The Econometrics Journal

Econometrics Journal (2010), volume **13**, pp. 63–94.
doi: 10.1111/j.1368-423X.2009.00302.x

Unit root inference in panel data models where the time-series dimension is fixed: a comparison of different tests

EDITH MADSEN[†]

[†]*Department of Economics and CAM, University of Copenhagen, Øster Farimagsgade 5, DK-1353 Copenhagen K, Denmark*
E-mail: edith.madsen@econ.ku.dk

First version received: September 2008; final version accepted: September 2009

Summary The objective of the paper is to investigate and compare the performance of some of the unit root tests in micropanels, which have been suggested in the literature. The framework is a first-order autoregressive panel data model allowing for heterogeneity in the intercept but not in the autoregressive parameter. The tests are all based on usual t-statistics corresponding to least squares estimators of the autoregressive parameter resulting from different transformations of the observed variables. The performance of the tests is investigated and compared by deriving the local power of the tests when the autoregressive parameter is local-to-unity. The results show that the assumption concerning the initial values is extremely important in this matter. The outcome of a simulation experiment demonstrates that the local power of the tests provides a good approximation to their actual power in finite samples.

Keywords: *Dynamic panel data model*, *Initial values*, *Local alternatives*, *Unit roots*.

1. INTRODUCTION

In this paper, we investigate unit root inference in panel data models where the cross-section dimension is much larger than the time-series dimension. So we consider traditional micropanels. At present there is a large econometric literature dealing with unit root testing in panel data models which has developed during the last 10 years. Contrary to the previous literature on dynamic panel data models, a large part of this new literature considers macropanels where the cross-section and time-series dimensions are similar in magnitude. Banerjee (1999), Baltagi and Kao (2000) and Breitung and Pesaran (2008) review many of the contributions to the literature on unit root testing in panel data models. Reviews of the literature on dynamic micropanels are provided in Hsiao (1986), Baltagi (1995) and Arellano (2003) of which only the latter discusses the issue of unit roots.

The analysis in this paper is done within the framework of a first-order autoregressive panel data model allowing for individual-specific levels. This means that we are testing the null hypothesis of each time-series process being a random walk without drift against the alternative hypothesis of each time-series process being stationary with individual-specific levels but the same autoregressive parameter for all cross-section units. In the autoregressive panel data model there are two sources of persistency. One is the autoregressive mechanism which is the same for all cross-section units and the other is the unobserved individual-specific means. Everything else being equal, a high value of the autoregressive parameter means that more persistency is attributed to the autoregressive mechanism. The null hypothesis means that the effect of unanticipated shocks will persist over time, whereas the alternative hypothesis means that the effect will eventually disappear as time goes by. The hypothesis is of interest since many economic variables at the individual level, such as income of individuals and firm-level variables, are found to be highly persistent over time.

The main contribution of the paper is to provide analytical results about the performance of some of the unit root tests which have been suggested in the literature. This is done by deriving the limiting distributions of the corresponding test statistics under local alternatives when the autoregressive parameter is local-to-unity. The results are used to compare the performance of the different tests in terms of their local power. In addition, the results reveal how the local power of the tests is affected by the nuisance parameters of the data-generating process (DGP). So far the power properties of unit root tests in micropanels have been investigated and compared in simulation studies; see, for example, Bond et al. (2002) and Hall and Mairesse (2005). However, the outcome of these might depend on the particular choice of nuisance parameters in the simulation set-up in a non-transparent way. Therefore, it seems to be a useful contribution within this research area. The paper by Breitung (2000) is related to this paper as it compares the local power of some of the unit root tests in macropanels.

We consider three different unit root tests. They are all based on *t*-statistics corresponding to different least-squares (LS) estimators of the autoregressive parameter. The reason that this is not a trivial testing problem is the presence of the individual-specific (incidental) intercepts. Without the presence of these parameters standard testing theory implies that the *t*-statistic based on the OLS estimator of the autoregressive parameter in the original model gives a test which is optimal asymptotically. This is the first test we consider and we would expect it to perform well in terms of having high power when there is no or little variation in the individual-specific intercepts. On the other hand, when the variation in the individual-specific intercepts is high the OLS estimator has a substantial positive asymptotic bias and therefore the OLS unit root test is expected to have low power in this case. The other two tests we consider are both invariant with respect to adding an individual-specific constant to all variables but they differ in terms of the way in which the invariance with respect to this type of transformation is obtained. In other words, they use different ways of removing the individual-specific means from the variables. One subtracts the initial values from all variables and is suggested by Breitung and Meyer (1994) and the other subtracts the respective individual-specific time-series means of the variables from both sides of the equation and is suggested by Harris and Tzavalis (1999). The Breitung–Meyer test and the Harris–Tzavalis test are panel data versions of the unit root tests in single time series suggested by Schmidt and Phillips (1992) and Dickey and Fuller (1979), respectively. The Breitung–Meyer estimator of the autoregressive parameter is consistent under the null hypothesis whereas the Harris–Tzavalis estimator (the within-group estimator) is inconsistent and therefore a bias adjustment is necessary. Both estimators are inconsistent under the alternative hypothesis meaning that the removal of the individual-specific means that cause the inconsistency of the OLS estimator leads to new sources of asymptotic bias.

From the description above, it is not straightforward to determine which test is best in terms of having the highest power. It turns out that the initial values are crucial for the performance of the tests in terms of asymptotic power under local alternatives. In general it is always important to be aware of the power properties when applying a statistical test in practice and if there are several tests to choose from it is especially important to understand how their performance is affected by nuisance parameters in order to choose the best testing procedure. Even if it is not the case that one test outperforms the others for all values of the nuisance parameters it is important to understand under which assumptions the tests are likely to have high or low power. The importance of the initial values when testing the unit root hypothesis in micropanels is a result which is also found for single time series and macropanels; see Müller and Elliott (2003) and Harris et al. (2008), respectively.

An important finding is that under mean stationary alternatives the local power of the Breitung–Meyer test is always higher than the local power of the Harris–Tzavalis test. This result is similar to findings in Moon et al. (2007) where they derive the power envelope of unit tests in macropanels. In the case with individual-specific intercepts they find that within the class of tests that are invariant with respect to individual-specific constants the macropanel version of the Breitung–Meyer test has asymptotic power equal to the power envelope whereas the macropanel version of the Harris–Tzavalis test suggested by Levin et al. (2002) has lower asymptotic power than the power envelope. This result is different from the findings for single time-series versions of these unit root tests where the Schmidt–Phillips test is close to being optimal for values of the autoregressive parameter close to unity whereas the Dickey–Fuller test is close to being optimal for values of the autoregressive parameter close to zero; see Hwang and Schmidt (1996). An important difference is that in macropanels it is possible to find tests that are uniformly most powerful, whereas this is not possible in single time series. In addition, we find that when there is no or little variation in the individual-specific means the local power of the OLS test is higher than the local power of the Breitung–Meyer test. This result implies that the estimation of the individual-specific means causes a decrease in local power since the number of observations over time which contains information about the individual-specific means remains constant and hence it makes a difference whether or not the individual-specific means are estimated.

The paper is organized as follows. In Section 2, the basic model is specified. In Section 3, we investigate and compare the different unit root tests described above. This is done by deriving the limiting distributions of the corresponding test statistics under local alternatives. In Section 4, the analytical results are illustrated in a simulation study. In Section 5, we provide some concluding remarks. Proofs are provided in the Appendix.

2. THE MODEL AND ASSUMPTIONS

We consider the first-order autoregressive panel data model with individual-specific intercepts defined by

$$y_{it} = \rho y_{it-1} + (1-\rho)\alpha_i + \varepsilon_{it} \quad \text{for } i = 1, \dots, N \text{ and } t = 1, \dots, T, \tag{2.1}$$

where $-1 < \rho \leq 1$ and for every $i = 1, \dots, N$ the sequence $\{\varepsilon_{it}\}_{t=1}^{\infty}$ is white noise. For notational convenience we assume that the initial values y_{i0} are observed such that the actual number of observations over time equals $T + 1$. The model provides a framework for testing the null hypothesis of each time-series process being a random walk against the alternative hypothesis of each time-series process being stationary with an individual-specific level. To specify the model further the assumptions below are imposed.

ASSUMPTION 2.1. *ε_{it} is independent across i, t with $E(\varepsilon_{it}) = 0$, $E(\varepsilon_{it}^2) = \sigma_{i\varepsilon}^2$ and $E(\varepsilon_{it}^4) = E(\varepsilon_{is}^4)$ for all $t, s = 1, \ldots, T$. In addition, ε_{it} is independent of α_i and y_{i0}.*

ASSUMPTION 2.2. *α_i is i.i.d. across i with $E(\alpha_i) = 0$, $E(\alpha_i^2) = \sigma_\alpha^2$ and $E(\alpha_i^4) < \infty$.*

ASSUMPTION 2.3. *For $-1 < \rho \leq 1$ the initial values satisfy $y_{i0} = \alpha_i + \sqrt{\tau}\varepsilon_{i0}$, where ε_{i0} is independent of α_i and independent across i with $E(\varepsilon_{i0}) = 0$ and $E(\varepsilon_{i0}^2) = \sigma_{i\varepsilon}^2$.*

ASSUMPTION 2.4. *The following hold: (i) $E|\varepsilon_{it}|^{4+\delta} < K < \infty$ for some $\delta > 0$ and all $i = 1, \ldots, N$, $t = 0, 1, \ldots, T$. (ii) $\frac{1}{N}\sum_{i=1}^N \sigma_{i\varepsilon}^2 \to \sigma_{2\varepsilon} > 0$ as $N \to \infty$. (iii) $\frac{1}{N}\sum_{i=1}^N \sigma_{i\varepsilon}^4 \to \sigma_{4\varepsilon}$ as $N \to \infty$. (iv) $\frac{1}{N}\sum_{i=1}^N E(\varepsilon_{it}^4) \to m_4$ as $N \to \infty$.*

Assumption 2.1 states that the errors ε_{it} are independent over cross-section units and time and allowed to be heteroscedastic over cross-section units but not over time. Further, they are independent of the individual-specific term α_i and the initial value y_{i0}. The assumption about independency over time is stronger than the usual assumption about ε_{it} being serially uncorrelated. It is a simplifying assumption made in order to derive the asymptotic properties of the test statistics in Section 3. Assumption 2.2 states that the α_i's are i.i.d. across cross-section units and again it is made in order to simplify the derivation of the results in the next section. Note that the assumption that $E(\alpha_i) = 0$ means that we interpret the model in (2.1) as describing the behaviour of the observed variables after having subtracted the overall or the time-specific means. In practice, it means that as a starting point we subtract either the overall or the time-specific sample means from all observed variables. This type of transformation maintains the asymptotic properties of LS estimators and related statistics such that we can consider the model in (2.1) with i.i.d. mean zero terms as the starting point after having subtracted the cross-section sample means from all variables. A similar result is shown in detail in Madsen (2005) within the framework of a pure cross-section analysis. Assumption 2.3 specifies the initial values and implies that they are such that the time-series processes for y_{it} become mean stationary that is $E(y_{it}|\alpha_i) = \alpha_i$ for all $t = 0, 1, 2, \ldots$. It implies that it is possible to remove the individual-specific means from the observed variables by simple linear transformations. The parameter τ describes the dispersion of the initial deviation from the stationary level. If the initial values are such that the time-series processes are covariance stationary then $\tau = 1/(1 - \rho^2)$. This condition is only meaningful when $-1 < \rho < 1$. We see that as ρ approaches unity then the parameter τ tends to infinity such that all variables are dominated by the initial deviation from the individual-specific mean. In the next section we will formalize this property as it turns out to be important for the results in this paper. Note that ε_{it} is independent of ε_{i0} by Assumption 2.1. Finally, Assumption 2.4 is a technical assumption which enables us to derive the asymptotic properties of the statistics of interest by applying standard asymptotic theory. The assumption states that the innovations ε_{it} have uniformly bounded moments of order slightly greater than four and that the cross-section average of their variances, squared variances and fourth-order moments have well-defined limits as the cross-section dimension N tends to infinity. Note that when the errors ε_{it} are homoscedastic across units then $\sigma_{4\varepsilon} = \sigma_{2\varepsilon}^2$. Assumption 2.4(iv) is only required in relation to the test statistic suggested by Harris and Tzavalis (1999), as this is the only statistic of the ones considered in this paper which depends on fourth-order moments. Also note that $\sigma_{4\varepsilon} - \sigma_{2\varepsilon}^2 = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N (\sigma_{i\varepsilon}^2 - \sigma_{2\varepsilon})^2 \geq 0$ and $m_4 - \sigma_{4\varepsilon} = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N E((\varepsilon_{it}^2 - \sigma_{i\varepsilon}^2)^2) \geq 0$ such that $\sigma_{2\varepsilon}^2 \leq \sigma_{4\varepsilon} \leq m_4$.

3. THE TEST STATISTICS AND THEIR ASYMPTOTIC PROPERTIES

We consider the testing problem where the null hypothesis and the alternative hypothesis are given by

$$H_0 : \rho = 1 \qquad H_A : |\rho| < 1. \tag{3.1}$$

In the following, we consider local alternatives where ρ is modelled as being local-to-unity. More specifically, we consider local-to-unity sequences for ρ defined by

$$\rho_N = 1 - \frac{c}{N^k} \quad \text{for } k, c > 0. \tag{3.2}$$

This means that as the sample size N increases, the value of the parameter ρ is in an N^{-k} neighbourhood of unity. So instead of deriving asymptotic representations based on ρ being constant as N increases we derive asymptotic representations based on $c = (1 - \rho_N)N^k$ being constant as N increases. The idea is that these representations will provide good approximations to the actual distributions of the relevant statistics. With one exception the LS estimators of ρ considered in this paper converge weakly to normal distributions at the rate $\sqrt{N}$ and therefore we consider local-to-unity sequences for ρ with $k = \frac{1}{2}$. In one situation, the LS estimator must be normalized differently in order to converge weakly to a non-degenerate distribution under the local alternative and the local-to-unity sequence is defined accordingly. Note that $c = 0$ corresponds to the null hypothesis of ρ being unity.

In the next section we show that the local power of the different unit root tests that we consider depend on c and possibly the other parameters in the model. For a test that has non-trivial power against a local alternative where $\rho_N = 1 - c/\sqrt{N}$ this means that when reducing $(1 - \rho)$ by half (for example, from $\rho = 0.950$ to $\rho = 0.975$) the number of observations in the cross-section dimension N must be four times as large in order to attain the same level of local power. On the other hand, when a test has power against a local alternative where $\rho_N = 1 - c/N$ this means that when reducing $(1 - \rho)$ by half the number of observations in the cross-section dimension N only has to be twice as large in order to attain the same level of local power.

