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

Edith Madsen

Centre for Applied Microeconometrics (CAM)
Department of Economics, University of Copenhagen, Denmark

This version: April 2009

Abstract

The objective of the paper is to investigate and compare the performance of some of the unit root tests in micro-panels 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; Unit roots; Local alternatives; Initial values

JEL classification: C12; C23
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 micro-panels. At present there is a large econometric literature dealing with unit root testing in panel data models which has developed during the last ten years. Contrary to the previous literature on dynamic panel data models, a large part of this new literature considers macro-panels where the cross-section and time-series dimensions are similar in magnitude. Banerjee (1999), Baltagi & Kao (2000) and Breitung & Pesaran (2008) review many of the contributions to the literature on unit root testing in panel data models. Reviews of the literature on dynamic micro-panels 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 micro-panels have been investigated and compared in simulation studies, see for example Bond, Nauges & Windmeijer (2002) and Hall & Mairesse (2005). However, the outcome of these might depend on the particular choice of nuisance parameters in the simulation setup 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 macro-panels.

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 & 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 & 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 & Phillips
(1992) and Dickey & 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 micro-panels is a result which is also found for single time series and macro-panels,
see Müller & Elliott (2003) and Harris, Harvey, Leybourne & Sakkas (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, Perron & Phillips (2007) where they derive the power envelope of unit tests in macro-
panels. 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 macro-panel version of the Breitung-Meyer
test has asymptotic power equal to the power envelope whereas the macro-panel version of the Harris-
Tzavalis test suggested by Levin, Lin & Chu (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 & Schmidt (1996). An important difference is that in macro-panels 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.

2 The model and assumptions

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

\[ y_{it} = \rho y_{i(t-1)} + (1 - \rho) \alpha_i + \varepsilon_{it} \quad \text{for } i = 1, ..., N \text{ and } t = 1, ..., T \]  

(1)

where \(-1 < \rho \leq 1\) and for every \(i = 1, ..., N\) 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 1** \(\varepsilon_{it}\) is independent across \(i, t\) with \(E(\varepsilon_{it}) = 0\), \(E(\varepsilon_{it}^2) = \sigma_{\varepsilon}^2\) and \(E(\varepsilon_{it}^4) = E(\varepsilon_{is}^4)\) for all \(t, s = 1, ..., T\). In addition \(\varepsilon_{it}\) is independent of \(\alpha_i\) and \(y_{i0}\).

**Assumption 2** \(\alpha_i\) is iid across \(i\) with \(E(\alpha_i) = 0\), \(E(\alpha_i^2) = \sigma_{\alpha}^2\) and \(E(\alpha_i^4) < \infty\).

**Assumption 3** For \(-1 < \rho \leq 1\) the initial values satisfy \(y_{i0} = \alpha_i + \sqrt{T}\varepsilon_{i0}\) where \(\varepsilon_{i0}\) is independent of \(\alpha_i\) and independent across \(i\) with \(E(\varepsilon_{i0}) = 0\) and \(E(\varepsilon_{i0}^2) = \sigma_{\varepsilon}^2\).

**Assumption 4** The following hold:

(i) \(E|\varepsilon_{it}|^{4+\delta} < K < \infty\) for some \(\delta > 0\) and all \(i = 1, ..., N, t = 0, 1, ..., T\)

(ii) \(\frac{1}{N} \sum_{i=1}^{N} \sigma_{\varepsilon}^2 \rightarrow \sigma_{2\varepsilon} > 0\) as \(N \rightarrow \infty\)

(iii) \(\frac{1}{N} \sum_{i=1}^{N} \sigma_{\varepsilon}^4 \rightarrow \sigma_{4\varepsilon} > 0\) as \(N \rightarrow \infty\)

(iv) \(\frac{1}{N} \sum_{i=1}^{N} E(\varepsilon_{it}^4) \rightarrow m_4\) as \(N \rightarrow \infty\)
Assumption 1 states that the errors $\varepsilon_{it}$ are independent over cross-section units and time and allowed to be heteroskedastic over cross-section units but not over time. Further, they are independent of the individual-specific term $\alpha_i$ and the initial value $y_{i0}$. The assumption about independency over time is stronger than the usual assumption about $\varepsilon_{it}$ 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 states that the $\alpha_i$‘s are iid 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 (1) as describing the behavior 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. Assumption 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 $\tau$ 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 $\rho$ approaches unity then the parameter $\tau$ 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 $\varepsilon_{it}$ is independent of $\varepsilon_{i0}$ by Assumption 1. Finally, Assumption 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 $\varepsilon_{it}$ 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 $\varepsilon_{it}$ are homoskedastic across units then $\sigma_{4\varepsilon} = \sigma_{2\varepsilon}^2$. Assumption 4 (iv) is only required in relation to the test statistic suggested by Harris & Tzavalis (1999), as this is the only statistic of the ones considered in this paper which depends on fourth order moments. Also note that $\sigma_{4\varepsilon} - \sigma_{2\varepsilon}^2 = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} (\sigma_{4\varepsilon}^2 - \sigma_{2\varepsilon})^2 \geq 0$ and $m_4 - \sigma_{4\varepsilon} = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} E \left( (\varepsilon_{it}^2 - \sigma_{2\varepsilon}^2)^2 \right) \geq 0$ 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

\[ H_0 : \rho = 1 \quad H_A : |\rho| < 1 \]  

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

\[ \rho_N = 1 - \frac{c}{Nk} \quad \text{for } k, c > 0 \]
This means that as the sample size $N$ increases, the value of the parameter $\rho$ is in a $N^{-k}$ neighborhood of unity. So instead of deriving asymptotic representations based on $\rho$ 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 $\rho$ considered in this paper converge weakly to normal distributions at the rate $\sqrt{N}$ and therefore we consider local-to-unity sequences for $\rho$ 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 $\rho$ 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 4 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). We consider the following two situations

(i) : $\tau$ is fixed
(ii) : $\tau = \kappa \frac{1}{1 - \rho^2}$ for $\kappa > 0$

(i) means that the initial deviation from the stationary level is described by a parameter $\tau$ that remains constant as $\rho$ 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) $\tau$ depends on $\rho$ and goes to infinity as $\rho$ approaches unity and it is not defined for $\rho$ equal to unity. This means that the two formulations are fundamentally different.

