic results about AR(1) processes (see also Hamilton (1994)). To find the limiting distribution in (129) we show where the three sums converge:

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

Taking first $T\to\infty$,

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

and thus

$$\frac{1}{\sqrt{N}}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T-p-1} z_{it}(1-\varphi_{NT})a_i \rightarrow 0. \tag{134}$$

$$\frac{1}{\sqrt{N}}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T-p-1} z_{it}u_{it+p+1} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T-p-1} z_{it}u_{it+p+1}. \tag{135}$$

As $T \rightarrow \infty$,

$$\frac{1}{T}\sum_{t=1}^{T-p-1} z_{it}u_{it+p+1} \rightarrow \frac{1}{2}\sigma^2\left\{[W_i(1)]^2-1\right\} \tag{136}$$

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, thus, $E\{[W_i(1)]^2\} = 1$ and $Var\{[W_i(1)]^2\} = 2$. Then as $N \rightarrow \infty$ (135) becomes:

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{2}\sigma^2\left\{[W_i(1)]^2-1\right\} \rightarrow N(0,\frac{\sigma^4}{2}) \tag{137}$$

Finally, the limit of the denominator comes from:

$$\frac{1}{N}\frac{1}{T^2}\sum_{i=1}^{N}\sum_{t=1}^{T-p-1} z_{it}z_{it+p} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T-p-1} z_{it}z_{it+p} \tag{138}$$

Then

$$z_{it+p} = \varphi_{NT}^{p}z_{it} + \sum_{k=1}^{p}\varphi_{NT}^{k-1}(1-\varphi_{NT})a_i + \sum_{k=1}^{p}\varphi_{NT}^{k-1}u_{it+(p-k-1)} \tag{139a}$$

Plugging (139a) into (138) and acknowledging (132):

$$\frac{1}{T^2}\sum_{t=1}^{T-p-1}\varphi_{NT}^{p}z_{it}^2 = \frac{1}{T^2}\sum_{t=1}^{T-p-1}z_{it}^2 + o(T) \rightarrow \sigma^2\int_0^1[W_i(r)]^2dr \tag{140}$$

$$\frac{1}{T^2}\sum_{t=1}^{T-p-1}z_{it}\sum_{k=1}^{p}\varphi_{NT}^{k-1}(1-\varphi_{NT})a_i \rightarrow 0 \tag{141}$$

$$\frac{1}{T^2}\sum_{t=1}^{T-p-1}z_{it}\sum_{k=1}^{p}\varphi_{NT}^{k-1}u_{it+(p-k-1)} = \frac{1}{T^2}\sum_{t=1}^{T-p-1}z_{it}\sum_{k=1}^{p}u_{it+(p-k-1)} + o(T) \rightarrow 0 \tag{142}$$

Inserting (140), (141) and (142) into (138) and as $N \rightarrow \infty$:

$$\frac{1}{N}\sum_{i=1}^{N}\sigma^2\int_0^1[W_i(r)]^2dr \rightarrow -\frac{\sigma^2}{2} \tag{143}$$

The moment in (143) is found in Levin et al. (2002).

Combining (134), (137) and (143)

$$
\begin{aligned}
T\sqrt{N}(\hat{\varphi}_{IV}-\varphi_{NT}) &\rightarrow N(0,2)\\
T\sqrt{N}(\hat{\varphi}_{IV}-1) &\rightarrow N(-c,2).
\end{aligned}
$$

Proof of Corollary 2 Assuming $\Gamma_N=\sigma_u^2 I_T$ an important identity is

$$2tr({A_{IV}}^2)=tr(\Lambda'\Pi_p\Lambda),\ \text{for each } p \tag{144}$$

The following identities apply:

$$
\begin{aligned}
tr(\Lambda'\Pi_p\Lambda) &= \frac{1}{2}(T^2-T),\ \text{for } p=0\\
tr(\Lambda'\Pi_p\Lambda) &= \frac{[\frac{1}{2}(T-2)(T-1)]^2}{\frac{1}{2}T(T-3)+1},\ \text{for } p=1\\
tr(\Lambda'\Pi_p\Lambda) &= \frac{T^2}{2}-\frac{5T}{2}+3,\ \text{for } p=2\\
tr(\Lambda'\Pi_p\Lambda) &= \frac{T^2}{2}-\frac{7T}{2}+6,\ \text{for } p=3 \qquad (145)
\end{aligned}
$$

When a MA(1) process is assumed for the error term, De Wachter, Harris and Tzavalis (2007) find that:

$$
\begin{aligned}
D(\theta,T) &= D_{1,IV}\theta^2+D_{2,IV}\theta+D_{1,IV} \qquad (146)\\
R(\theta,T) &= R_{1,IV}\theta^4+R_{2,IV}\theta^3+R_{3,IV}\theta^2+R_{2,IV}\theta+R_{1,IV}
\end{aligned}
$$

where

$$
\begin{aligned}
D_{1,IV} &= \frac{T^2}{2}-\frac{3T}{2}+1\\
D_{2,IV} &= T^2-4T+4 \qquad (147)\\
R_{1,IV} &= \frac{1}{2}T(T-3)+1 \qquad (148)\\
R_{2,IV} &= 2T(T-5)+12\\
R_{3,IV} &= 3T(T-5)+20 \qquad (149)
\end{aligned}
$$