It turns out that the assumption being made about the variation of the initial deviation from the mean stationary level is crucial for the limiting distributions of the different statistics under the local alternative defined by (3.2). We consider the following two situations:

$$(i): \quad \tau \text{ is fixed,} \tag{3.3}$$

$$(ii): \quad \tau = \kappa \frac{1}{1 - \rho^2} \quad \text{for } \kappa > 0. \tag{3.4}$$

(i) means that the initial deviation from the stationary level is described by a parameter τ that remains constant as ρ approaches unity. (ii) means that the variance of the initial deviation from the stationary level is proportional to the variance of the autoregressive process. The specification in (ii) contains the case where the time-series processes are covariance stationary ($\kappa = 1$) which in particular implies that the variances of the observed variables are constant over time. In (ii) τ depends on ρ and goes to infinity as ρ approaches unity and it is not defined for ρ equal to unity. This means that the two formulations are fundamentally different.

Under the local-to-unity sequence for ρ given by $\rho_N = 1 - c/N^k$ the formulations in (3.3)–(3.4) correspond to

$$(i):\quad \tau_N = \tau \quad \text{for } \tau \geq 0, \tag{3.5}$$

$$(ii):\quad \tau_N = bN^k + o(N^k) \quad \text{for } k, b > 0. \tag{3.6}$$

Note that (3.6) is more general than (3.4) since b and c might take values independently of each other. Equation (3.4) corresponds to (3.6) with $b = \kappa/(2c)$, see Lemma A.1 in Section A.1 of the Appendix, implying that the parameters b and c are not independent of each other but are on a specific path in the parameter space. In (3.5) τ is a fixed parameter such that the term $\sqrt{\tau}\varepsilon_{i0}$ is of the same order of magnitude as the remaining terms in the expression for the variables y_{it}. On the other hand, in (3.6) this term dominates the behaviour of the variables y_{it} asymptotically as N tends to infinity since then we have $\sqrt{\tau}\varepsilon_{i0} = O_P(N^{k/2})$. The interpretation is that the behaviour of the observed variables $y_{i0}, \ldots, y_{iT}$ is dominated by the initial deviation from the mean stationary level $(y_{i0} - \alpha_i) = \sqrt{\tau}\varepsilon_{i0}$. In a time-series framework the assumption about the initial values being such that the time series become covariance stationary seems very natural as it implies that the initial values are of the same order of magnitude as the remaining term describing the observed variables as the number of observations over time goes to infinity. In a panel data framework this is not the case since it implies that the initial values are of a higher order of magnitude. This also means that the results about how the initial values affect the test statistics in single time series might not carry over to macro- and micropanels.

In this paper, we will focus on the cases where τ is fixed (corresponding to $b = 0$) and covariance stationarity (corresponding to $b = 1/(2c)$). It could be the case that $b = \kappa/(2c)$ (the variance of the initial values is proportional to the variance of the autoregressive process) or $b > 0$ and independent of c (the variance of the initial values is very high but does not depend on the value of the autoregressive parameter) and we shortly discuss how this affects our results.

3.1. OLS

The equation in (2.1) can be rewritten as the following regression model:

$$\begin{aligned} y_{it} &= \rho y_{it-1} + v_{it} \\ v_{it} &= (1-\rho)\alpha_i + \varepsilon_{it} \end{aligned} \qquad \text{for } i = 1, \ldots, N \text{ and } t = 1, \ldots, T.$$

The OLS estimator of the autoregressive parameter ρ is defined by

$$\hat{\rho}_{\text{OLS}} = \left(\sum_{i=1}^{N} y_{i,-1}' y_{i,-1} \right)^{-1} \left(\sum_{i=1}^{N} y_{i,-1}' y_i \right), \tag{3.7}$$

where $y_i = (y_{i1}, \ldots, y_{iT})'$ and $y_{i,-1} = (y_{i0}, \ldots, y_{iT-1})'$. The estimator is consistent when $\rho = 1$ whereas inconsistent when $|\rho| < 1$. In the latter case, the inconsistency is attributable to the term α_i which appears in both the regressor y_{it-1} and the regression error v_{it}. As α_i appears with the factor $(1 - \rho)$ in v_{it} the covariance between the regressor and the regression error is positive and decreases towards zero as ρ approaches unity. Now the regressor y_{it-1} can be expressed as the sum of the two independent terms α_i and $(y_{it-1} - \alpha_i)$ which are the stationary level and the deviation from the stationary level, respectively. If the variability of the two terms are of similar order as ρ approaches unity, the asymptotic bias of $\hat{\rho}_{\text{OLS}}$ is positive and decreases towards zero as ρ approaches unity. This describes the situation where the variance of the initial deviation from the mean stationary level is fixed. On the other hand, if the behaviour of y_{it-1} is dominated by the term $(y_{it-1} - \alpha_i)$ as ρ approaches unity, the asymptotic bias of $\hat{\rho}_{\text{OLS}}$ will be zero when ρ approaches unity. This describes the situation where the initial values are such that the time-series processes become covariance stationary.

The discussion above is formalized by the results given in Proposition 3.1 below. The proposition provides the limiting distribution of the OLS estimator $\hat{\rho}_{\text{OLS}}$ under both the null hypothesis when ρ is unity and local alternatives when ρ is local-to-unity. We consider different local alternatives depending on the assumption being made about the initial values as given by equations (3.5)–(3.6).

PROPOSITION 3.1. *Under Assumptions 2.1–2.4 and the local-to-unity sequence for ρ given by $\rho_N = 1 - c/\sqrt{N}$ for $c \geq 0$ and when τ is fixed, the limiting distribution of the OLS estimator $\hat{\rho}_{\text{OLS}}$ is given by*

$$\sqrt{N}\left(\hat{\rho}_{\text{OLS}} - \rho_N\right) \xrightarrow{w} N\left(c\frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \left(\tau + \frac{T-1}{2}\right)\sigma_{2\varepsilon}}, \frac{1}{T}\frac{\sigma_\alpha^2\sigma_{2\varepsilon} + \left(\tau + \frac{T-1}{2}\right)\sigma_{4\varepsilon}}{\left(\sigma_\alpha^2 + \left(\tau + \frac{T-1}{2}\right)\sigma_{2\varepsilon}\right)^2}\right) \quad \text{as } N \to \infty. \tag{3.8}$$

Under Assumptions 2.1–2.4 and the local-to-unity sequence for ρ given by $\rho_N = 1 - c/N$ for $c \geq 0$ and when $\tau_N = bN + o(N)$ for $b > 0$, the limiting distribution of the OLS estimator $\hat{\rho}_{\text{OLS}}$ is given by

$$N\left(\hat{\rho}_{\text{OLS}} - \rho_N\right) \xrightarrow{w} N\left(0, \frac{1}{T}\frac{\sigma_{4\varepsilon}}{b\sigma_{2\varepsilon}^2}\right) \quad \text{as } N \to \infty. \tag{3.9}$$

The proposition shows that in the unit root case when $c = 0$ and τ is fixed, the estimator $\hat{\rho}_{\text{OLS}}$ is $\sqrt{N}$-consistent and its limiting variance is decreasing in τ, T and $\sigma_{2\varepsilon}^2/\sigma_{4\varepsilon}$. Under the local alternative the estimator $\hat{\rho}_{\text{OLS}}$ has an asymptotic bias of order $1/\sqrt{N}$ which is always positive and increasing in c and $\sigma_\alpha^2/\sigma_{2\varepsilon}$ and decreasing in τ and T. The limiting variance of $\hat{\rho}_{\text{OLS}}$ is decreasing in τ, T and $\sigma_{2\varepsilon}^2/\sigma_{4\varepsilon}$ and increasing in $\sigma_\alpha^2/\sigma_{2\varepsilon}$ and does not depend on the location parameter c. On the other hand, when the variables are dominated by the initial deviation from the mean stationary level the estimator $\hat{\rho}_{\text{OLS}}$ is N-consistent for all values of c. In this case $\hat{\rho}_{\text{OLS}}$ estimates the parameter ρ very precisely also when its true value is close to unity. Further, the limiting variance of $\hat{\rho}_{\text{OLS}}$ is decreasing in b, T and $\sigma_{2\varepsilon}^2/\sigma_{4\varepsilon}$. The covariance stationary local alternative corresponds to $b = 1/(2c)$ such that the limiting variance is increasing in c. This rather surprising result is explained as follows. Under this assumption about the initial values that in particular holds under covariance stationarity the behaviour of y_{it} for $t = 0, \ldots, T$ is dominated by the initial deviation from the stationary level $(y_{i0} - \alpha_i)$. More specifically, the variation of $(y_{i0} - \alpha_i)$ is of order N under the local-to-unity sequence for ρ given by $1 - c/N$ for $c > 0$, see the result in (3.6), whereas the variation of the remaining terms in y_{it} is bounded as N tends to infinity. This implies that the numerator in (3.7) must be normalized by N in order to converge in distribution and the denominator in (3.7) must be normalized by N^2 in order to converge in probability. The consistency is a result of the term $(y_{i0} - \alpha_i)$, which dominates the behaviour of the regressor, being independent of the term α_i. This indicates that the asymptotic representation in (3.9) is only appropriate when the variances of α_i and ε_{it} are much smaller than the variance of $(y_{i0} - \alpha_i)$. Once the variances are of similar magnitude, the asymptotic representation in (3.8) is expected to provide a better approximation to the actual distribution of $\hat{\rho}_{\mathrm{OLS}}$.

The unit root test based on the usual t-statistic is obtained by normalizing $(\hat{\rho}_{\mathrm{OLS}} - 1)$ appropriately. For this purpose we need a consistent estimator of the limiting variance of $\hat{\rho}_{\mathrm{OLS}}$ and we use White's heteroscedastic-consistent estimator; see White (1980). Under the covariance stationary local alternative this estimator must be normalized differently in order to be consistent. Letting $k = \frac{1}{2}$ and $k = 1$ refer to the situations where $\hat{\rho}_{\mathrm{OLS}}$ converges in distribution at the rate $\sqrt{N}$ and N, respectively, White's heteroscedastic-consistent estimator of the limiting variance of $\hat{\rho}_{\mathrm{OLS}}$ is given by the following expression:

$$\hat{V}_{\mathrm{OLS}}(k) = \left(\frac{1}{N^{2k}} \sum_{i=1}^{N} y_{i,-1}' y_{i,-1}\right)^{-1} \frac{1}{N^{2k}} \sum_{i=1}^{N} y_{i,-1}' \hat{v}_i \hat{v}_i' y_{i,-1} \left(\frac{1}{N^{2k}} \sum_{i=1}^{N} y_{i,-1}' y_{i,-1}\right)^{-1},$$

where the vector of residuals is $\hat{v}_i = y_i - \hat{\rho}_{\mathrm{OLS}} y_{i,-1}$. The t-statistic is then defined as

$$t_{\mathrm{OLS}} = \hat{V}_{\mathrm{OLS}}(k)^{-\frac{1}{2}} N^k (\hat{\rho}_{\mathrm{OLS}} - 1).$$

Note that $\hat{V}_{\mathrm{OLS}}(k)^{-\frac{1}{2}} N^k$ does not depend on k since the normalization factors cancel out. This means that the test statistic t_{OLS} and also asymptotic confidence intervals do not depend on the actual normalization. This is a desirable feature since we might not know which assumption is appropriate for the initial values. The proposition below provides the limiting distribution of the t-statistic.

PROPOSITION 3.2. *Under Assumptions 2.1–2.4 and the local-to-unity sequence for ρ given by $\rho_N = 1 - c/\sqrt{N}$ for $c \geq 0$ and when τ is fixed, the limiting distribution of the OLS t-statistic t_{OLS} is given by*

$$t_{\mathrm{OLS}} \xrightarrow{w} N\left(-c\left(\tau + \frac{T-1}{2}\right)\sqrt{\left(\frac{\sigma_\alpha^2}{\sigma_{2\varepsilon}} + \left(\tau + \frac{T-1}{2}\right)\frac{\sigma_{4\varepsilon}}{\sigma_{2\varepsilon}^2}\right)^{-1} T}, 1\right) \quad \text{as } N \to \infty. \tag{3.10}$$

Under Assumptions 2.1–2.4 and the local-to-unity sequence for ρ given by $\rho_N = 1 - c/N$ for $c \geq 0$ and when $\tau_N = bN + o(N)$ for $b > 0$, the limiting distribution of the OLS t-statistic t_{OLS} is given by

$$t_{\mathrm{OLS}} \xrightarrow{w} N\left(-c\sqrt{b\frac{\sigma_{2\varepsilon}^2}{\sigma_{4\varepsilon}} T}, 1\right) \quad \text{as } N \to \infty. \tag{3.11}$$

The proposition shows that in both cases under the null hypothesis of a unit root the t-statistic t_{OLS} is asymptotically standard normal. So unit root inference is carried out by employing critical values from the standard normal distribution. Furthermore, the proposition shows that when τ is fixed, the local power is increasing in c, τ, T and $\sigma_{2\varepsilon}^2/\sigma_{4\varepsilon}$ (the location parameter is shifted to the left-hand side when these parameters increase) and decreasing in $\sigma_\alpha^2/\sigma_{2\varepsilon}$ (the location parameter is shifted to the right-hand side when $\sigma_\alpha^2/\sigma_{2\varepsilon}$ increases). Under the covariance stationary alternative when $b = 1/(2c)$, the local power only depends on c, T and $\sigma_{2\varepsilon}^2/\sigma_{4\varepsilon}$ and is increasing in these parameters. When $b > 0$ and independent of the value of c we find that for a fixed value of c the local power is increasing in b. This result is similar to the finding in single time series where the local power of the Dickey–Fuller unit root test turns out to be an increasing function of the initial value and in addition is an optimal test when the initial values are high; see Müller and Elliott (2003). The reason why we obtain a similar result here for a test statistic, which is not invariant with respect to individual-specific constants, is that in a single time series the estimation of a constant does not affect the test statistics asymptotically as the time-series dimension goes to infinity. Note that as discussed above, the limiting distribution in (3.11) will only provide a good approximation to the actual distribution of the *t*-statistic when the behaviour of y_{it} is dominated by the initial deviation from the mean stationary level.