Under the local-to-unity sequence for $\rho$ given by $\rho = 1 - c/N^k$ the formulations in (4)-(5) correspond to

(i) : $\tau_N = \tau$ for $\tau \geq 0$
(ii) : $\tau_N = b N^k + o(N^k)$ for $k, b > 0$

Note that (7) is more general than (5) since $b$ and $c$ might take values independently of each other. (5) corresponds to (7) with $b = \kappa/(2c)$, see Lemma 1 in Appendix A.1, implying that the parameters $b$
and $c$ are not independent of each other but are on a specific path in the parameter space. In (6) $\tau$ is a fixed parameter such that term $\sqrt{T}\varepsilon_{i0}$ is of the same order of magnitude as the remaining terms in the expression for the variables $y_{it}$. On the other hand, in (7) this term dominates the behavior of the variables $y_{it}$ asymptotically as $N$ tends to infinity since then we have $\sqrt{T}\varepsilon_{i0} = O_p\left(N^{k/2}\right)$. The interpretation is that the behavior of the observed variables $y_{i0}, \ldots, y_{iT}$ is dominated by the initial deviation from the mean stationary level $(y_{i0} - \alpha_i) = \sqrt{T}\varepsilon_{i0}$. 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 micro-panel.

In this paper we will focus on the cases where $\tau$ 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 (1) can be rewritten as the following regression model

$$
y_{it} = \rho y_{i,t-1} + v_{it}
$$

$$
v_{it} = (1 - \rho) \alpha_i + \varepsilon_{it}
$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$  \hspace{1cm} (8)

The OLS estimator of the autoregressive parameter $\rho$ is defined by

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

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 $\alpha_i$ which appears in both the regressor $y_{i,t-1}$ and the regression error $v_{it}$. As $\alpha_i$ 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 $\rho$ approaches unity. Now the regressor $y_{i,t-1}$ can be expressed as the sum of the two independent terms $\alpha_i$ and $(y_{i,t-1} - \alpha_i)$ 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 $\rho$ approaches unity, the asymptotic bias of $\hat{\rho}_{OLS}$ is positive and decreases towards zero as $\rho$ approaches unity. This describes the situation the variance of the initial deviation from the mean stationary level is fixed. On the other hand, if the behavior of $y_{i,t-1}$ is dominated by the term $(y_{i,t-1} - \alpha_i)$ as $\rho$ approaches unity, the asymptotic bias of $\hat{\rho}_{OLS}$ will be zero when $\rho$ 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 1 below. The proposition provides the limiting distribution of the OLS estimator \( \hat{\rho}_{\text{OLS}} \) under both the null hypothesis when \( \rho \) is unity and local alternatives when \( \rho \) is local-to-unity. We consider different local alternatives depending on the assumption being made about the initial values as given by equations (6)-(7).

**Proposition 1** Under Assumption 1-4 and the local-to-unity sequence for \( \rho \) given by \( \rho_N = 1 - c/\sqrt{N} \) for \( c \geq 0 \) and when \( \tau \) is fixed, the limiting distribution of the OLS estimator \( \hat{\rho}_{\text{OLS}} \) is given by

\[
\sqrt{N} (\hat{\rho}_{\text{OLS}} - \rho_N) \xrightarrow{w} N \left( c \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + (\tau + \frac{1}{2}) \sigma_2^2} \frac{1}{T} \frac{\sigma_\alpha^2 \sigma_2 \tau + (\tau + \frac{1}{2}) \sigma_4}{\sigma_\alpha^2 + (\tau + \frac{1}{2}) \sigma_2^2} \right) \quad \text{as } N \to \infty
\]

Under Assumption 1-4 and the local-to-unity sequence for \( \rho \) 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

\[
N (\hat{\rho}_{\text{OLS}} - \rho_N) \xrightarrow{w} N \left( 0, \frac{\sigma_4^2}{b \sigma_2^4} \right) \quad \text{as } N \to \infty
\]

The proof of Proposition 1 is given in Appendix A.2. The proposition shows that in the unit root case when \( c = 0 \) and \( \tau \) is fixed, the estimator \( \hat{\rho}_{\text{OLS}} \) is \( \sqrt{N} \)-consistent and its limiting variance is decreasing in \( \tau \), \( T \) and \( \sigma_2^2/\sigma_4 \). 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 \( \sigma_\alpha^2/\sigma_2^2 \) and decreasing in \( \tau \) and \( T \). The limiting variance of \( \hat{\rho}_{\text{OLS}} \) is decreasing in \( \tau \), \( T \) and \( \sigma_2^2/\sigma_4 \) and increasing in \( \sigma_\alpha^2/\sigma_2^2 \) 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 \( \rho \) 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 \( \sigma_2^2/\sigma_4 \). 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 behavior 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 \( \rho \) given by \( 1 - c/N \) for \( c > 0 \), see the result in (7), whereas the variation of the remaining terms in \( y_{it} \) is bounded as \( N \) tends to infinity. This implies that the numerator in (9) must be normalized by \( N \) in order to converge in distribution and the denominator in (9) 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 behavior of the regressor, being independent of the term \( \alpha_i \). This indicates that the asymptotic representation in (11) is only appropriate when the variances of \( \alpha_i \) and \( \varepsilon_{it} \) are much smaller than the variance of \( (y_{i0} - \alpha_i) \). Once the variances are of similar magnitude, the asymptotic representation in (10) is expected to provide a better approximation to the actual distribution of \( \hat{\rho}_{\text{OLS}} \).