Proof of Corollary 3 Substitute in (95) the following identities to derive the final results:
$p=0:$

$$
\begin{aligned}
tr(\Lambda'\Psi_{p,WG}) &= 0 \\
tr(\Psi_{p,WG}\Lambda) &= 0 \\
tr(F'Q) &= -\frac{T^2}{6} + \frac{T}{2} - \frac{1}{3} \\
tr(\Lambda' Q\Lambda) &= \frac{T^2-1}{6} \\
tr(A_{WG}^2) &= tr(Y^2) - tr(Y_\Psi^2) \\
tr(Y^2) &= \frac{1}{2}tr((\Lambda' Q)^2) + \frac{1}{2}tr(\Lambda' Q\Lambda) \\
tr((\Lambda' Q)^2) &= -\frac{T^2}{12} + \frac{T}{2} - \frac{5}{12} \\
tr(Y_\Psi^2) &= \frac{T}{3} + \frac{1}{6T} - \frac{1}{2}
\end{aligned} \tag{150}
$$

$p = 1:$ With a different p the results that contain matrix $\Psi_{p,WG}$ change.

$$
\begin{aligned}
tr(\Lambda'\Psi_{p,WG}) &= \frac{T-1}{2} \\
tr(\Psi\Lambda_p) &= -\frac{T}{2} - \frac{1}{T} + \frac{3}{2} \\
tr(Y_\Psi^2) &= \frac{1}{2}tr(F^2) + \frac{1}{2}tr(F'F) \\
tr(F^2) &= -\frac{1}{2T} + \frac{1}{2} \\
tr(F'F) &= T + \frac{5}{2T} - \frac{1}{T^2} - \frac{5}{2}
\end{aligned} \tag{151}
$$

$p = 2:$

$$
\begin{aligned}
tr(\Psi_{p,WG}\Lambda) &= -\frac{(T-2)^2}{T} \\
tr(\Lambda'\Psi_{p,WG}) &= \frac{(T-1)^2}{T} \\
tr(F^2) &= -\frac{1}{3}T - \frac{25}{6T} + \frac{2}{T^2} + \frac{5}{2} \\
tr(F'F) &= \frac{5}{3}T + \frac{65}{6T} - \frac{7}{T^2} - \frac{13}{2}
\end{aligned} \tag{152}
$$

$p = 3:$

$$\begin{aligned} tr(\Psi_{p,WG}\Lambda) &= -\frac{3T}{2} - \frac{10}{T} + \frac{15}{2} \\ tr(\Lambda'\Psi_{p,WG}) &= \frac{3T}{2} + \frac{4}{T} - \frac{9}{2} \\ tr(F^2) &= -\frac{2}{3}T - \frac{77}{6T} + \frac{10}{T^2} + 11 \\ tr(F'F) &= \frac{7}{3}T + \frac{175}{6T} - \frac{26}{T^2} - \frac{25}{2} \end{aligned} \tag{153}$$

$p = 1$, u_i **is a MA(1) process.** Denote θ the moving average parameter, then the local power function has the representation:

$$\begin{aligned} tr(\Lambda'Q\Lambda\Gamma) &= (\frac{T^2-1}{6})\theta^2 + (\frac{1}{3}T^2 - T + \frac{2}{3})\theta + (\frac{T^2-1}{6}) \\ tr(F'Q\Gamma) &= (-\frac{T^2}{6} + \frac{T}{2} - \frac{1}{3})\theta^2 + (-\frac{T^2}{3} + \frac{3T}{2} + \frac{1}{T} - \frac{13}{6})\theta + (-\frac{T^2}{6} + \frac{T}{2} - \frac{1}{3}) \\ tr(\Lambda'\Psi\Gamma) &= \frac{1}{2}(T-1)\theta^2 + (-\frac{2}{T} + 1)\theta + \frac{1}{2}(T-1) \\ tr(\Psi\Lambda\Gamma) &= (-\frac{T}{2} - \frac{1}{T} + \frac{3}{2})\theta^2 + (-T - \frac{4}{T} + 4)\theta + (-\frac{T}{2} - \frac{1}{T} + \frac{3}{2}) \\ 2tr((A_{WG}\Gamma)^2) &= R_1\theta^4 + R_2\theta^3 + R_3\theta^2 + R_2\theta + R_1 \\ R_1 &= \frac{T^2}{12} - \frac{T}{2} - \frac{2}{T} + \frac{1}{T^2} + \frac{17}{12} \\ R_2 &= \frac{T^2}{3} - \frac{10T}{3} - \frac{80}{3T} + \frac{20}{T^2} + \frac{41}{3} \\ R_3 &= \frac{T^2}{2} - 5T - \frac{45}{T} + \frac{38}{T^2} + \frac{43}{2}. \end{aligned} \tag{154}$$

Proof of Corollary 4 Under the assumptions of Corollary 4 the following identity holds:

$$tr(\Lambda'Q^*\Lambda\Gamma) + tr(F'Q^*\Gamma) = tr(\Phi_p^{WG}\Lambda\Gamma) + tr(\Lambda'\Phi_p^{WG}\Gamma) \tag{155}$$

Proof of Corollary 5 Immediate from the proof of theorem 4.

Proof of Corollary 6 Under the assumptions of Corollary 6

$$E(u_i^* u_i^{*\prime}) = \begin{pmatrix} q & r & s & & & 0 \\ r & q & r & s & & \\ s & r & q & r & s & \\ & \cdot & \cdot & \cdot & \cdot & \\ & & \cdot & \cdot & \cdot & \cdot \\ 0 & & & s & r & q \end{pmatrix} \tag{156}$$

where $q = (2 + \theta^2) - 2\theta$, $r = 2\theta - (1 + \theta^2)$ and $s = -\theta$. The following identities also hold:

$$tr(\Lambda^{*\prime}\Pi_p^*\Lambda^*) = T - p - 3 \tag{157}$$

$$tr((A_{IV}^*\Theta)^2) = T - p - 2 \tag{158}$$

Also,

$$\begin{aligned} P_1 &= 2(T - 3) \\ P_2 &= -2(2T - 8) \\ P_3 &= 2(4T - 15) \end{aligned}$$