To explore the results of the proposition in more detail let us consider the case where $T + 1 = 5$ and the following values of the nuisance parameters: $\sigma^2_{2\varepsilon}/\sigma_{4\varepsilon} = 1$ and $\sigma^2_\alpha/\sigma_{2\varepsilon} = 1$. For $\tau = 1$ and $\rho = 0.95$ then N must be approximately 150 in order for the OLS unit root test to obtain a power level of 0.5 for a one-sided alternative at the nominal level of 5%. As explained in the previous section this immediately implies that for $\tau = 1$ and $\rho = 0.975$ then N must be approximately 600 in order to attain the power level of 0.5. The numbers for N in this example would be higher if $\sigma^2_\alpha/\sigma^2_\varepsilon > 1$. Under the covariance stationary alternative we find that N must only be approximately 30 and 60, respectively, in order for this test to attain the power level of 0.5 in the alternatives $\rho = 0.95$ and $\rho = 0.975$, respectively. So the test is very powerful against this alternative.

Altogether, the advantage of using the OLS unit root test is that it is expected to have high power under the covariance stationary alternative and when the variation in the initial deviation from the stationary levels are very high even for values of ρ very close to unity. However, if this is not the case the power of the test for values of ρ close to unity is expected to be low when $\sigma^2_\alpha/\sigma_{2\varepsilon}$ is high. This will be most evident for small values of T.

3.2. Breitung–Meyer

Subtracting the initial value y_{i0} from both sides of the equation in (2.1) yields the following regression model:

$$\begin{aligned} y_{it} - y_{i0} &= \rho\,(y_{it-1} - y_{i0}) + \tilde{v}_{it} \\ \tilde{v}_{it} &= (\rho - 1)\,(y_{i0} - \alpha_i) + \varepsilon_{it} \end{aligned} \qquad \text{for } i = 1, \ldots, N \text{ and } t = 1, \ldots, T.$$

The LS estimator of ρ obtained from this regression equation is defined by

$$\hat{\rho}_0 = \left(\sum_{i=1}^{N} \tilde{y}'_{i,-1}\tilde{y}_{i,-1}\right)^{-1}\left(\sum_{i=1}^{N} \tilde{y}'_{i,-1}\tilde{y}_i\right), \tag{3.12}$$

where $\tilde{y}_i = y_i - y_{i0}\iota_T$, $\tilde{y}_{i,-1} = y_{i,-1} - y_{i0}\iota_T$ and ι_T is a $T \times 1$ vector of ones. Again the estimator is consistent when $\rho = 1$, whereas inconsistent when $|\rho| < 1$. In the latter case, its asymptotic bias equals $\frac{1}{2}(1 - \rho)$ under the assumption about covariance stationarity; see Breitung and Meyer (1994). As an example, this means that the asymptotic bias equals 0.050, 0.025 and 0.005 when ρ equals 0.90, 0.95 and 0.99, respectively. The inconsistency is attributable to the term $(y_{i0} - \alpha_i)$ as it appears in both the regressor $(y_{it-1} - y_{i0})$ and the regression error $\tilde{v}_{it}$. The covariance between the regressor and the regression error decreases towards zero as ρ approaches unity when the variance of $(y_{i0} - \alpha_i)$ is kept constant. However, the decrease might be offset if the variance of $(y_{i0} - \alpha_i)$ increases as ρ approaches unity. This is exactly what happens when the initial values are such that the time-series processes become covariance stationary.

Proposition 3.3 below provides the limiting distribution of the Breitung–Meyer estimator $\hat{\rho}_0$ under both the null hypothesis when ρ is unity and the mean stationary local alternative when ρ is local-to-unity. In this case, the local alternatives are the same irrespective of the assumption about the dispersion of the initial deviation from the stationary level.

PROPOSITION 3.3. *Under Assumptions 2.1–2.4 and the local-to-unity sequence for ρ given by $\rho_N = 1 - c/\sqrt{N}$ for $c \geq 0$ and when $\tau_N = b\sqrt{N} + o(\sqrt{N})$ for $b \geq 0$, the limiting distribution of the Breitung–Meyer estimator $\hat{\rho}_0$ is given by*

$$\sqrt{N}\,(\hat{\rho}_0 - \rho_N) \overset{w}{\to} N\left(c^2 b, \frac{\sigma_{4\varepsilon}}{\sigma_{2\varepsilon}^2}\frac{2}{T(T-1)}\right) \text{ as } N \to \infty. \tag{3.13}$$

The proposition shows that in the unit root case and under the mean stationary local alternative when τ is fixed $\hat{\rho}_0$ is $\sqrt{N}$-consistent. Under the covariance stationary local alternative $\hat{\rho}_0$ has a positive asymptotic bias of order $1/\sqrt{N}$. The limiting variance of $\hat{\rho}_0$ does not depend on the assumption being made about the initial values and it is a simple function of T and $\sigma_{2\varepsilon}^2/\sigma_{4\varepsilon}$ and decreasing in both. As indicated above, the results follow by using that when the variance of $(y_{i0} - \alpha_i)$ is of order less than $\sqrt{N}$, the asymptotic bias disappears under the local alternative. This is the case when τ is fixed. On the contrary, under covariance stationarity this is not the case, as the variance of $(y_{i0} - \alpha_i)$ in this case is of order $\sqrt{N}$; see the result in (3.6).

As before, White's heteroscedastic-consistent estimator of the limiting variance of $\hat{\rho}_0$ is given by the following expression:

$$\hat{V}_0 = \left(\frac{1}{N}\sum_{i=1}^{N}\tilde{y}_{i,-1}'\tilde{y}_{i,-1}\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}\tilde{y}_{i,-1}'\hat{\tilde{v}}_i\hat{\tilde{v}}_i'\tilde{y}_{i,-1}\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{y}_{i,-1}'\tilde{y}_{i,-1}\right)^{-1},$$

where the vector of residuals is $\hat{\tilde{v}}_i = \tilde{y}_i - \hat{\rho}_0\tilde{y}_{i,-1}$. The t-statistic is then defined as

$$t_0 = \hat{V}_0^{-\frac{1}{2}}\sqrt{N}\,(\hat{\rho}_0 - 1)\,.$$

When the errors ε_{it} are homoscedastic across units such that $\sigma_{4\varepsilon} = \sigma_{2\varepsilon}^2$, the limiting variance of $\hat{\rho}_0$ is a function of T only. Therefore, it is possible to use a normalized coefficient statistic when testing the unit root hypothesis. The statistic is defined in the following way:

$$\bar{t}_0 = \sqrt{\frac{T(T-1)}{2}}\sqrt{N}\,(\hat{\rho}_0 - 1).$$

The proposition below provides the limiting distributions of the test statistics defined above.

PROPOSITION 3.4. *Under Assumptions 2.1–2.4 and the local-to-unity sequence for ρ given by $\rho_N = 1 - c/\sqrt{N}$ for $c \geq 0$ and when $\tau_N = b\sqrt{N} + o(\sqrt{N})$ for $b \geq 0$, the limiting distribution of the Breitung–Meyer t-statistic t_0 is given by*

$$t_0 \overset{w}{\to} N\left(-c\,(1 - cb)\sqrt{\frac{\sigma_{2\varepsilon}^2}{\sigma_{4\varepsilon}}\frac{T(T-1)}{2}}, 1\right) \text{ as } N \to \infty. \tag{3.14}$$

The limiting distribution of the normalized coefficient statistic $\bar{t}_0$ is given by

$$\bar{t}_0 \xrightarrow{w} N\left(-c(1-cb)\sqrt{\frac{T(T-1)}{2}}, \frac{\sigma_{4\varepsilon}}{\sigma_{2\varepsilon}^2}\right) \quad \text{as } N \to \infty. \tag{3.15}$$

The proposition shows that under the null hypothesis of a unit root, the t-statistic t_0 is asymptotically standard normal. So again unit root inference is carried out by employing critical values from the standard normal distribution. Further, the proposition shows that the local power of the test is increasing in c, T and $\sigma_{2\varepsilon}^2/\sigma_{4\varepsilon}$ when $b < 1/c$. In particular, this means that the local power is monotonically increasing in c when $b = 0$ (τ is fixed) and when $b = 1/(2c)$ (covariance stationarity). When $\sigma_{2\varepsilon}^2/\sigma_{4\varepsilon} = 1$ the local power of a one-sided test at the nominal 5% level against these two alternatives is given by $\Phi(-1.645 + c\sqrt{T(T-1)/2})$ and $\Phi(-1.645 + c/2\sqrt{T(T-1)/2})$, respectively, where $\Phi(\cdot)$ denotes the cdf of the standard normal distribution. This means that for a specific value of ρ we need four times as many cross-section observations in the covariance stationary alternative as in the fixed τ alternative in order to obtain the same level of local power. When $b > 1/c$ then the local power of the test is monotonically decreasing in c and the local power is in fact less than the nominal size of the test for all values of c. This would be the case when the variance of the initial deviation from the stationary level is more than two times $\sigma_{i\varepsilon}^2/(1-\rho^2)$. Also we see that for a fixed value of c the local power is decreasing in b. These findings are similar to the results in Müller and Elliott (2003) and Harris et al. (2008) for unit root tests in single time series and macropanels, respectively. Note that for a fixed value of $b > 0$ which is not linked to the value of c then the local power is a non-monotonic function of c that tends to zero as c goes to infinity. Figure 1 shows the local power as a function of c for different values of b when $T + 1 = 5$ and $\sigma_{2\varepsilon}^2/\sigma_{4\varepsilon} = 1$ and we see the findings described above. Also note that there is a power loss associated with having cross-sectional heterogeneity in the error terms since a test based on $y_{it}/\sigma_{i\varepsilon}$ would have higher power.

The test based on the normalized coefficient statistic $\bar{t}_0$ is asymptotically equivalent to the test based on the t-statistic t_0 when $\sigma_{4\varepsilon} = \sigma_{2\varepsilon}^2$. When this is not the case, the test based on the normalized coefficient statistic will be distorted when employing critical values from the standard normal distribution. In a one-sided test it will reject the null hypothesis of a unit root too often since $\sigma_{2\varepsilon}^2 < \sigma_{4\varepsilon}$. So unless there is any prior knowledge about the error terms being homoscedastic over cross-section units the unit root test should be based on the t-statistic. Note, that if $\sigma_{2\varepsilon}^2 < \sigma_{4\varepsilon}$ such that there is a difference between the local power of the two tests, this difference decreases as T increases. However, the size distortion is not affected by T and hence it remains as T increases.

The advantage of using the Breitung–Meyer unit root test is that the local power only depends on one nuisance parameter. Further, under mean stationarity the test is invariant with respect to the individual-specific levels. This means that the size of the test is invariant with respect to the initial values and the power of the test is invariant with respect to the individual-specific term α_i. On the other hand, the test is sensitive to the assumptions on the initial values through the initial deviation from the stationary level and the local power can be quite low if the variation in this term is very high. This is in contrast to the OLS unit root test where we found the opposite result.

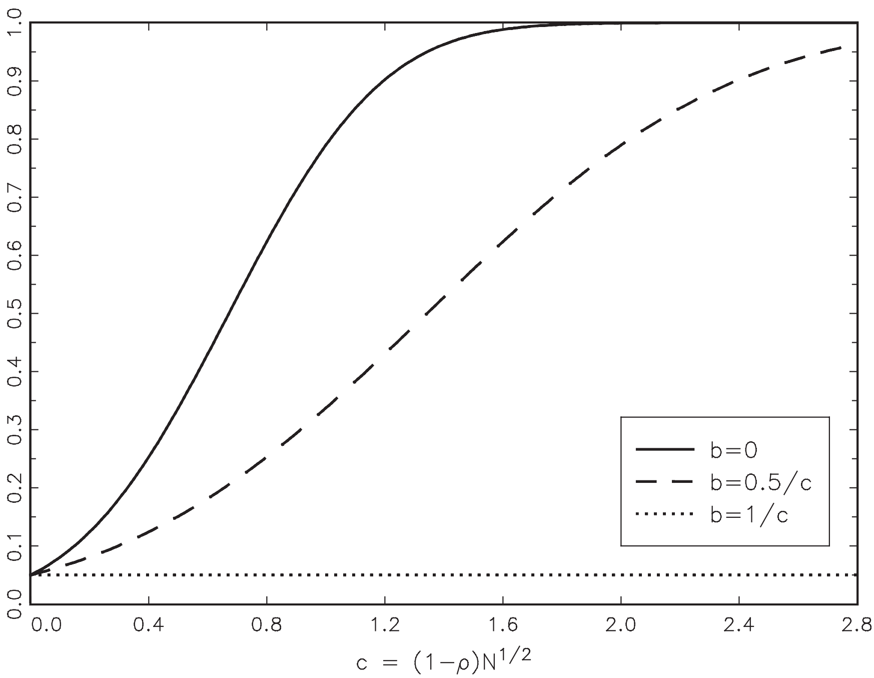

Figure 1. The local power of the Breitung–Meyer unit root test.

3.3. *Harris–Tzavalis*

The within-group transformation of the original model is obtained by subtracting the individual time series means from the variables in equation (2.1). This yields the following regression model:

$$\begin{aligned} y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it} &= \rho\left(y_{it-1} - \frac{1}{T}\sum_{t=1}^{T} y_{it-1}\right) + w_{it} \\ w_{it} &= \varepsilon_{it} - \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it} \end{aligned} \qquad \text{for } i = 1, \ldots, N \text{ and } t = 1, \ldots, T.$$

The within-group estimator of ρ is then defined by

$$\hat{\rho}_{WG} = \left(\sum_{i=1}^{N} y'_{i,-1} Q_T y_{i,-1}\right)^{-1} \left(\sum_{i=1}^{N} y'_{i,-1} Q_T y_i\right), \tag{3.16}$$

where Q_T is a $T \times T$ symmetric and idempotent matrix defined as $Q_T = I_T - \frac{1}{T}\iota_T\iota'_T$, where I_T is the $T \times T$ identity matrix and $\iota_T\iota'_T$ is a $T \times T$ matrix of ones. It is well-known that this estimator is inconsistent when $-1 < \rho < 1$. The asymptotic bias is often referred to as the Nickell-bias since Nickell (1981) is the first to provide an analytical expression for it. Under the assumption about the time-series processes being covariance stationary, the asymptotic bias is a function of ρ and T which is always negative when $0 < \rho < 1$ and decreases numerically as T increases. Harris and Tzavalis (1999) show that the asymptotic bias of the within-group estimator equals $-3/(T+1)$ when $\rho = 1$. As this expression does not depend on any nuisance parameters, their idea is to base a unit root test on the bias adjusted within-group estimator. Proposition 3.5 below provides the limiting distribution of $\hat{\rho}_{WG}$ under both the null hypothesis of a unit root and the mean stationary local alternative when ρ is local-to-unity.