The unit root test based on the usual \( t \)-statistic is obtained by normalizing \( (\hat{\rho}_{\text{OLS}} - 1) \) appropriately. For this purpose we need a consistent estimator of the limiting variance of \( \hat{\rho}_{\text{OLS}} \) and we use White’s heteroskedastic 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 $\hat{\rho}_{OLS}$ converges in distribution at the rate $\sqrt{N}$ and $N$ respectively, White’s heteroskedastic consistent estimator of the limiting variance of $\hat{\rho}_{OLS}$ is given by the following expression

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

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

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

Note that $\hat{V}_{OLS}(k)^{-\frac{1}{2}} N^k$ 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 2** Under Assumption 1-4 and the local-to-unity sequence for $\rho$ given by $\rho_N = 1 - c/\sqrt{N}$ for $c \geq 0$ and when $\tau$ is fixed, the limiting distribution of the OLS t-statistic $t_{OLS}$ is given by

$$
t_{OLS} \overset{w}{\rightarrow} N \left( -c \left( \tau + \frac{T-1}{2} \right) \sqrt{\left( \frac{\sigma_o^2}{\sigma_2^2} + \left( \tau + \frac{T-1}{2} \right) \frac{\sigma_4^2}{\sigma_2^2} \right)^{-1}}, 1 \right) \quad \text{as } N \rightarrow \infty
$$

Under Assumption 1-4 and the local-to-unity sequence for $\rho$ 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

$$
t_{OLS} \overset{w}{\rightarrow} N \left( -c \sqrt{\frac{\sigma_o^2}{\sigma_2^2}}, 1 \right) \quad \text{as } N \rightarrow \infty
$$

The proof of Proposition 2 is given in Appendix A.2. 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 $\tau$ is fixed, the local power is increasing in $c$, $\tau$, $T$ and $\sigma_2^2/\sigma_4^2$ (the location parameter is shifted to the left when these parameters increase) and decreasing in $\sigma_o^2/\sigma_2^2$ (the location parameter is shifted to the right when $\sigma_o^2/\sigma_2^2$ increases). Under the covariance stationary alternative when $b = 1/(2c)$, the local power only depends on $c$, $T$ and $\sigma_2^2/\sigma_4^2$ 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 value are high, see Müller & 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 (15) will only provide a good approximation to the actual distribution of the t-statistic when the behavior 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: \( \sigma^2_{2r}/\sigma_{4e} = 1 \) and \( \sigma^2_{dr}/\sigma_{2e} = 1 \). 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 \( \sigma^2_{dr}/\sigma_{2e} > 1 \). 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 \( \rho \) very close to unity. However, if this is not the case the power of the test for values of \( \rho \) close to unity is expected to be low when \( \sigma^2_{dr}/\sigma_{2e} \) 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 (1) yields the following regression equation:

\[
y_{it} - y_{i0} = \rho (y_{i(t-1)} - y_{i0}) + \tilde{v}_{it}
\]

\[
\tilde{v}_{it} = (\rho - 1) (y_{i0} - \alpha_i) + \tilde{\epsilon}_{it}
\]

for \( i = 1, \ldots, N \) and \( t = 1, \ldots, T \) \hspace{1cm} (16)

The LS estimator of \( \rho \) obtained from this regression equation is defined by

\[
\hat{\rho}_0 = \left( \sum_{i=1}^{N} \tilde{y}_{i,-1} \tilde{y}_{i,-1} \right)^{-1} \left( \sum_{i=1}^{N} \tilde{y}_{i,-1} \tilde{y}_{i} \right)
\]

(17)

where \( \tilde{y}_{i} = y_i - y_{i0}T \), \( \tilde{y}_{i,-1} = y_{i,-1} - y_{i0}T \) and \( T \) 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 & Meyer (1994). As an example, this means that the asymptotic bias equals 0.050, 0.025 and 0.005 when \( \rho \) 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 \((y_{i(t-1)} - y_{i0})\) and the regression error \(v_{it}\). The covariance between the regressor and the regression error decreases towards zero as \( \rho \) 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 \( \rho \) approaches unity. This is exactly what happens when the initial values are such that the time-series processes become covariance stationary.

Proposition 3 below provides the limiting distribution of the Breitung-Meyer estimator \( \hat{\rho}_0 \) under both the null hypothesis when \( \rho \) is unity and the mean stationary local alternative when \( \rho \) 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 Under Assumption 1-4 and the local-to-unity sequence for \( \rho \) given by \( \rho_N = 1 - c/\sqrt{N} \) for \( c \geq 0 \) and when \( \tau_N = b\sqrt{N} + o\left(\sqrt{N}\right) \) for \( b \geq 0 \), the limiting distribution of the Breitung-Meyer estimator \( \hat{\rho}_0 \) is given by

\[
\sqrt{N} (\hat{\rho}_0 - \rho_N) \xrightarrow{w} N \left( c^2 b, \frac{\sigma_{4e}}{\sigma_{2e}^2} \frac{2}{T(T-1)} \right) \quad \text{as } N \to \infty \tag{18}
\]

The proof of Proposition 3 is given in Appendix A.3. The proposition shows that in the unit root case and under the mean stationary local alternative when \( \tau \) 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 \( 1/N \). 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 \( \sigma_{2e}/\sigma_{4e} \) and decreasing in both. As indicated above, the results follow by using that when the variance of \( (y_{0i} - \alpha_i) \) is of order less than \( \sqrt{N} \), the asymptotic bias disappears under the local alternative. This is the case when \( \tau \) is fixed. On the contrary, under covariance stationarity this is not the case, as the variance of \( (y_{0i} - \alpha_i) \) in this case is of order \( pN \), see the result in (7).