PROPOSITION 3.5. *Under Assumptions 2.1–2.4 and the local-to-unity sequence for ρ given by $\rho_N = 1 - c/\sqrt{N}$ for $c \geq 0$ and when $\tau_N = b\sqrt{N} + o(\sqrt{N})$ for $b \geq 0$, the limiting distribution of the adjusted within-group estimator $\hat{\rho}_{WG}$ is given by*

$$\sqrt{N}\left(\hat{\rho}_{WG} - \rho_N + \frac{3}{T+1}\right) \xrightarrow{w} N\left(-c\frac{T-2}{2(T+1)} + c^2 b\frac{3T}{2(T+1)}, \frac{k_1 m_4 + k_2 \sigma_{4\varepsilon}}{\sigma_{2\varepsilon}^2}\right) \quad \text{as } N \to \infty, \tag{3.17}$$

where

$$k_1 = \frac{12(T-2)(2T-1)}{5T(T-1)(T+1)^3} \qquad k_2 = \frac{3(17T^3 - 44T^2 + 77T - 24)}{5T(T-1)(T+1)^3}.$$

The proposition shows that except in the unit root case, the adjusted within-group estimator has an asymptotic bias of order $1/\sqrt{N}$ under the local alternative. The bias is negative when τ is fixed and positive under covariance stationarity. This means that the adjustment is respectively too big and too small. The limiting variance of $\hat{\rho}_{WG}$ is the same in the unit root case and under the local alternatives. It depends on fourth-order moments of the errors ε_{it} through the term m_4. As $k_1 < k_2$ the fourth-order moments receive less weight than the squared second-order moments.

Harris and Tzavalis (1999) assume that the errors ε_{it} are i.i.d. normally distributed across i such that $\sigma_{4\varepsilon} = \sigma_{2\varepsilon}^2$ and $m_4 = 3\sigma_{2\varepsilon}^2$. In this case, the limiting variance of $\hat{\rho}_{WG}$ only depends on T and is given by the following expression:

$$\tilde{V}_{WG} = 3k_1 + k_2 = \frac{3(17T^2 - 20T + 17)}{5(T-1)(T+1)^3}.$$

Therefore, Harris and Tzavalis (1999) suggest using the normalized coefficient statistic as a unit root test statistic. It is defined as follows:

$$\bar{t}_{WG} = \tilde{V}_{WG}^{-\frac{1}{2}}\sqrt{N}\left(\hat{\rho}_{WG} - 1 + \frac{3}{T+1}\right).$$

However, as before it is also possible to use the usual t-statistic as a test statistic. White's heteroscedasticity-consistent estimator of the limiting variance of the bias adjusted within-group estimator is given by the following expression:

$$\hat{V}_{WG} = \left(\frac{1}{N}\sum_{i=1}^{N} y_{i,-1}' Q_T y_{i,-1}\right)^{-1} \frac{1}{N}\sum_{i=1}^{N} y_{i,-1}' Q_T \hat{\omega}_i \hat{\omega}_i' Q_T y_{i,-1} \left(\frac{1}{N}\sum_{i=1}^{N} y_{i,-1}' Q_T y_{i,-1}\right)^{-1},$$

where the vector of residuals is $\hat{w}_i = Q_T y_i - \hat{\rho}_{WG} Q_T y_{i,-1}$. The bias-adjusted within-group t-statistic is then defined in the following way:

$$t_{WG} = \hat{V}_{WG}^{-\frac{1}{2}} \sqrt{N} \left(\hat{\rho}_{WG} - 1 + \frac{3}{T+1} \right).$$

The limiting distributions of these test statistics are given in Proposition 3.6 below.

PROPOSITION 3.6. *Under Assumptions 2.1–2.4 and the local-to-unity sequence for ρ given by $\rho_N = 1 - c/\sqrt{N}$ for $c \geq 0$ and when $\tau_N = b\sqrt{N} + o(\sqrt{N})$ for $b \geq 0$, the limiting distribution of the adjusted within-group t-statistic t_{WG} is given by*

$$t_{WG} \xrightarrow{w} N\left(-c\,(1-cb)\,\frac{3T}{2\,(T+1)} \sigma_{2\varepsilon} \sqrt{(k_1 m_4 + k_2 \sigma_{4\varepsilon})^{-1}}, 1 \right) \quad as\ N \to \infty. \tag{3.18}$$

The limiting distribution of the Harris–Tzavalis normalized coefficient statistic $\bar{t}_{WG}$ is given by

$$\bar{t}_{WG} \xrightarrow{w} N\left(-c\,(1-cb)\,\frac{3T}{2\,(T+1)} \sqrt{\frac{5(T-1)(T+1)^3}{3\left(17T^2 - 20T + 17\right)}}, \frac{\tilde{k}_1 m_4 + \tilde{k}_2 \sigma_{4\varepsilon}}{\sigma_{2\varepsilon}^2} \right) \quad as\ N \to \infty, \tag{3.19}$$

where

$$\tilde{k}_1 = \frac{4(T-2)(2T-1)}{T(17T^2 - 20T + 17)} \qquad \tilde{k}_2 = \frac{17T^3 - 44T^2 + 77T - 24}{T(17T^2 - 20T + 17)}.$$

Once again unit root inference based on the adjusted t-statistic t_{WG} can be carried out by employing critical values from the standard normal distribution. We also note that the parameters c and b appear in a similar manner as in the limiting distribution of Breitung–Meyer test statistic and therefore the results here are similar to the results in Section 3.2. In particular, we find that the local power of the Harris–Tzavalis test is increasing in c, T, $\sigma_{2\varepsilon}^2/\sigma_{4\varepsilon}$ and $\sigma_{2\varepsilon}^2/m_4$ when $b < 1/c$. This means that the local power is monotonically increasing in c when $b = 0$ (τ is fixed) and when $b = 1/(2c)$ (covariance staionarity). Also as in Section 3.2 the location parameter in the first case is twice as large as in the second case such that four times as many cross-section observations are necessary when $b = 1/(2c)$ in order to obtain the same level of local power as when $b = 0$ for a specific value of ρ. The unit root test based on the Harris–Tzavalis normalized coefficient statistic $\bar{t}_{WG}$ is asymptotically equivalent to the test based on the t-statistic t_{WG} when the errors ε_{it} are normally distributed and homoscedastic across units. If at least one of these assumptions is violated, the test is likely to be distorted when employing critical values from the standard normal distribution. The test will reject the null hypothesis too often when $\sigma_{2\varepsilon}^2 < \sigma_{4\varepsilon}$ and when the excess kurtosis of ε_{it} is positive, i.e. $m_4 > 3\sigma_{2\varepsilon}^2$. Therefore, the Harris–Tzavalis normalized coefficient statistic should not be used for unit root inference unless the underlying assumptions have been verified.

As with the Breitung–Meyer unit root test, the Harris–Tzavalis unit root test is invariant with respect to the individual-specific levels. However, the local power of the Harris–Tzavalis test depends on more nuisance parameters. A more serious disadvantage of this test is that the bias adjustment of the within-group estimator $\hat{\rho}_{WG}$ depends crucially on the errors ε_{it} being homoscedastic over time. If this assumption is violated, the Harris–Tzavalis unit root test is likely to be distorted. To avoid this problem, Kruiniger and Tzavalis (2001) suggest using an estimator of the asymptotic bias in the adjustment of $\hat{\rho}_{WG}$. In the unit root case, the estimator of the asymptotic bias is consistent. However, in this paper we only investigate the performance of the unit root tests when the errors ε_{it} are homoscedastic over time. Therefore, we do not consider this different bias adjustment in detail but we note that it is available.

3.4. *Comparison of the tests*

Below we list the main findings about the local power of the tests. They follow immediately from the results in Proposition 3.2, 3.4 and 3.6.

(1) When $b \leq 1/c$ the local power of the Breitung–Meyer test is always higher than the local power of the Harris–Tzavalis test. This follows by using that $\sigma_{4\varepsilon} \leq m_4$ such that $(k_1 m_4 + k_2 \sigma_{4\varepsilon})^{-1} \leq \sigma_{4\varepsilon}^{-1}(k_1 + k_2)^{-1} = \sigma_{4\varepsilon}^{-1} \frac{5T(T-1)(T+1)^3}{51T^3-108T^2+171T-48} \leq \sigma_{4\varepsilon}^{-1} \frac{5T(T-1)(T+1)^3}{51T^3-108T^2+171T-48}$. This gives $(\frac{3T}{2(T+1)})^2 (k_1 m_4 + k_2 \sigma_{4\varepsilon})^{-1} \leq \sigma_{4\varepsilon}^{-1} \frac{T(T-1)}{2} \cdot \frac{45T^2(T+1)}{2(51T^3-108T^2+171T-48)} \leq \sigma_{4\varepsilon}^{-1} \frac{T(T-1)}{2}$ since $0 < \frac{45T^2(T+1)}{2(51T^3-108T^2+171T-48)} < 1$.

(2) When τ is fixed the local power of the OLS test is higher than the local power of the Breitung–Meyer test when $\frac{\sigma_\alpha^2}{\sigma_{2\varepsilon}} < \frac{\sigma_{4\varepsilon}}{\sigma_{2\varepsilon}^2} \tau (1 + \frac{2\tau}{T-1})$.

(3) When $\tau = \frac{1}{1-\rho^2}$ (covariance stationarity) the local power of the OLS test is higher than the local power of the Breitung–Meyer test when $\rho > \frac{T-5}{T-1}$.

(4) When τ is fixed, $\sigma_{2\varepsilon}^2 / \sigma_{4\varepsilon} = 1$ and $m_4 = 3\sigma_{2\varepsilon}^2$ the local power of the OLS test is higher than the local power of the Harris–Tzavalis test when $\frac{\sigma_\alpha^2}{\sigma_{2\varepsilon}} < (\frac{4(17T^2-20T+17)}{15T(T-1)(T+1)}(\tau + \frac{T-1}{2}) - 1)(\tau + \frac{T-1}{2})$.

Figure 2 below illustrates some of these results. In each figure, the local power of one-sided tests at the 5% nominal level based on the t-statistics is graphed as a function of $c = (1-\rho)\sqrt{N}$. It is calculated for the following parameter values: $\tau = 1$, $\sigma_{2\varepsilon}^2 / \sigma_{4\varepsilon} = 1$ and $m_4 = 3\sigma_{2\varepsilon}^2$. The figures correspond to the value of $T + 1$ being 5 or 10 and the value of $\sigma_\alpha^2 / \sigma_{2\varepsilon}$ being 1 or 10. For this choice of parameters, the local power of the Breitung–Meyer test and the Harris–Tzavalis test only depends on T. As an example, the local power of the Breitung–Meyer test is obtained as $\Phi(-1.645 + c\sqrt{T(T-1)/2})$ where Φ denotes the cdf of the standard normal; see Proposition 3.4. The figures show that the local power of the Breitung–Meyer test is higher than the local power of the Harris–Tzavalis test for all values of c. When $\sigma_\alpha^2 / \sigma_{2\varepsilon} = 1$ the local power of the OLS test is highest for all values of c, whereas when $\sigma_\alpha^2 / \sigma_{2\varepsilon} = 10$ the local power of the OLS test is lowest for all values of c. More specifically, when $\sigma_\alpha^2 / \sigma_{2\varepsilon} = 1$ and $T + 1 = 5$ then in order to attain a local power level equal to 0.5 for a given value of ρ we need approximately 1.2 (Breitung–Meyer test) and 1.6 (Harris–Tzavalis test) as many cross-section observations in order to do so compared to when using the OLS test. When $\sigma_\alpha^2 / \sigma_{2\varepsilon} = 10$ and $T + 1 = 5$ we need approximately 1.4 (Harris–Tzavalis test) and 3.0 (OLS test) as many cross-section observations compared to when using the Breitung–Meyer test in order to attain a local power level of 0.5.

4. SIMULATION EXPERIMENTS

In this section the analytical results obtained in Section 3 are illustrated in a simulation experiment. The simulated model is the following:

$$y_{i0} = \alpha_i + \varepsilon_{i0},$$
$$y_{it} = \rho y_{it-1} + (1-\rho)\alpha_i + \varepsilon_{it},$$

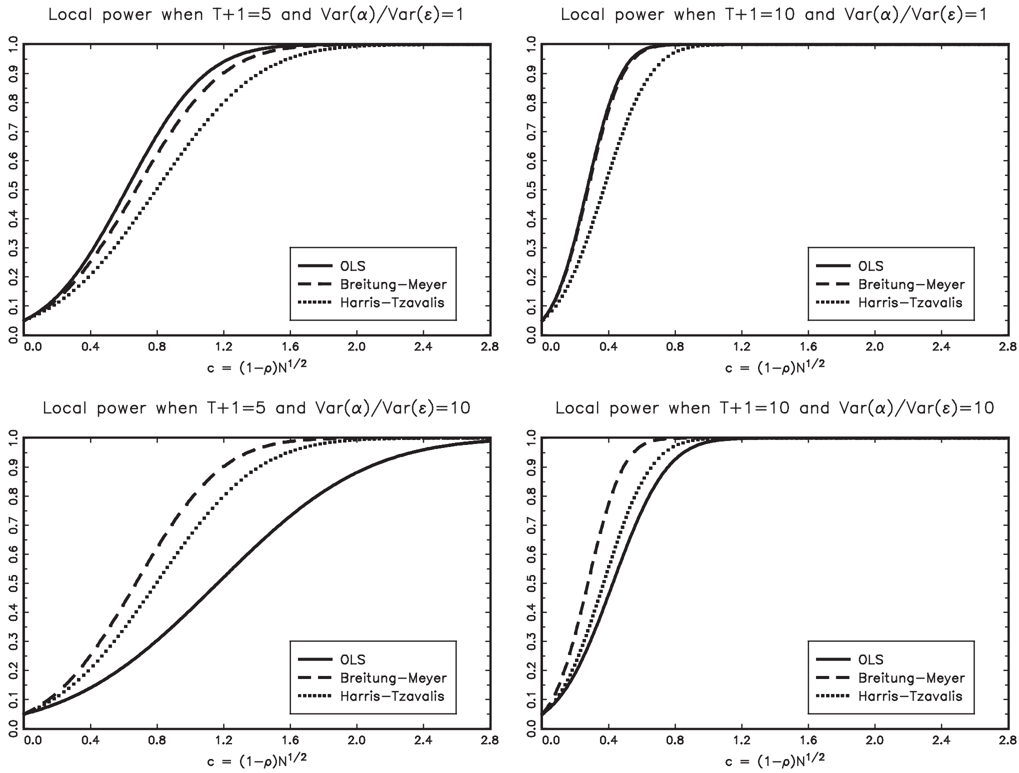

Figure 2. Comparison of the local power under mean stationarity.

with

$$\varepsilon_{it} \sim \text{i.i.d.}N\,(0,\,1) \qquad \alpha_i \sim \text{i.i.d.}N\left(0,\,\sigma_\alpha^2\right) \qquad \varepsilon_{i0} \sim \text{i.i.d.}N\,(0,\,\tau)\,.$$

We consider different values of T, N and ρ which are $T+1 = 5, 10, 15$, $N = 100, 250, 500$ and $\rho = 0.90, 0.95, 0.99, 1.00$. The results are based on 5000 replications of the model. In Tables 1 and 2, we report the empirical rejection probabilities of one-sided unit root tests based on the t-statistics with the critical value taken from the standard normal distribution at the nominal 5% significance level. For comparison the analytical rejection probabilities (i.e. the local power) are reported in parentheses. We consider different simulation set-ups where the value of σ_α^2 is either 1 or 10. This parameter will only affect the OLS test as the two other tests do not depend on this parameter under the alternatives considered here. Further, the simulation set-ups depend on the variance of initial error term τ. Table 1 corresponds to the unit root case with $\tau = 1$ and Table 2 corresponds to the covariance stationary alternative with $\tau = 1/(1-\rho^2)$.

In Table 1, we see that the empirical size of all tests is close to the nominal size of 0.05 and the empirical power is quite high even for values of ρ close to unity such as $\rho = 0.95$. Further, the increase in power can be quite dramatic when increasing $T+1$ from 5 to 10. For example, when $\rho = 0.99$ and $N = 500$ the power of the Breitung–Meyer test increases from 0.15 to 0.37, the power of the Harris–Tzavalis test increases from 0.13 to 0.25, and the power of the OLS test increases from 0.15 to 0.38 when $\sigma_\alpha^2 = 1$ and from 0.09 to 0.21 when $\sigma_\alpha^2 = 10$. When comparing

Table 1. Empirical and analytical (in brackets) rejection probabilities when $\tau = 1$.

ρ	$T+1$	N	OLS, $\sigma_\alpha^2 = 1$	OLS, $\sigma_\alpha^2 = 10$	Breitung–Meyer	Harris–Tzavalis
0.900	5	100	0.803 (0.848)	0.370 (0.409)	0.684 (0.790)	0.562 (0.667)
0.900	5	250	0.989 (0.995)	0.666 (0.723)	0.957 (0.987)	0.865 (0.949)
0.900	5	500	1.000 (1.000)	0.896 (0.935)	0.999 (1.000)	0.991 (0.999)
0.900	10	100	1.000 (1.000)	0.943 (0.987)	0.999 (1.000)	0.947 (0.998)
0.900	10	250	1.000 (1.000)	0.999 (1.000)	1.000 (1.000)	1.000 (1.000)
0.900	10	500	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)
0.900	15	100	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	0.998 (1.000)
0.900	15	250	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)
0.900	15	500	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)
0.950	5	100	0.369 (0.379)	0.184 (0.174)	0.318 (0.337)	0.277 (0.272)
0.950	5	250	0.650 (0.680)	0.282 (0.299)	0.572 (0.615)	0.463 (0.498)
0.950	5	500	0.887 (0.910)	0.452 (0.475)	0.823 (0.863)	0.694 (0.750)
0.950	10	100	0.880 (0.922)	0.567 (0.615)	0.855 (0.912)	0.620 (0.723)
0.950	10	250	0.998 (0.999)	0.886 (0.922)	0.997 (0.999)	0.910 (0.971)
0.950	10	500	1.000 (1.000)	0.989 (0.996)	1.000 (1.000)	0.995 (1.000)
0.950	15	100	0.996 (0.999)	0.891 (0.956)	0.993 (0.999)	0.856 (0.962)
0.950	15	250	1.000 (1.000)	0.999 (1.000)	1.000 (1.000)	0.996 (1.000)
0.950	15	500	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)
0.990	5	100	0.097 (0.084)	0.078 (0.066)	0.094 (0.081)	0.092 (0.075)
0.990	5	250	0.115 (0.111)	0.081 (0.078)	0.111 (0.104)	0.103 (0.094)
0.990	5	500	0.150 (0.148)	0.091 (0.092)	0.145 (0.136)	0.128 (0.119)
0.990	10	100	0.171 (0.151)	0.119 (0.104)	0.157 (0.148)	0.131 (0.116)
0.990	10	250	0.257 (0.249)	0.155 (0.151)	0.242 (0.243)	0.178 (0.174)
0.990	10	500	0.380 (0.391)	0.213 (0.218)	0.367 (0.381)	0.253 (0.260)
0.990	15	100	0.253 (0.248)	0.175 (0.165)	0.253 (0.245)	0.177 (0.168)
0.990	15	250	0.434 (0.451)	0.275 (0.280)	0.426 (0.446)	0.273 (0.286)
0.990	15	500	0.677 (0.694)	0.419 (0.442)	0.665 (0.687)	0.418 (0.454)
1.000	5	100	0.057 (0.050)	0.057 (0.050)	0.062 (0.050)	0.063 (0.050)
1.000	5	250	0.054 (0.050)	0.054 (0.050)	0.054 (0.050)	0.059 (0.050)
1.000	5	500	0.055 (0.050)	0.055 (0.050)	0.055 (0.050)	0.057 (0.050)
1.000	10	100	0.064 (0.050)	0.064 (0.050)	0.062 (0.050)	0.064 (0.050)
1.000	10	250	0.056 (0.050)	0.056 (0.050)	0.060 (0.050)	0.061 (0.050)
1.000	10	500	0.055 (0.050)	0.055 (0.050)	0.049 (0.050)	0.056 (0.050)
1.000	15	100	0.055 (0.050)	0.055 (0.050)	0.058 (0.050)	0.063 (0.050)
1.000	15	250	0.058 (0.050)	0.058 (0.050)	0.053 (0.050)	0.056 (0.050)
1.000	15	500	0.050 (0.050)	0.050 (0.050)	0.048 (0.050)	0.053 (0.050)

Table 2. Empirical and analytical (in brackets) rejection probabilities when $\tau = 1/(1-\rho^2)$.

ρ	$T+1$	N	OLS, $\sigma_\alpha^2 = 1$	OLS, $\sigma_\alpha^2 = 10$	Breitung–Meyer	Harris–Tzavalis
0.900	5	100	0.997 (0.998)	0.883 (0.998)	0.336 (0.337)	0.301 (0.272)
0.900	5	250	1.000 (1.000)	0.997 (1.000)	0.604 (0.615)	0.519 (0.498)
0.900	5	500	1.000 (1.000)	1.000 (1.000)	0.851 (0.863)	0.760 (0.750)
0.900	10	100	1.000 (1.000)	0.998 (1.000)	0.888 (0.912)	0.737 (0.723)
0.900	10	250	1.000 (1.000)	1.000 (1.000)	0.998 (0.999)	0.976 (0.971)
0.900	10	500	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	0.999 (1.000)
0.900	15	100	1.000 (1.000)	1.000 (1.000)	1.000 (0.999)	0.965 (0.962)
0.900	15	250	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)
0.900	15	500	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)
0.950	5	100	0.934 (0.935)	0.759 (0.935)	0.167 (0.151)	0.157 (0.130)
0.950	5	250	1.000 (1.000)	0.981 (1.000)	0.258 (0.249)	0.231 (0.205)
0.950	5	500	1.000 (1.000)	1.000 (1.000)	0.383 (0.391)	0.325 (0.314)
0.950	10	100	1.000 (0.999)	0.979 (0.999)	0.431 (0.442)	0.320 (0.299)
0.950	10	250	1.000 (1.000)	1.000 (1.000)	0.759 (0.766)	0.559 (0.549)
0.950	10	500	1.000 (1.000)	1.000 (1.000)	0.947 (0.956)	0.809 (0.804)
0.950	15	100	1.000 (1.000)	0.999 (1.000)	0.744 (0.770)	0.556 (0.525)
0.950	15	250	1.000 (1.000)	1.000 (1.000)	0.976 (0.983)	0.867 (0.855)
0.950	15	500	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	0.986 (0.985)
0.990	5	100	0.427 (0.409)	0.379 (0.409)	0.078 (0.064)	0.074 (0.062)
0.990	5	250	0.722 (0.723)	0.668 (0.723)	0.080 (0.073)	0.082 (0.069)
0.990	5	500	0.933 (0.935)	0.888 (0.935)	0.094 (0.085)	0.090 (0.079)
0.990	10	100	0.695 (0.683)	0.632 (0.683)	0.101 (0.089)	0.096 (0.078)
0.990	10	250	0.957 (0.956)	0.931 (0.956)	0.130 (0.121)	0.105 (0.098)
0.990	10	500	0.998 (0.999)	0.996 (0.999)	0.162 (0.165)	0.130 (0.126)
0.990	15	100	0.860 (0.842)	0.809 (0.842)	0.133 (0.121)	0.118 (0.096)
0.990	15	250	0.994 (0.994)	0.988 (0.994)	0.193 (0.187)	0.145 (0.135)
0.990	15	500	1.000 (1.000)	1.000 (1.000)	0.269 (0.282)	0.195 (0.189)

the different tests we see the results described in Section 3.4. To summarize, the power of the Breitung–Meyer test is always higher than the power of the Harris–Tzavalis test, and the OLS test has the highest (lowest) power of the three tests when $\sigma_\alpha^2 = 1$ ($\sigma_\alpha^2 = 10$). Finally, we see that the empirical rejection probabilities are quite close to the analytical rejection probabilities. This demonstrates that the local power provides a good approximation to the actual power.

In Table 2, the most striking result is that the OLS test has very high power even for values of ρ very close to unity such as $\rho = 0.99$. According to the analytical results in Section 3.1, this will be the case unless the variability of the variable of interest is dominated by the variability of the individual-specific term. This is also the main conclusion from the simulation studies in the papers by Bond et al. (2002) and Hall and Mairesse (2005) where the time-series processes are covariance stationary in the simulation set-ups. The empirical power of the OLS test is always higher than that of the Breitung–Meyer test and the Harris–Tzavalis test. In addition, the empirical power of the Breitung–Meyer test is always higher that of the Harris–Tzavalis test and compared to Table 1 the empirical power of these tests is lower. These findings are all in accordance with the analytical results in Section 3. Again, we see that the empirical power is quite close to the analytical power except for the OLS test with $\sigma_\alpha^2 = 10$. As explained in Section 3.1, this is to be expected.

5. CONCLUSIONS

In this paper, we have investigated the performance of some of the unit root tests in micropanels which have been suggested in the literature. To do this we have derived the asymptotic power of the tests under local alternatives. One of the main findings is that the initial values are very important for the performance of the tests. This result also holds for unit root tests in single time series and macropanels. The results show that the OLS unit root test is very powerful when the variation of the initial deviation from the mean stationary level is high and in fact the local power is increasing in the parameter describing this feature. However, this test is not invariant with respect to adding individual-specific means to all variables and the results show that its power can be very low when the variation in the individual-specific means is high. The Breitung–Meyer test and the Harris–Tzavalis test are invariant with respect to this type of transformation and another main finding is that the local power of the Breitung–Meyer test is always higher than the local power of the Harris–Tzavalis test. Since the Harris–Tzavalis test relies on rather strong assumptions such as the error terms having homoscedastic variances in order to perform the bias adjustment the results show that the Breitung–Meyer test is to be preferred. This result is confirmed by findings from macropanels; see Moon et al. (2007).

In future research it would be interesting to investigate whether and under which conditions the tests considered in this paper are optimal by deriving the local power of optimal tests. This could be done in a more general framework where the AR parameter can differ across cross-section units under the alternative hypothesis. Results from macropanels suggest that in this case some of the tests considered here might be optimal (the OLS test without incidental intercepts and the Breitung–Meyer test with incidental intercepts); see Moon et al. (2007). These results are also interesting in relation to the type of panel data unit root test suggested by Im et al. (2003). Their test statistic is based on the cross-section average of individual-specific Dickey–Fuller test statistics as opposed to the pooled test statistics considered here in this paper. In macropanels the Im–Pesaran–Shin test appears to have substantially lower power than the optimal tests and that might also be the case in micropanels.

ACKNOWLEDGMENTS

I wish to thank J. Breitung, M. Browning, S. Johansen, H. C. Kongsted, two anonymous referees and co-editor Pierre Perron for useful comments and suggestions. I also wish to thank the Danish National Research Foundation for its support through the Centre for Applied Microeconometrics (CAM).

REFERENCES

Arellano, M. (2003). *Panel Data Econometrics*. Oxford: Oxford University Press.

Baltagi, B. H. (1995). *Econometric Analysis of Panel Data*. New York: John Wiley.

Baltagi, B. H. (2000). *Nonstationary Panels, Panel Cointegration and Dynamic Panels, Advances in Econometrics, Volume 15*. Amsterdam: Elsevier.

Baltagi, B. H. and C. Kao (2000). Nonstationary panels, cointegration in panels and dynamic panels: a survey. In B. H. Baltagi (Ed.), *Nonstationary Panels, Panel Cointegration, and Dynamic Panels, Advances in Econometrics*, *Volume 15*, 7–52. Amsterdam: Elsevier.

Banerjee, A. (1999). Panel data unit roots and cointegration: an overview. *Oxford Bulletin of Economics and Statistics 61*, 607–629.

Bond, S., C. Nauges and F. Windmeijer (2002). Unit roots and identification in autoregressive panel data models: a comparison of alternative tests. Working Paper, University of Bristol.

Breitung, J. (2000). The local power of some unit root tests for panel data. In B. H. Baltagi (Ed.), *Nonstationary Panels, Panel Cointegration, and Dynamic Panels, Advances in Econometrics, Volume 15*, 161–77. Amsterdam: Elsevier.

Breitung, J. and W. Meyer (1994). Testing for unit roots in panel data: are wages on different bargaining levels cointegrated? *Applied Economics 26*, 353–61.

Breitung, J. and M. H. Pesaran (2008). Unit roots and cointegration in panels. In L. Mátayás and P. Sevestre (Eds.), *The Econometrics of Panel Data, Advanced Studies in Theoretical and Applied Econometrics, Volume 46*, 279–322. Berlin: Springer.

Dickey, D. and W. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association 74*, 427–31.

Hall, B. H. and J. Mairesse (2005). Testing for unit roots in panel data: an exploration using real and simulated data. In D. Andrews and J. Stock (Eds.), *Identification and Inference in Econometric Models: Essays in Honor of Thomas J. Rothenberg*. Cambridge: Cambridge University Press.

Harris, D., D. I. Harvey, S. J. Leybourne and N. D. Sakkas (2008). Local asymptotic power of the Im–Pesaran–Shin panel unit root test and the impact of initial observations. Granger Centre Discussion Paper No. 08/02, University of Nottingham.

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*, 201–26.

Hsiao, C. (1986). *Analysis of Panel Data*. Cambridge: Cambridge University Press.