As before White’s heteroskedastic consistent estimator of the limiting variance of \( \hat{\rho}_0 \) is given by the following expression

\[
\hat{V}_0 = \left( \frac{1}{N} \sum_{i=1}^N \tilde{y}_{i-1} \tilde{y}_{i-1} \right)^{-1} \frac{1}{N} \sum_{i=1}^N \tilde{v}_i \tilde{v}_i \left( \frac{1}{N} \sum_{i=1}^N \tilde{y}_{i-1} \tilde{y}_{i-1} \right)^{-1} \tag{19}
\]

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

\[
t_0 = \hat{V}_0^{-1/2} \sqrt{N} (\hat{\rho}_0 - 1) \tag{20}
\]

When the errors \( \varepsilon_{it} \) are homoskedastic across units such that \( \sigma_{4e} = \sigma_{2e}^2 \), 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

\[
\tilde{t}_0 = \sqrt{\frac{T(T-1)}{2}} \frac{1}{\sqrt{N}} \sqrt{N} (\hat{\rho}_0 - 1) \tag{21}
\]

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

Proposition 4 Under Assumption 1-4 and the local-to-unity sequence for \( \rho \) given by \( \rho_N = 1 - c/\sqrt{N} \) for \( c \geq 0 \) and when \( \tau_N = b\sqrt{N} + o\left(\sqrt{N}\right) \) for \( b \geq 0 \), the limiting distribution of the Breitung-Meyer t-statistic \( t_0 \) is given by

\[
t_0 \xrightarrow{w} N \left( -c (1 - cb), \frac{\sigma_{2e}^2 T(T-1)}{2\sigma_{4e}} \right) \quad \text{as } N \to \infty \tag{22}
\]

The limiting distribution of the normalized coefficient statistic \( \tilde{t}_0 \) is given by

\[
\tilde{t}_0 \xrightarrow{w} N \left( -c (1 - cb), \frac{\sigma_{2e}^2 T(T-1)}{2\sigma_{4e}} \right) \quad \text{as } N \to \infty \tag{23}
\]
The proof of Proposition 4 is given in Appendix A.3. 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 \( \sigma_{2e}^2/\sigma_{4e} \) when \( b < 1/c \). In particular, this means that the local power is monotonically increasing in \( c \) when \( b = 0 \) (\( \tau \) is fixed) and when \( b = 1/(2c) \) (covariance stationarity). When \( \sigma_{2e}^2/\sigma_{4e} = 1 \) the local power of a one-sided test at the nominal 5\% level against these two alternatives is given by \( \Phi \left( -1.645 + c\sqrt{T(T-1)/2} \right) \) and \( \Phi \left( -1.645 + c/2\sqrt{T(T-1)/2} \right) \), respectively, where \( \Phi (\cdot) \) denotes the cdf of the standard normal distribution. This means that for a specific value of \( \rho \) we need four times as many cross-section observations in the covariance stationary alternative as in the fixed \( \tau \) 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 \( \sigma_{2e}^2 / (1 - \rho^2) \).

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 & Elliott (2003) and Harris, Harvey, Leybourne & Sakkas (2008) for unit root tests in single time series and macro-panels, 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 \( b \) 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 \( y_{it}/\sigma_{4e} \) would have higher power.

Figure 1: The local power of the Breitung-Meyer unit root test as a function of \( c \) for different values of \( b \) when \( T + 1 = 5 \) and \( \sigma_{2e}^2/\sigma_{4e} = 1 \)

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 \( \sigma_{4e} = \sigma_{2e}^2 \). 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 \( \sigma_{2e}^2 < \sigma_{4e} \). So unless there is any prior knowledge about the error terms being homoskedastic over cross-section units the unit root test should be based on the \( t \)-statistic. Note, that if \( \sigma_{2e}^2 < \sigma_{4e} \) 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 \( \alpha_i \). 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.

### 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 (1). This yields the following regression model

\[
y_{it} - \frac{1}{T} \sum_{t=1}^{T} y_{it} = \rho \left( y_{i(t-1)} - \frac{1}{T} \sum_{t=1}^{T} y_{i(t-1)} \right) + w_{it} \\
w_{it} = \varepsilon_{it} - \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it}
\]

for \( i = 1, ..., N \) and \( t = 1, ..., T \) (24)

The within-group estimator of \( \rho \) is then defined by

\[
\hat{\rho}_{WG} = \left( \sum_{i=1}^{N} y_{i,-1}Q_{Tr}y_{i,-1} \right)^{-1} \left( \sum_{i=1}^{N} y_{i,-1}Q_{Tr}y_{i} \right)
\]

where \( Q_T \) is a \( T \times T \) symmetric and idempotent matrix defined as \( Q_T = I_T - \frac{1}{T} T_{t}T' \) where \( I_T \) is the \( T \times T \) identity matrix and \( T_{t}T' \) 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 \( \rho \) and \( T \) which is always negative when \( 0 < \rho < 1 \) and decreases numerically as \( T \) increases. Harris & 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 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 \( \rho \) is local-to-unity.