Hwang, J. and P. Schmidt (1996). Alternative methods of detrending and the power of unit root tests. *Journal of Econometrics 71*, 227–48.

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

Kruiniger, H. and E. Tzavalis (2001). Testing for unit roots in short dynamic panels with serially correlated and heteroscedastic disturbance terms. Working Paper, Department of Economics, Queen Mary, University of London.

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

Madsen, E. (2005). Estimating cointegrating relations from a cross section. *Econometrics Journal*, 380–405.

Moon, H. R., B. Perron and P. C. B. Phillips (2007). Incidental trends and the power of panel unit root tests. *Journal of Econometrics 141*, 416–59.

Müller, U. and G. Elliott (2003). Tests for unit roots and the initial condition. *Econometrica 71*, 1269–86.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica 49*, 1417–26.

Schmidt, P. and P. C. B. Phillips (1992). LM tests for a unit root in the presence of deterministic trends. *Oxford Bulletin of Economics and Statistics 54*, 257–87.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica 48*, 817–38.

White, H. (2001). *Asymptotic Theory for Econometricians*. San Diego: Academic Press.

APPENDIX: PROOFS OF RESULTS

This appendix contains the proofs of the propositions in Section 3. The proofs are all based on standard asymptotic theory; see for example White (2001). The notation '$X_N \overset{as}{=} Y_N$' means that $X_N - Y_N \overset{P}{\to} 0$ as $N \to \infty$, i.e. X_N and Y_N, are asymptotically equivalent as $N \to \infty$. We start out with some results that will be used in the following.

A.1. Preliminary lemmas and results

LEMMA A.1. *Under the local-to-unity sequence for ρ given by $\rho_N = 1 - c/N^k$ for $k, c > 0$, the following hold:*

$$\rho_N^t = 1 - t\frac{c}{N^k} + o(N^{-k}), \tag{A.1}$$

$$\frac{1}{1-\rho_N^2} = \frac{N^k}{2c} + o(N^k). \tag{A.2}$$

Proof: The binomial formula yields

$$\rho_N^t = \left(1 - \frac{c}{N^k}\right)^t = 1 - t\frac{c}{N^k} + \frac{t(t-1)}{2!}\frac{c^2}{N^{2k}} - \frac{t(t-1)(t-2)}{3!}\frac{c^3}{N^{3k}} + \cdots + \frac{(-c)^t}{N^{kt}}$$

and the results follow directly. □

For $-1 < \rho \leq 1$, the following expression for y_{it} is obtained by recursive substitution in (2.1):

$$y_{it} = (1-\rho^t)\alpha_i + \rho^t y_{i0} + \rho^{t-1}\varepsilon_{i1} + \cdots + \varepsilon_{it} \quad \text{for } t = 1, \ldots, T.$$

Inserting the expression for the initial value given in Assumption 2.3 yields

$$y_{it} = \alpha_i + \rho^t\sqrt{\tau}\varepsilon_{i0} + \rho^{t-1}\varepsilon_{i1} + \cdots + \varepsilon_{it} \quad \text{for } t = 0, \ldots, T.$$

Using stacked notation, equation (2.1) can be expressed as

$$y_i = \rho y_{i,-1} + v_i.$$

Expressions for the regressor $y_{i,-1}$ and the regression error v_i are given by

$$y_{i,-1} = \alpha_i\iota_T + C_T(\rho)\varepsilon_i + A_T(\rho)\sqrt{\tau}\varepsilon_{i0}, \tag{A.3}$$

$$v_i = (1-\rho)\alpha_i\iota_T + \varepsilon_i, \tag{A.4}$$

where ι_T is a $T \times 1$ vector of ones and $C_T(\rho)$ is the $T \times T$ matrix and $A_T(\rho)$ is the $T \times 1$ vector defined as

$$C_T(\rho) = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & \vdots & \vdots \\ \rho & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ \rho^{T-2} & \cdots & \rho & 1 & 0 \end{bmatrix} \qquad A_T(\rho) = \begin{bmatrix} 1 \\ \rho \\ \rho^2 \\ \vdots \\ \rho^{T-1} \end{bmatrix}.$$

Note that $C_T(\rho_N) = C_T(1) + O(N^{-k})$ and $A_T(\rho_N) = \iota_T + O(N^{-k})$ when $\rho_N = 1 - c/N^k$ according to Lemma A.1.

In the following, we will use notation like $y_{i,-1}(\rho)$ and $u_i(\rho)$ to indicate that these variables depend on the value of the parameter ρ. In Lemma A.2 below we provide results that are used to prove the propositions in Sections 3.1–3.3.

LEMMA A.2. *Consider the sequence $\{x_i(\rho), u_i(\rho)\}_{i=1}^N$ of independent variables where $x_i(\rho)$ and $u_i(\rho)$ are $k \times 1$ variables with mean zero and finite fourth-order moments for all values of ρ. PART A: If the following hold for the sequence ρ_N*

$$\frac{1}{N}\sum_{i=1}^{N} E(x_i(1)'x_i(1)) \to m_{XX} \quad \textit{as } N \to \infty, \tag{A.5}$$

$$E((x_i(\rho_N) - x_i(1))'x_i(\rho_N)) = o(1), \tag{A.6}$$

$$E((x_i(\rho_N) - x_i(1))'x_i(1)) = o(1), \tag{A.7}$$

then

$$\frac{1}{N}\sum_{i=1}^{N} x_i(\rho_N)'x_i(\rho_N) \xrightarrow{P} m_{XX} \quad \textit{as } N \to \infty.$$

PART B: If the following hold for the sequence ρ_N

$$E(x_i(1)'u_i(1)) = 0, \tag{A.8}$$

$$\frac{1}{N}\sum_{i=1}^{N} \text{Var}(x_i(1)'u_i(1)) \to \Sigma \quad \textit{as } N \to \infty, \tag{A.9}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} E(x_i(\rho_N)'(u_i(\rho_N) - u_i(1))) \to \mu_1 \quad \textit{as } N \to \infty, \tag{A.10}$$

$$\text{Var}(x_i(\rho_N)'(u_i(\rho_N) - u_i(1))) = o(1), \tag{A.11}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} E((x_i(\rho_N) - x_i(1))'u_i(1)) \to \mu_2 \quad \textit{as } N \to \infty, \tag{A.12}$$

$$\text{Var}((x_i(\rho_N) - x_i(1))'u_i(1)) = o(1), \tag{A.13}$$

then

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} x_i(\rho_N)'u_i(\rho_N) \xrightarrow{w} N(\mu_1 + \mu_2, \Sigma) \quad \textit{as } N \to \infty.$$

PART C: Let ρ_N be a sequence such that $\sqrt{N}(\rho_N - 1) = O(1)$ and let $\hat{\rho}$ be a sample statistic such that $\sqrt{N}(\hat{\rho} - \rho_N) = O_P(1)$. Then

$$\hat{V}(\rho) \equiv \frac{1}{N}\sum_{i=1}^{N} x_i(\rho)'\hat{u}_i(\rho)\hat{u}_i(\rho)'x_i(\rho) \xrightarrow{P} \Sigma \quad \textit{as } N \to \infty,$$

where

$$\hat{u}_i(\rho) = (\rho_N - \hat{\rho})x_i(\rho)' + u_i(\rho).$$

Proof: PART A: We have that

$$\frac{1}{N}\sum_{i=1}^{N} x_i(\rho)'x_i(\rho) = \frac{1}{N}\sum_{i=1}^{N} x_i(1)'x_i(1) + \frac{1}{N}\sum_{i=1}^{N}(x_i(\rho) - x_i(1))'x_i(1) + \frac{1}{N}\sum_{i=1}^{N}(x_i(\rho) - x_i(1))'x_i(\rho).$$

The Law of Large Numbers together with the condition in (A.5) imply that the first term in the second line of the expression above converges in probability to m_{XX} as $N \to \infty$. Together with the assumption about existence of fourth-order moments the condition in (A.5) is sufficient to give this result. Using the same arguments the last two terms in the second line of the expression above converge in probability to zero as $N \to \infty$ since their means converge to zero according to the conditions in (A.6)–(A.7). Altogether this proves that

$$\frac{1}{N}\sum_{i=1}^{N} x_i(\rho_N)'x_i(\rho_N) \overset{as}{=} \frac{1}{N}\sum_{i=1}^{N} x_i(1)'x_i(1) \overset{P}{\to} m_{XX} \quad \text{as } N \to \infty.$$

PART B: We have that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} x_i(\rho)'u_i(\rho) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} x_i(1)'u_i(1) + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(x_i(\rho) - x_i(1))'u_i(1) + \frac{1}{\sqrt{N}}\sum_{i=1}^{N} x_i(\rho)'(u_i(\rho) - u_i(1)).$$

The Central Limit Theorem and the Law of Large Numbers together with the conditions in (A.8)–(A.13) and the existence of fourth-order moments give

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N} x_i(1)'u_i(1) \overset{w}{\to} N(0, \Sigma) && \text{as } N \to \infty,\\
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(x_i(\rho_N) - x_i(1))'u_i(1) \overset{P}{\to} \mu_1 && \text{as } N \to \infty,\\
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N} x_i(\rho_N)'(u_i(\rho_N) - u_i(1)) \overset{P}{\to} \mu_2 && \text{as } N \to \infty.
\end{aligned}$$

In particular, the conditions in (A.10)–(A.11) together with independency across i imply the following as $N \to \infty$ which give the result in the second equation above:

$$\begin{aligned}
E\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(x_i(\rho_N) - x_i(1))'u_i(1)\right) &\to \mu_1,\\
\mathrm{Var}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(x_i(\rho_N) - x_i(1))'u_i(1)\right) &= \frac{1}{N}\sum_{i=1}^{N}\mathrm{Var}((x_i(\rho_N) - x_i(1))'u_i(1)) \to 0.
\end{aligned}$$

Altogether this proves that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} x_i(\rho_N)'u_i(\rho_N) \overset{w}{\to} N(\mu_1 + \mu_2, \Sigma) \quad \text{as } N \to \infty.$$

PART C: We use the following definitions:

$$\hat{V}(\rho) = \frac{1}{N}\sum_{i=1}^{N} x_i(\rho)'\hat{u}_i(\rho)\hat{u}_i(\rho)'x_i(\rho),$$
$$V(\rho) = \frac{1}{N}\sum_{i=1}^{N} E(x_i(\rho)'u_i(\rho)u_i(\rho)'x_i(\rho)).$$

We have that

$$\hat{V}(\rho_N) - V(\rho_N) = o_P(1).$$

This follows since the terms $x_i(\rho)$ and $u_i(\rho)$ have finite fourth-order moments together with the assumption that $(\hat{\rho} - \rho_N) = o_P(1)$. Because that $V(\rho_N) - V(1) \to 0$ as $N \to \infty$ this proves the result since

$$V(1) = \frac{1}{N}\sum_{i=1}^{N} E(x_i(1)'u_i(1)u_i(1)'x_i(1)) \to \Sigma \quad \text{as } N \to \infty.$$

□

A.2. Proofs of the propositions in Section 3.1: OLS

Using the equation in (3.7) we have that

$$N^k(\hat{\rho}_{\text{OLS}} - \rho) = \left(\frac{1}{N^{2k}}\sum_{i=1}^{N} y_{i,-1}'y_{i,-1}\right)^{-1} \frac{1}{N^k}\sum_{i=1}^{N} y_{i,-1}'v_i \quad \text{for } k > 0.$$

Proposition 3.1 now follows by the results in Lemma A.3 below.

LEMMA A.3. *Under Assumptions 2.1–2.4 and the local-to-unity sequence for ρ given by $\rho_N = 1 - c/\sqrt{N}$ for $c \geq 0$ and when τ is fixed, then the following results hold:*

$$\frac{1}{N}\sum_{i=1}^{N} y_{i,-1}'y_{i,-1} \xrightarrow{P} T\left(\sigma_\alpha^2 + \left(\tau + \frac{T-1}{2}\right)\sigma_{2\varepsilon}\right) \quad \text{as } N \to \infty, \tag{A.14}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i,-1}'v_i \xrightarrow{w} N\left(cT\sigma_\alpha^2, T\left(\sigma_\alpha^2\sigma_{2\varepsilon} + \left(\tau + \frac{T-1}{2}\right)\sigma_{4\varepsilon}\right)\right) \quad \text{as } N \to \infty. \tag{A.15}$$

Under Assumptions 2.1–2.4 and the local-to-unity sequence for ρ_N given by $\rho_N = 1 - c/N$ for $c \geq 0$ and when $\tau_N = bN + o(N)$ for $b > 0$, then the following results hold:

$$\frac{1}{N^2}\sum_{i=1}^{N} y_{i,-1}'y_{i,-1} \xrightarrow{P} bT\sigma_{2\varepsilon} \quad \text{as } N \to \infty, \tag{A.16}$$

$$\frac{1}{N}\sum_{i=1}^{N} y_{i,-1}'v_i \xrightarrow{w} N(0, bT\sigma_{4\varepsilon}) \quad \text{as } N \to \infty. \tag{A.17}$$

Proof: The following results will be used below:

$$\begin{aligned}
\mathrm{tr}\{C_T(\rho)\} &= 0,\\
A_T(1)'A_T(1) &= \iota_T'\iota_T = T,\\
\mathrm{tr}\{C_T(1)'C_T(1)\} &= \mathrm{tr}\left\{C_T(1)'C_T(1)\right\} = \frac{T(T-2)}{2},\\
A_T(\rho_N) - A_T(1) &= O(N^{-k}),\\
C_T(\rho_N) - C_T(1) &= O(N^{-k}).
\end{aligned}$$

The first part of Lemma A.3 follows by using Lemma A.2 with the following definitions of $x_i(\rho)$ and $u_i(\rho)$:

$$\begin{aligned}
x_i(\rho) &= y_{i,-1}(\rho) = \alpha_i\iota_T + A_T(\rho)\sqrt{\tau}\varepsilon_{i0} + C_T(\rho)\varepsilon_i,\\
u_i(\rho) &= v_i(\rho) = (1-\rho)\alpha_i\iota_T + \varepsilon_i,
\end{aligned}$$

where expressions for $y_{i,-1}$ and u_i are given in equations (A.3)–(A.4). This gives that

$$\begin{aligned}
x_i(\rho) - x_i(1) &= (A_T(\rho) - A_T(1))\sqrt{\tau}\varepsilon_{i0} + (C_T(\rho) - C_T(1))\varepsilon_i,\\
u_i(\rho) - u_i(1) &= (1-\rho)\alpha_i\iota_T.
\end{aligned}$$

The sequences $x_i(\rho)$ and $u_i(\rho)$ are both independent across i with finite fourth-order moments.

We prove the result in (A.14) by using Part A in Lemma A.2. We have the following:

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} E(x_i(1)'x_i(1)) &= \frac{1}{N}\sum_{i=1}^{N} E\left(\alpha_i^2\right)\iota_T'\iota_T + \frac{1}{N}\sum_{i=1}^{N}\sigma_{i\varepsilon}^2\tau A_T(1)'A_T(1) + \frac{1}{N}\sum_{i=1}^{N}\sigma_{i\varepsilon}^2\mathrm{tr}\left\{C_T(1)'C_T(1)\right\}\\
&\to \sigma_\alpha^2 T + \sigma_{2\varepsilon}T\left(\tau + \frac{T-1}{2}\right) \quad \text{as } N\to\infty,
\end{aligned}$$

where we have used that α_i, ε_{i0} and ε_i are independent of each other with mean zero. We also have that for all ρ:

$$\begin{aligned}
E(x_i(\rho)'(x_i(\rho_N) - x_i(1))) &= \sigma_{i\varepsilon}^2\tau A_T(\rho)'(A_T(\rho_N) - A_T(1))\\
&\quad + \sigma_{i\varepsilon}^2\,\mathrm{tr}\left\{C_T(\rho)'(C_T(\rho_N) - C_T(\rho_N))\right\} = O(1/\sqrt{N}).
\end{aligned}$$

This means that the conditions in Part A of Lemma A.2 are satisfied such that

$$\frac{1}{N}\sum_{i=1}^{N} x_i(\rho_N)'x_i(\rho_N) \xrightarrow{P} T\left(\sigma_\alpha^2 + \left(\tau + \frac{T-1}{2}\right)\sigma_{2\varepsilon}\right) \quad \text{as } N\to\infty.$$

This proves the result in (A.14).

We prove the result in (A.15) by using Part B in Lemma A.2. The mean and variance of $x_i(1)'u_i(1)$ are given by

$$\begin{aligned}
E(x_i(1)'u_i(1)) &= \sigma_{i\varepsilon}^2\,\mathrm{tr}\{C_T(1)'\} = 0,\\
\mathrm{Var}(x_i(1)'u_i(1)) &= \sigma_{i\varepsilon}^2 E(x_i(1)'x_i(1)) = \sigma_{i\varepsilon}^2\left(T\sigma_\alpha^2 + \left(T\tau + \frac{T(T-1)}{2}\right)\sigma_{i\varepsilon}^2\right),
\end{aligned}$$

such that

$$\frac{1}{N}\sum_{i=1}^{N}\mathrm{Var}(x_i(1)'u_i(1)) \to T\left(\sigma_\alpha^2\sigma_{2\varepsilon} + \left(\tau + \frac{T-1}{2}\right)\sigma_{4\varepsilon}\right) \quad \text{as } N\to\infty.$$

In addition, we have the following results concerning means:

$$E\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}x_i(\rho_N)'(u_i(\rho_N)-u_i(1))\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(1-\rho_N)\iota_T'\iota_T E\left(\alpha_i^2\right) = cT\sigma_\alpha^2,$$
$$E\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(x_i(\rho_N)-x_i(1))'u_i(1)\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma_{i\varepsilon}^2\,\mathrm{tr}\{C_T(\rho)-C_T(1)\} = 0.$$

For the variances we have the following:

$$\mathrm{Var}(x_i(\rho_N)'(u_i(\rho_N)-u_i(1))) \le O(1/N),$$
$$\mathrm{Var}((x_i(\rho_N)-x_i(1))'u_i(1)) \le O(1/N).$$

This holds since

$$\mathrm{Var}(x_i(\rho_N)'(u_i(\rho_N)-u_i(1))) \le E\left((x_i(\rho_N)'x_i(\rho_N))^2\right)^{1/2}E(((u_i(\rho_N)-u_i(1))'\left(u_i(\rho_N)-u_i(1)\right))^2\big)^{1/2} = O(1/N),$$

where the inequality sign in the expression above follows by Schwarz's inequality and the equality sign follows by using that for $du_i(\rho_N) = u_i(\rho_N) - u_i(1)$ we have that $E((du_i(\rho_N)'du_i(\rho_N))^2) = O(1/N^2)$ and $E((x_i(\rho_N)'x_i(\rho_N))^2) = O(1)$. The second result follows by using similar arguments. This means that the conditions in Part B of Lemma A.2 are satisfied such that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}x_i(\rho_N)'u_i(\rho_N) \xrightarrow{w} N\left(cT\sigma_\alpha^2, T\left(\sigma_\alpha^2\sigma_{2\varepsilon} + \left(\tau + \frac{T-1}{2}\right)\sigma_{4\varepsilon}\right)\right) \quad \text{as } N \to \infty.$$

This proves the result in (A.15).

The second part of Lemma A.3 follows by repeating the steps above but with the following definitions of $x_i(\rho_N)$ and $u_i(\rho_N)$:

$$x_i(\rho_N) = A_T(\rho_N)\sqrt{b}\varepsilon_{i0} \overset{as}{=} y_{i,-1}(\rho_N)/\sqrt{N},$$
$$u_i(\rho_N) = (1-\rho_N)\iota_T\alpha_i + \varepsilon_i.$$

We have that

$$E(x_i(1)'x_i(1)) = \sigma_{i\varepsilon}^2 bA_T(1)'A_T(1)' = \sigma_{i\varepsilon}^2 bT,$$
$$E(x_i(\rho)'(x_i(\rho_N)-x_i(1))) = \sigma_{i\varepsilon}^2 bA_T(\rho)'(A_T(\rho_N)-A_T(1)) = O(1/N).$$

Such that

$$\frac{1}{N^2}\sum_{i=1}^{N}y_{i,-1}(\rho_N)'y_{i,-1}(\rho_N) \overset{as}{=} \frac{1}{N}\sum_{i=1}^{N}x_i(\rho_N)'x_i(\rho_N) \xrightarrow{P} \sigma_{2\varepsilon}bT \quad \text{as } N \to \infty.$$

This proves the result in (A.16). Using that $x_i(\rho)$ and $u_i(\rho)$ are independent for all values of ρ we have

$$E(x_i(1)'u_i(1)) = 0,$$
$$\mathrm{Var}(x_i(1)'u_i(1)) = E\left((x_i(1)'u_i(1))^2\right) = \sigma_{i\varepsilon}^4 bT,$$
$$E(x_i(\rho_N)'(u_i(\rho_N)-u_i(1))) = 0,$$
$$E((x_i(\rho_N)-x_i(1))'u_i(1)) = 0.$$

Altogether this implies that

$$\frac{1}{N}\sum_{i=1}^{N}y_{i,-1}(\rho_N)'v_i(\rho_N) \overset{as}{=} \frac{1}{\sqrt{N}}\sum_{i=1}^{N}x_i(\rho_N)'u_i(\rho_N) \xrightarrow{w} N(0, bT\sigma_{4\varepsilon}) \quad \text{as } N \to \infty,$$

which proves the result in (A.17). □

Proof of Proposition 3.2: Part C in Lemma A.2 immediately implies that $\hat{V}_{\mathrm{OLS}}(k)$ is a consistent estimator as $N \to \infty$ of the variance in the limiting distribution of $N^k(\hat{\rho}_{\mathrm{OLS}} - \rho_N)$ when ρ_N is local-to-unity. Then using the expression for t_{OLS} we have that

$$t_{\mathrm{OLS}} = \hat{V}_{\mathrm{OLS}}(k)^{-\frac{1}{2}} N^k(\hat{\rho}_{\mathrm{OLS}} - 1) = \hat{V}_{\mathrm{OLS}}(k)^{-\frac{1}{2}} N^k(\hat{\rho}_{\mathrm{OLS}} - \rho_N) - c\hat{V}_{\mathrm{OLS}}(k)^{-\frac{1}{2}}.$$

Proposition 3.2 now follows by the results already obtained. □

A.3. Proofs of the propositions in Section 3.2: Breitung–Meyer

Using the expressions for $y_{i,-1}$ and v_i given in (A.3)–(A.4) we have that

$$\begin{aligned} \tilde{y}_{i,-1} &= y_{i,-1} - \iota_T y_{i0} = (A_T(\rho) - \iota_T)\sqrt{\tau}\varepsilon_{i0} + C_T(\rho)\varepsilon_i, \\ \tilde{v}_i &= v_i + (\rho - 1)\iota_T y_{i0} = (\rho - 1)\sqrt{\tau}\iota_T\varepsilon_{i0} + \varepsilon_i. \end{aligned}$$

Using the equation in (3.12) we have

$$\sqrt{N}(\hat{\rho}_0 - \rho) = \left(\frac{1}{N}\sum_{i=1}^{N} \tilde{y}_{i,-1}'\tilde{y}_{i,-1}\right)^{-1} \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \tilde{y}_{i,-1}'\tilde{v}_i.$$

Proof of Proposition 3.3: Follows by the results in Lemma A.4 below. □

LEMMA A.4. *Let Assumptions 2.1–2.4 be satisfied. When $\rho_N = 1 - c/\sqrt{N}$ and $\tau_N = b\sqrt{N} + o(\sqrt{N})$ for $b, c \geq 0$, then the following results hold:*

$$\frac{1}{N}\sum_{i=1}^{N} \tilde{y}_{i,-1}'\tilde{y}_{i,-1} \xrightarrow{p} \sigma_{2\varepsilon}\frac{T(T-1)}{2} \quad \text{as } N \to \infty, \tag{A.18}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \tilde{y}_{i,-1}'\tilde{v}_i \xrightarrow{w} N\left(c^2 b\sigma_{2\varepsilon}\frac{T(T-1)}{2}, \sigma_{4\varepsilon}\frac{T(T-1)}{2}\right) \quad \text{as } N \to \infty. \tag{A.19}$$

Proof: In the following we will use that

$$\begin{aligned} A_T(\rho_N) - A_T(1) &= (\rho_N - 1)\tilde{A}_T + o(1/\sqrt{N}), \\ \iota_T'\tilde{A}_T &= \frac{T(T-1)}{2}, \end{aligned}$$

where the $T \times 1$ vector $\tilde{A}_T$ is defined as $\tilde{A}_T = (0, 1, 2, \ldots, T-1)'$.

We use the following specifications:

$$\begin{aligned} x_i(\rho) &= \tilde{y}_{i,-1}(\rho) = (A_T(\rho) - \iota_T)\sqrt{\tau}\varepsilon_{i0} + C_T(\rho)\varepsilon_i, \\ u_i(\rho) &= \tilde{v}_i(\rho) = (\rho - 1)\sqrt{\tau}\iota_T\varepsilon_{i0} + \varepsilon_i, \end{aligned}$$

such that

$$\begin{aligned} x_i(\rho) - x_i(1) &= (A_T(\rho) - \iota_T)\sqrt{\tau}\varepsilon_{i0} + (C_T(\rho) - C_T(1))\varepsilon_i, \\ u_i(\rho) - u_i(1) &= (\rho - 1)\sqrt{\tau}\iota_T\varepsilon_{i0}. \end{aligned}$$

It follows immediately that

$$\frac{1}{N}\sum_{i=1}^{N} E(x_i(1)'x_i(1)) = \frac{1}{N}\sum_{i=1}^{N} \sigma_{i\varepsilon}^2 \operatorname{tr}\{C_T(1)'C_T(1)\} \to \sigma_{2\varepsilon}\frac{T(T-1)}{2} \quad \text{as } N \to \infty.$$

We also have

$$E((x_i(\rho_N) - x_i(1))'x_i(\rho_N)) = \sigma_{i\varepsilon}^2(\tau_N(A_T(\rho_N) - \iota_T)'(A_T(\rho_N) - \iota_T) + \mathrm{tr}\{(C_T(\rho_N) - C_T(1))'C_T(\rho_N)\}) = O(1/\sqrt{N}).$$

This holds since $\tau_N(A_T(\rho_N) - \iota_T)'(A_T(\rho_N) - \iota_T) = c^2b/\sqrt{N}\tilde{A}_T'\tilde{A}_T + o(1/\sqrt{N}) = O(1/\sqrt{N})$. Altogether this gives the result in (A.18) according to Part A of Lemma A.2.

In order to prove the result in (A.19) we will show that the conditions in Part B of Lemma A.2 are satisfied. We have that

$$\begin{aligned}
E(x_i(1)'u_i(1)) &= \sigma_{i\varepsilon}^2 \,\mathrm{tr}\{C_T(1)'\} = 0,\\
\mathrm{Var}(x_i(1)'u_i(1)) &= \sigma_{i\varepsilon}^4 \frac{T(T-1)}{2},\\
E(x_i(\rho_N)'(u_i(\rho_N) - u_i(1))) &= (\rho_N - 1)\tau_N\sigma_{i\varepsilon}^2\iota_T'(A_T(\rho_N) - A_T(1)),\\
E((x_i(\rho_N) - x_i(1))'u_i(1)) &= 0,\\
\mathrm{Var}(x_i(\rho_N)'(u_i(\rho_N) - u_i(1))) &\leq O(1/\sqrt{N}),\\
\mathrm{Var}((x_i(\rho_N) - x_i(1))'u_i(1)) &\leq O(1/\sqrt{N}),
\end{aligned}$$

such that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N E(x_i(\rho_N)'(u_i(\rho_N) - u_i(1))) \to c^2b\frac{T(T-1)}{2}\sigma_{2\varepsilon} \quad \text{as } N \to \infty,$$

where we have used that $\iota_T'(A_T(\rho_N) - A_T(1)) = -c\iota_T'\tilde{A}_T/\sqrt{N} + o(1/\sqrt{N})$ and $(\rho_N - 1)\tau_N = -cb + o(\sqrt{N})$. Altogether by Part B of Lemma A.2 this proves the result in (A.19). □

Proof of Proposition 3.4: Part C in Lemma A.2 immediately implies that $\hat{V}_0$ is a consistent estimator as $N \to \infty$ of the variance in the limiting distribution of $\sqrt{N}(\hat{\rho}_0 - \rho_N)$ when ρ_N is local-to-unity. Then using the expressions for t_0 and $\bar{t}_0$ we have that

$$\begin{aligned}
t_0 &= \hat{V}_0^{-\frac{1}{2}}\sqrt{N}(\hat{\rho}_0 - 1) = \hat{V}_0^{-\frac{1}{2}}\sqrt{N}(\hat{\rho}_0 - \rho_N) - c\hat{V}_0^{-\frac{1}{2}},\\
\bar{t}_0 &= \sqrt{\frac{T(T-1)}{2}}\sqrt{N}(\hat{\rho}_0 - 1) = \sqrt{\frac{T(T-1)}{2}}\sqrt{N}(\hat{\rho}_0 - \rho_N) - c\sqrt{\frac{T(T-1)}{2}}.
\end{aligned}$$