**Proposition 5** Under Assumption 1-4 and the local-to-unity sequence for \( \rho \) 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

\[
\sqrt{N} \left( \hat{\rho}_{WG} - \rho_N + \frac{3}{T+1} \right) \xrightarrow{w} N \left( -c, \frac{T-2}{2(T+1)} + c^2 b^2 \frac{3T}{2(T+1)} \frac{k_1m_1 + k_2 \sigma_{4e}}{\sigma_{2e}^2} \right) \quad \text{as } N \to \infty
\]
where
\[
k_1 = \frac{12(T - 2)(2T - 1)}{5T(T - 1)(T + 1)^3} \quad k_2 = \frac{3(17T^3 - 447T^2 + 77T - 24)}{5T(T - 1)(T + 1)^3}
\] (27)

The proof of Proposition 5 is given in Appendix A.4. 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 \(\tau\) 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 \(\varepsilon_{it}\) through the term \(m_4\). As \(k_1 < k_2\) the fourth order moments receive less weight than the squared second order moments.

Harris & Tzavalis (1999) assume that the errors \(\varepsilon_{it}\) are iid normally distributed across \(i\) such that \(\sigma_{4T} = \sigma_{2T}^2\) and \(m_4 = 3\sigma_{2T}^2\). In this case, the limiting variance of \(\hat{\rho}_{WG}\) only depends on \(T\) and is given by the following expression
\[
\hat{V}_{WG} = 3k_1 + k_2 = \frac{3(17T^2 - 20T + 17)}{5(T - 1)(T + 1)^3}
\] (28)

Therefore, Harris & Tzavalis (1999) suggest using the normalized coefficient statistic as a unit root test statistic. It is defined as follows
\[
\hat{t}_{WG} = \hat{V}_{WG}^{-\frac{1}{2}} \sqrt{N} \left( \hat{\rho}_{WG} - 1 + \frac{3}{T + 1} \right)
\] (29)

However, as before it is also possible to use the usual \(t\)-statistic as a test statistic. White’s heteroskedasticity consistent estimator of the limiting variance of the bias adjusted within-group estimator is given by the following expression
\[
\hat{V}_{WG} = \left( \frac{1}{N} \sum_{i=1}^{N} y_{i, -1}Q_T y_{i, -1} \right)^{-1} \left( \frac{1}{N} \sum_{i=1}^{N} y_{i, -1}Q_T \hat{w}_i \hat{w}_i' Q_T y_{i, -1} \right)^{-1}
\] (30)

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
\[
t_{WG} = V_{WG}^{-\frac{1}{2}} \sqrt{N} \left( \hat{\rho}_{WG} - 1 + \frac{3}{T + 1} \right)
\] (31)

The limiting distributions of these test statistics are given in Proposition 6 below.

**Proposition 6** Under Assumption 1-4 and the local-to-unity sequence for \(\rho\) given by \(\rho_N = 1 - c/\sqrt{N}\) for \(c \geq 0\) and when \(\tau_N = b\sqrt{N} + a\left(\sqrt{N}\right)\) for \(b \geq 0\), the limiting distribution of the adjusted within-group \(t\)-statistic \(t_{WG}\) is given by
\[
t_{WG} \overset{w}{\rightarrow} N \left( -c(1 - cb) \frac{3T}{2(T + 1)} \sigma_{2T} \sqrt{\frac{k_1m_4 + k_2\sigma_{4T}}{\sigma_{2T}^2}}, 1 \right) \text{ as } N \rightarrow \infty
\] (32)

The limiting distribution of the Harris-Tzavalis normalized coefficient statistic \(\hat{t}_{WG}\) is given by
\[
\hat{t}_{WG} \overset{w}{\rightarrow} N \left( -c(1 - cb) \frac{3T}{2(T + 1)} \sqrt{\frac{5(T - 1)(T + 1)^3}{3(17T^2 - 20T + 17)}}, \frac{\hat{k}_1m_4 + \hat{k}_2\sigma_{4T}}{\sigma_{2T}^2} \right) \text{ as } N \rightarrow \infty
\] (33)
where

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

(34)

The proof of Proposition 6 is given in Appendix A.4. 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 \( c, T, \sigma_{2x}/\sigma_{4x} \) and \( \sigma_{2x}^2/m_4 \) when \( b < 1/c \). This means that the local power is monotonically increasing in \( c \) when \( b = 0 \) (\( \tau \) is fixed) and when \( b = 1/(2c) \) (covariance stationarity). 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 \( \rho \). The unit root test based on the Harris-Tzavalis normalized coefficient statistic \( \tilde{i}_{WG} \) is asymptotically equivalent to test based on the \( t \)-statistic \( t_{WG} \) when the errors \( \varepsilon_{it} \) are normally distributed and homoskedastic across units. If at least one of these assumptions are 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 \( \sigma_{2x}^2 < \sigma_{4x} \) and when the excess kurtosis of \( \varepsilon_{it} \) is positive, i.e. \( m_4 > 3\sigma_{2x}^2 \). 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 \( \varepsilon_{it} \) being homoskedastic over time. If this assumption is violated the Harris-Tzavalis unit root test is likely to be distorted. To avoid this problem, Kruiniger & 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 tests when the errors \( \varepsilon_{it} \) are homoskedastic 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 2, 4 and 6.

- 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 \( m_4 \geq \sigma_{4x} \) such that \( (k_1 m_4 + k_2 \sigma_{4x})^{-1} \leq \sigma_{4x}^{-1} (k_1 + k_2)^{-1} \).

- When \( \tau \) is fixed the local power of the OLS test is higher than the local power of the Breitung-Meyer test when \( \frac{\sigma_{2x}^2}{\sigma_{4x}^2} \leq \frac{\sigma_{4x}^2}{\sigma_{2x}^2} \left( 1 + \frac{2\tau}{T-1} \right) \).
• 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 $\rho > \frac{T-5}{T-1}$.

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

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: $\tau = 1$, $\sigma_{2c}^2/\sigma_{4c} = 1$ and $m_4 = 3\sigma_{2c}^2$. The figures correspond to the value of $T+1$ being 5 or 10 and the value of $\sigma_{2c}^2/\sigma_{2c}$ 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 $\Phi \left( -1.645 + c \sqrt{T(T-1)}/2 \right)$ where $\Phi$ denotes the cdf of the standard normal, see Proposition 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 $\sigma_{2c}^2/\sigma_{2c} = 1$ the local power of the OLS test is highest for all values of $c$, whereas when $\sigma_{2c}^2/\sigma_{2c} = 10$ the local power of the OLS test is lowest for all values of $c$. More specifically, when $\sigma_{2c}^2/\sigma_{2c} = 10$ and $T+1 = 5$ then in order to attain a local power level equal to 0.5 for a given value of $\rho$ 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 $\sigma_{2c}^2/\sigma_{2c} = 10$ 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.

Figure 2: Comparison of the local power under mean stationarity
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

\[ y_{i0} = \alpha_i + \varepsilon_{i0} \]  \hspace{2cm} (35)  
\[ y_{it} = \rho y_{i(t-1)} + (1 - \rho) \alpha_i + \varepsilon_{it} \]  \hspace{2cm} (36)

with

\[ \varepsilon_{it} \sim \text{iid} N(0,1) \]
\[ \alpha_i \sim \text{iid} N(0, \sigma^2_\alpha) \]
\[ \varepsilon_{i0} \sim \text{iid} N(0, \tau) \]  \hspace{2cm} (37)

We consider different values of \( T, N \) and \( \rho \) 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 Table 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 brackets. We consider different simulation setups where the value of \( \tau \) 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 setups depend on the variance of initial error term \( \tau \). 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 \( \rho \) 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 \( \sigma^2_\alpha = 1 \) and from 0.09 to 0.21 when \( \sigma^2_\alpha = 10 \). When comparing 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 \( \sigma^2_\alpha = 1 \) (\( \sigma^2_\alpha = 10 \)). 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 \( \rho \) 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 Hall & Mairesse (2005) and Bond, Nauges & Windmeijer (2002) where the time-series processes are covariance stationary in the simulation setups. 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 $\sigma_2^2 = 10$. 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 micro-panels 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 macro-panels. The results show that the OLS unit root test is a 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 homoskedastic variances over term 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 macro-panels, see Moon, Perron & Phillips (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 macro-panels suggest that in this case some of the test considered here might be optimal (the OLS test without incidental intercepts and the Breitung-Meyer test with incidental intercepts), see Moon, Perron & Phillips (2007). These results are also interesting in relation to the type of panel data unit root test suggested by Im, Pesaran & Shin (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 macro-panels the Im-Peseran-Shin test appears to have substantially lower power than the optimal tests and that might also be the case in micro-panels.

Acknowledgements

I wish to thank J. Breitung, M. Browning, S. Johansen, H.C. Kongsted, two anonymous referees and coeditor Pierre Perron for useful comments and suggestions. I also wish to thank the Danish National Research Foundation for its support through Centre for Applied Microeconometrics (CAM).
Table 1: Empirical and analytical (in brackets) rejection probabilities when $\tau = 1$

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

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

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 "$X_N \overset{a.s.}{=} Y_N$" means that $X_N - Y_N \xrightarrow{P} 0$ 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 1 Under the local-to-unity sequence for $\rho$ given by $\rho_N = 1 - c/N^k$ for $k, c > 0$ the following hold

$$\rho_N^t = 1 - t \frac{c}{N^k} + o\left(\frac{1}{N^{k}}\right)$$ (38)

$$\frac{1}{1 - \rho_N^2} = \frac{N^k}{2c} + o\left(\frac{1}{N^{k}}\right)$$ (39)

Proof: The binomial formula yields

$$\rho_N = \left(1 - \frac{c}{N^k}\right)^t = 1 - t \frac{c}{N^k} + \frac{t(t-1)}{2!} \frac{c^2}{N^{2k}} - \frac{t(t-1)(t-2)}{3!} \frac{c^3}{N^{3k}} + \cdots + \left(\frac{-c}{N^k}\right)^t$$ (40)

and the results follow directly.

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

$$y_{it} = (1 - \rho^t) \alpha_i + \rho^t y_{i0} + \rho^{t-1} \varepsilon_{1i} + \cdots + \varepsilon_{it} \quad \text{for } t = 1, \ldots, T$$ (41)

Inserting the expression for the initial value given in Assumption 3 yields

$$y_{it} = \alpha_i + \rho^t \sqrt{T} \varepsilon_{i0} + \rho^{t-1} \varepsilon_{1i} + \cdots + \varepsilon_{it} \quad \text{for } t = 0, \ldots, T$$ (42)

Using stacked notation equation (1) can be expressed as

$$y_i = \rho y_{i,-1} + v_i$$ (43)

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

$$y_{i,-1} = \alpha_i + C_T(\rho) \varepsilon_i + A_T(\rho) \sqrt{T} \varepsilon_{i0}$$ (44)

$$v_i = (1 - \rho) \alpha_{i+1} T + \varepsilon_i$$ (45)

where $\iota_T$ 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

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

$$A_T(\rho) = \begin{bmatrix} 1 \\ \rho \\ \rho^2 \\ \vdots \\ \rho^{T-1} \end{bmatrix}$$ (46)
Note that $C_T (\rho_N) = C_T (1) + O (N^{-k})$ and $A_T (\rho_N) = \nu_T + O (N^{-k})$ when $\rho_N = 1 - c/N^k$ according to Lemma 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 $\rho$. In Lemma 2 below we provide results that are used to prove the propositions in Section 3.1-3.3.