Proposition 3.4 now follows by the results already obtained. □

A.4. Proofs of the propositions in Section 3.3: Harris–Tzavalis

Using the expressions for $y_{i,-1}$ and v_i given in (A.3)–(A.4) and that $Q_T\iota_T = 0$ we have

$$\begin{aligned}
Q_T y_{i,-1} &= Q_T C_T(\rho)\varepsilon_i + Q_T A_T(\rho)\sqrt{\tau}\varepsilon_{i0},\\
Q_T v_i &= Q_T\varepsilon_i,
\end{aligned}$$

where $Q_T = I_T - \frac{1}{T}\iota_T\iota_T'$ is symmetric and idempotent. Using the expression for $\hat{\rho}_{WG}$ in (3.16) we have

$$\sqrt{N}\left(\hat{\rho}_{WG} - \rho + \frac{3}{T+1}\right) = \left(\frac{1}{N}\sum_{i=1}^N y_{i,-1}'Q_T y_{i,-1}\right)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^N\left(y_{i,-1}'Q_T\varepsilon_i + \frac{3}{T+1}y_{i,-1}'Q_T y_{i,-1}\right).$$

Proof of Proposition 3.5: Follows by the results in Lemma A.5 below. □

LEMMA A.5. *Let Assumptions 2.1–2.4 be satisfied. When $\rho_N = 1 - c/\sqrt{N}$ and $\tau_N = b\sqrt{N} + o(\sqrt{N})$ for $b, c \geq 0$, then the following results hold:*

$$\frac{1}{N}\sum_{i=1}^{N} y_{i,-1}' Q_T y_{i,-1} \xrightarrow{P} \sigma_{2\varepsilon}\frac{(T-1)(T+1)}{6} \quad \text{as } N \to \infty, \tag{A.20}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(y_{i,-1}' Q_T \varepsilon_i + \frac{3}{T+1} y_{i,-1}' Q_T y_{i,-1}\right) \xrightarrow{w} N(-c\sigma_{2\varepsilon}(a_1 - cba_2), g_1 m_4 + g_2 \sigma_{4\varepsilon}) \quad \text{as } N \to \infty, \tag{A.21}$$

where

$$a_1 = \frac{(T-1)(T-2)}{12} \qquad a_2 = \frac{T(T-1)}{4},$$

and

$$g_1 = \frac{(T-1)(T-2)(2T-1)}{15T(T+1)} \qquad g_2 = \frac{(T-1)(17T^3 - 44T^2 + 77T - 24)}{60T(T+1)}.$$

Proof: The following results will be used below:

$$\begin{aligned}
Q_T A_T(1) &= Q_T \iota_T = 0, \\
Q_T A_T(\rho) &= Q_T(A_T(\rho) - A_T(1)), \\
\mathrm{tr}\{C_T(1)' Q_T C_T(1)\} &= \frac{(T-1)(T+1)}{6}, \\
\mathrm{tr}\{C_T(1)' \iota_T \iota_T' C_T(1)\} &= \frac{T(T-1)(2T-1)}{6}, \\
\mathrm{tr}\{C_T(1)' Q_T\} &= -\frac{T-1}{2}.
\end{aligned}$$

The $T \times T$ matrix $\tilde{C}_T$ and the $T \times 1$ vector $\tilde{A}_T$ are defined as

$$\tilde{C}_T = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & \vdots & \vdots \\ 1 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ T-2 & \cdots & 1 & 0 & 0 \end{bmatrix} \qquad \tilde{A}_T = \begin{bmatrix} 0 \\ 1 \\ 2 \\ \vdots \\ T-1 \end{bmatrix}.$$

We will also use the following results:

$$\begin{aligned}
(A_T(\rho_N) - A_T(1))' Q_T A_T(\rho_N) &= (1-\rho_N)^2 \tilde{A}_T' Q_T \tilde{A}_T + o(1/N), \\
\mathrm{tr}\{C_T(1)' \tilde{C}_T\} &= \frac{T(T-1)(T-2)}{6}, \\
\iota_T' C_T(1)' \tilde{C}_T \iota_T &= \frac{T(T-1)(T-2)(3T-1)}{24}, \\
\iota_T' \tilde{C}_T{}' \iota_T &= \frac{T(T-1)(T-2)}{6}, \\
\tilde{A}_T' Q_T \tilde{A}_T &= \tilde{A}_T' \tilde{A}_T - \frac{1}{T}(\tilde{A}_T' \iota_T)^2 = \frac{T(T-1)(T+1)}{12}.
\end{aligned}$$

From above we have the following:

$$\sqrt{N}\left(\hat{\rho}_{WG}-\rho+\frac{3}{T+1}\right)=\left(\frac{1}{N}\sum_{i=1}^{N}x_i(\rho)'x_i(\rho)\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}x_i(\rho)'u_i(\rho)\right),$$

with

$$\begin{aligned}
x_i(\rho) &= Q_T y_{i,-1}(\rho) = Q_T A_T(\rho)\sqrt{\tau_N}\varepsilon_{i0} + Q_T C_T(\rho)\varepsilon_i,\\
u_i(\rho) &= Q_T\varepsilon_i + \frac{3}{T+1}Q_T y_{i,-1}(\rho) = Q_T\varepsilon_i + \frac{3}{T+1}\left(Q_T A_T(\rho)\sqrt{\tau}\varepsilon_{i0} + Q_T C_T(\rho)\varepsilon_i\right).
\end{aligned}$$

We have that

$$\begin{aligned}
x_i(1) &= Q_T C_T(1)\varepsilon_i,\\
u_i(1) &= Q_T\varepsilon_i + \frac{3}{T+1}Q_T C_T(1)\varepsilon_i,\\
x_i(\rho)-x_i(1) &= Q_T(A_T(\rho)-\iota_T)\sqrt{\tau}\varepsilon_{i0} + Q_T(C_T(\rho)-C_T(1))\varepsilon_i,\\
u_i(\rho)-u_i(1) &= \frac{3}{T+1}\left(Q_T(A_T(\rho)-\iota_T)\sqrt{\tau}\varepsilon_{i0} + Q_T(C_T(\rho)-C_T(1))\varepsilon_i\right).
\end{aligned}$$

The sequences $x_i(\rho_N)$ and $u_i(\rho_N)$ are both independent across i with finite fourth-order moments.

We have the following:

$$\frac{1}{N}\sum_{i=1}^{N}E(x_i(1)'x_i(1)) = \frac{1}{N}\sum_{i=1}^{N}\sigma_{i\varepsilon}^2\,\mathrm{tr}\{C_T(1)'Q_T C_T(1)\} \to \sigma_{2\varepsilon}\frac{(T-1)(T+1)}{6} \quad \text{as } N\to\infty,$$

and also

$$\begin{aligned}
E((x_i(\rho_N)-x_i(1))'x_i(\rho_N)) &= \sigma_{i\varepsilon}^2(\mathrm{tr}\{(C_T(\rho_N)-C_T(1))'Q_T C_T(\rho_N)\} + \tau_N(A_T(\rho_N)-A_T(1))'Q_T A_T(\rho_N))\\
&= O(1/\sqrt{N}).
\end{aligned}$$

This holds since $\tau_N = b\sqrt{N}+o(\sqrt{N})$ in combination with $(A_T(\rho_N)-A_T(1))'Q_T A_T(\rho_N) = O(1/N)$. Altogether this gives the result in (A.20) according to Part A of Lemma A.2.

We prove (A.21) by showing that the conditions in Part B of Lemma A.2 are satisfied. We have the following:

$$E(x_i(1)'u_i(1)) = \sigma_{i\varepsilon}^2\left(\mathrm{tr}\{C_T(1)'Q_T\} + \frac{3}{T+1}\mathrm{tr}\{C_T(1)'Q_T C_T(1)\}\right) = 0.$$

The following results can be found in Harris and Tzavalis (1999):

$$\begin{aligned}
E\left(\left(\varepsilon_i' C_T(1)'Q_T\varepsilon_i\right)^2\right) &= \frac{(2T-1)(T-1)}{6T}E\left(\varepsilon_{it}^4\right) + \frac{(T-1)(2T^2-4T+3)}{6T}\sigma_{i\varepsilon}^4,\\
E\left(\left(\varepsilon_i' C_T(1)'Q_T C_T(1)\varepsilon_i\right)^2\right) &= \frac{(T^2-1)(T^2+1)}{30T}E\left(\varepsilon_{it}^4\right) + \frac{(T^2-1)(T^2+1)(T-2)}{20T}\sigma_{i\varepsilon}^4,\\
E\left(\varepsilon_i' C_T(1)'Q_T\varepsilon_i\varepsilon_i' C_T(1)'Q_T C_T(1)\varepsilon_i\right) &= -\frac{(T^2-1)}{12}E\left(\varepsilon_{it}^4\right) - \frac{(T^2-1)(T-2)}{12}\sigma_{i\varepsilon}^4.
\end{aligned}$$

This gives that

$$\begin{aligned}\mathrm{Var}(x_i(1)'u_i(1)) &= E\left(\left(\varepsilon_i' C_T(1)' Q_T \varepsilon_i\right)^2\right) + \left(\frac{3}{T-1}\right)^2 E\left(\left(\varepsilon_i' C_T(1)' Q_T C_T(1)\varepsilon_i\right)^2\right)\\ &\quad + \frac{6}{T+1} E\left(\varepsilon_i' C_T(1)' Q_T \varepsilon_i \varepsilon_i' C_T(1)' Q_T C_T(1)\varepsilon_i\right)\\ &= g_1 E\left(\varepsilon_{it}^4\right) + g_2 \sigma_{i\varepsilon}^4,\end{aligned}$$

where g_1 and g_2 are defined in Lemma A.5. This implies that

$$\frac{1}{N}\sum_{i=1}^{N} \mathrm{Var}(x_i(1)'u_i(1)) \to g_1 m_4 + g_2 \sigma_{4\varepsilon} \quad \text{as } N \to \infty.$$

For the first mean term we have that

$$\begin{aligned}E\left((x_i(\rho_N) - x_i(1))'u_i(1)\right) &= \sigma_{i\varepsilon}^2 \,\mathrm{tr}\left\{\left(I_T + \frac{3}{T+1} C_T(1)\right)' Q_T \left(C_T(\rho_N) - C_T(1)\right)\right\}\\ &= \sigma_{i\varepsilon}^2 \,\mathrm{tr}\left\{\left(I_T + \frac{3}{T+1} C_T(1)\right)' Q_T \tilde{C}_T\right\}(\rho_N - 1) + o(1/\sqrt{N}),\end{aligned}$$

such that as $N \to \infty$:

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} E\left((x_i(\rho_N) - x_i(1))'u_i(1)\right) = -c\sigma_{2\varepsilon}\left(\mathrm{tr}\{Q_T \tilde{C}_T\} + \frac{3}{T+1}\mathrm{tr}\{C_T(1)' Q_T \tilde{C}_T\}\right).$$

For the second mean term we have that

$$\begin{aligned}&E(x_i(\rho_N)'(u_i(\rho_N) - u_i(1)))\\ &= \sigma_{i\varepsilon}^2 \frac{3}{T+1}\left(\mathrm{tr}\{C_T(\rho_N)' Q_T (C_T(\rho_N) - C_T(1))\} + \tau_N A_T(\rho_N)' Q_T (A_T(\rho_N) - A_T(1))\right)\\ &= \sigma_{i\varepsilon}^2 \frac{3}{T+1}\left(\mathrm{tr}\{C_T(1)' Q_T \tilde{C}_T\}(\rho_N - 1) + \tau_N(\rho_N - 1)^2 \tilde{A}_T' Q_T \tilde{A}_T\right) + o(1/\sqrt{N}),\end{aligned}$$

such that as $N \to \infty$:

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} E(x_i(\rho_N)'(u_i(\rho_N) - u_i(1))) \to -c\sigma_{2\varepsilon}\frac{3}{T+1}\left(\mathrm{tr}\{C_T(1)' Q_T \tilde{C}_T\} - cb\tilde{A}_T' Q_T \tilde{A}_T\right).$$

For the variance terms we have that

$$\begin{aligned}\mathrm{Var}((x_i(\rho_N) - x_i(1))'u_i(1)) &\le O(1/\sqrt{N}),\\ \mathrm{Var}(x_i(\rho_N)'(u_i(\rho_N) - u_i(1))) &\le O(1/\sqrt{N}).\end{aligned}$$

Altogether this gives the result in (A.21) according to Part B of Lemma A.2 since

$$\begin{aligned}&-c\sigma_{2\varepsilon}\left(\mathrm{tr}\{Q_T \tilde{C}_T\} + \frac{6}{T+1}\mathrm{tr}\left\{C_T(1)' Q_T \tilde{C}_T\right\} - cb\frac{3}{T+1}\tilde{A}_T' Q_T \tilde{A}_T\right)\\ &= -c\sigma_{2\varepsilon}\left(-\frac{(T-1)(T-2)}{6} + \frac{6(T-1)(T-2)}{24} - cb\frac{T(T-1)}{4}\right)\\ &= -c\sigma_{2\varepsilon}\left(\frac{(T-1)(T-2)}{12} - cb\frac{T(T-1)}{4}\right).\end{aligned}$$

□

Proof of Proposition 3.6: Part C in Lemma A.2 immediately implies that $\hat{V}_{WG}$ is a consistent estimator as $N \to \infty$ of the variance in the limiting distribution of $\sqrt{N}(\hat{\rho}_{WG} - \rho_N + 3/(T+1))$ when ρ_N is local-to-unity. Then using the expressions for t_{WG} and $\bar{t}_{WG}$ we have that

$$t_{WG} = \hat{V}_{WG}^{-\frac{1}{2}}\sqrt{N}\left(\hat{\rho}_{WG} - 1 + \frac{3}{T+1}\right) = \hat{V}_{WG}^{-\frac{1}{2}}\sqrt{N}\left(\hat{\rho}_{WG} - \rho_N + \frac{3}{T+1}\right) - c\hat{V}_{WG}^{-\frac{1}{2}},$$
$$\bar{t}_{WG} = \tilde{V}_{WG}^{-\frac{1}{2}}\sqrt{N}\left(\hat{\rho}_{WG} - 1 + \frac{3}{T+1}\right) = \tilde{V}_{WG}^{-\frac{1}{2}}\sqrt{N}\left(\hat{\rho}_{WG} - \rho_N + \frac{3}{T+1}\right) - c\tilde{V}_{WG}^{-\frac{1}{2}}.$$

Proposition 3.6 now follows by the results already obtained. □