**Lemma 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 $\rho$.

**PART A:** If the following hold for the sequence $\{x_i (\rho)\}_{i=1}^N$ 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 $\rho$.

\[
\frac{1}{N} \sum_{i=1}^N E (x_i (1)' x_i (1)) \rightarrow m_{XX} \quad \text{as } N \rightarrow \infty \quad (47)
\]

\[
E ((x_i (\rho_N) - x_i (1))' x_i (\rho_N)) = o(1) \quad (48)
\]

\[
E ((x_i (\rho_N) - x_i (1))' x_i (1)) = o(1) \quad (49)
\]

then

\[
\frac{1}{N} \sum_{i=1}^N x_i (\rho_N)' x_i (\rho_N) \xrightarrow{P} m_{XX} \quad \text{as } N \rightarrow \infty \quad (50)
\]

**PART B:** If the following hold for the sequence $\{x_i (\rho)\}_{i=1}^N$ 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 $\rho$.

\[
E (x_i (1)' u_i (1)) = 0 \quad (51)
\]

\[
\frac{1}{N} \sum_{i=1}^N \text{Var} (x_i (1)' u_i (1)) \rightarrow \Sigma \quad \text{as } N \rightarrow \infty \quad (52)
\]

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^N E (x_i (\rho_N)' (u_i (\rho_N) - u_i (1))) \rightarrow \mu_1 \quad \text{as } N \rightarrow \infty \quad (53)
\]

\[
\text{Var} (x_i (\rho_N)' (u_i (\rho_N) - u_i (1))) = o(1) \quad (54)
\]

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^N E ((x_i (\rho_N) - x_i (1))' u_i (1)) \rightarrow \mu_2 \quad \text{as } N \rightarrow \infty \quad (55)
\]

\[
\text{Var} ((x_i (\rho_N) - x_i (1))' u_i (1)) = o(1) \quad (56)
\]

then

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^N x_i (\rho_N)' u_i (\rho_N) \xrightarrow{w} N (\mu_1 + \mu_2, \Sigma) \quad \text{as } N \rightarrow \infty \quad (57)
\]

**PART C:** Let $\rho_N$ 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

\[
\hat{V} (\rho) \equiv \frac{1}{N} \sum_{i=1}^N x_i (\rho)' \hat{u}_i (\rho) \hat{u}_i (\rho)' x_i (\rho) \xrightarrow{P} \Sigma \quad \text{as } N \rightarrow \infty \quad (58)
\]

where

\[
\hat{u}_i (\rho) = (\rho_N - \hat{\rho}) x_i (\rho)' + u_i (\rho) \quad (59)
\]
Proof:

PART A: We have that
\[ \frac{1}{N} \sum_{i=1}^{N} x_i (\rho)' x_i (\rho) \]
\[ = \frac{1}{N} \sum_{i=1}^{N} x_i (1)' x_i (1) + \frac{1}{N} \sum_{i=1}^{N} (x_i (\rho) - x_i (1))'(x_i (1) + \frac{1}{N} \sum_{i=1}^{N} (x_i (\rho) - x_i (1))' x_i (\rho) \]  

(60)

The Law of Large Numbers together with the condition in (47) 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 (47) 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 (48)-(49). Altogether this proves that
\[ \frac{1}{N} \sum_{i=1}^{N} x_i (\rho_N)' x_i (\rho_N) \xrightarrow{a.s.} \frac{1}{N} \sum_{i=1}^{N} x_i (1)' x_i (1) \overset{P}{\to} m_{XX} \quad \text{as} \quad N \to \infty \]

(61)

PART B: We have that
\[ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i (\rho)' u_i (\rho) \]
\[ = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i (1)' u_i (1) + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (x_i (\rho) - x_i (1))' u_i (1) + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i (\rho)' (u_i (\rho) - u_i (1)) \]

(62)

The Central Limit Theorem and the Law of Large Numbers together with the conditions in (51)-(56) and the existence of fourth order moments give
\[ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i (1)' u_i (1) \xrightarrow{u} N (0, \Sigma) \quad \text{as} \quad N \to \infty \]

(63)

\[ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (x_i (\rho_N) - x_i (1))' u_i (1) \xrightarrow{P} \mu_1 \quad \text{as} \quad N \to \infty \]

(64)

\[ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i (\rho_N)' (u_i (\rho_N) - u_i (1)) \xrightarrow{P} \mu_2 \quad \text{as} \quad N \to \infty \]

(65)

In particular, the conditions in (53)-(54) together with independency across \( i \) imply the following which give the result in (64)
\[ E \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (x_i (\rho_N) - x_i (1))' u_i (1) \right) \xrightarrow{} \mu_1 \quad \text{as} \quad N \to \infty \]

(66)

\[ \text{Var} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (x_i (\rho_N) - x_i (1))' u_i (1) \right) = \frac{1}{N} \sum_{i=1}^{N} \text{Var} \left( (x_i (\rho_N) - x_i (1))' u_i (1) \right) \xrightarrow{} 0 \quad \text{as} \quad N \to \infty \]

(67)

Altogether this proves that
\[ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i (\rho_N)' u_i (\rho_N) \xrightarrow{u} N (\mu_1 + \mu_2, \Sigma) \quad \text{as} \quad N \to \infty \]

(68)

22
PART C: We use the following definitions

\[ \hat{V}(\rho) = \frac{1}{N} \sum_{i=1}^{N} x_i(\rho) \hat{u}_i(\rho) \hat{u}_i(\rho)^t x_i(\rho) \]  
\[ (69) \]

\[ V(\rho) = \frac{1}{N} \sum_{i=1}^{N} E(x_i(\rho)^t u_i(\rho) u_i(\rho)^t x_i(\rho)) \]  
\[ (70) \]

We have that

\[ \hat{V}(\rho_N) - V(\rho_N) = o_P(1) \]  
\[ (71) \]

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

\[ V(\rho_N) - V(1) \to 0 \text{ as } N \to \infty \]  
\[ (72) \]

\[ \blacksquare \]

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

Using the equation in (9) we have that

\[ N^k (\hat{\rho}_{OLS} - \rho) = \left( \frac{1}{N^{2k}} \sum_{i=1}^{N} y_{i,-1} y_{i,-1} \right)^{-1} \frac{1}{N^k} \sum_{i=1}^{N} y_{i,-1} v_i \text{ for } k > 0 \]  
\[ (73) \]

Proposition 1 now follows by the results in Lemma 3 below.

**Lemma 3** Under Assumption 1-4 and the local-to-unity sequence for \( \rho \) given by \( \rho_N = 1 - c/\sqrt{N} \) for \( c \geq 0 \) and when \( \tau \) is fixed, then the following results hold

\[ \frac{1}{N} \sum_{i=1}^{N} y_{i,-1} y_{i,-1} \overset{p}{\to} T \left( \sigma_0^2 + \left( \tau + \frac{T - 1}{2} \right) \sigma_{2\varepsilon} \right) \text{ as } N \to \infty \]  
\[ (74) \]

\[ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} y_{i,-1} v_i \overset{w}{\to} N \left( cT \sigma_0^2, T \left( \sigma_0^2 \sigma_{2\varepsilon} + \left( \tau + \frac{T - 1}{2} \right) \sigma_{4\varepsilon} \right) \right) \text{ as } N \to \infty \]  
\[ (75) \]

Under Assumption 1-4 and the local-to-unity sequence for \( \rho_N \) 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

\[ \frac{1}{N^2} \sum_{i=1}^{N} y_{i,-1} y_{i,-1} \overset{p}{\to} bT \sigma_{2\varepsilon} \text{ as } N \to \infty \]  
\[ (76) \]

\[ \frac{1}{N} \sum_{i=1}^{N} y_{i,-1} v_i \overset{w}{\to} N(0, bT \sigma_{4\varepsilon}) \text{ as } N \to \infty \]  
\[ (77) \]
Proof of Lemma 3:

The following results will be used below

\[
\begin{align*}
\text{tr} \{ C_T (\rho) \} &= 0 \quad (78) \\
A_T (1) A_T (1) &= I_T T = T \quad (79) \\
\text{tr} \{ C_T (1)' C_T (1) \} &= \text{tr} \{ C_T (1)' C_T (1) \} = \frac{T (T - 2)}{2} \quad (80) \\
A_T (\rho_N) - A_T (1) &= O (N^{-k}) \quad (81) \\
C_T (\rho_N) - C_T (1) &= O (N^{-k}) \quad (82)
\end{align*}
\]

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

\[
\begin{align*}
x_i (\rho) &= y_{i-1} (\rho) = \alpha_i I_T + A_T (\rho) \sqrt{T} \varepsilon_{i0} + C_T (\rho) \varepsilon_i \quad (83) \\
u_i (\rho) &= v_i (\rho) = (1 - \rho) \alpha_i I_T + \varepsilon_i \quad (84)
\end{align*}
\]

where expressions for \( y_{i-1} \) and \( u_i \) are given in equations (44)-(45). This gives that

\[
\begin{align*}
x_i (\rho) - x_i (1) &= (A_T (\rho) - A_T (1)) \sqrt{T} \varepsilon_{i0} + (C_T (\rho) - C_T (1)) \varepsilon_i \quad (85) \\
u_i (\rho) - u_i (1) &= (1 - \rho) \alpha_i I_T \quad (86)
\end{align*}
\]

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

We prove the result in (74) by using Part A in Lemma 2. We have the following

\[
\begin{align*}
\frac{1}{N} \sum_{i=1}^{N} E \left( x_i (1)' x_i (1) \right) &= \frac{1}{N} \sum_{i=1}^{N} E \left( \alpha_i^2 I_T I_T + \frac{1}{N} \sum_{i=1}^{N} \sigma_{i2}^2 T A_T (1)' A_T (1) + \frac{1}{N} \sum_{i=1}^{N} \sigma_{i2}^2 \text{tr} \{ C_T (1)' C_T (1) \} \right) \\
&\rightarrow \sigma_{i2}^2 T + \sigma_{22} T \left( \tau + \frac{T - 1}{2} \right) \quad \text{as } N \rightarrow \infty \quad (87)
\end{align*}
\]

where we have used that \( \alpha_i, \varepsilon_{i0} \) and \( \varepsilon_i \) are independent of each other with mean zero. We also have that for all \( \rho \)

\[
E \left( x_i (\rho)' (x_i (\rho_N) - x_i (1)) \right) = \sigma_{i2}^2 T (A_T (\rho_N) - A_T (1)) + \sigma_{i2}^2 \text{tr} \{ C_T (\rho)' (C_T (\rho_N) - C_T (\rho_N)) \} = O \left( \frac{1}{\sqrt{N}} \right) \quad (88)
\]

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

\[
\frac{1}{N} \sum_{i=1}^{N} x_i (\rho_N)' x_i (\rho_N) \rightarrow T \left( \sigma_{i2}^2 + \left( \tau + \frac{T - 1}{2} \right) \sigma_{22} \right) \quad \text{as } N \rightarrow \infty \quad (89)
\]

This proves the result in (74).
We prove the result in (75) by using Part B in Lemma 2. The mean and variance of $x_i(1)^Tu_i(1)$ are given by

$$E (x_i(1)^Tu_i(1)) = \sigma^2_{\xi} \text{tr} \{C_T (1)\}' = 0$$

$$\text{Var} (x_i(1)^Tu_i(1)) = \sigma^2_{\xi} E (x_i(1)^Tu_i(1)) = \sigma^2_{\xi} \left( T \sigma^2_{\alpha} + \left( \tau + \frac{T - 1}{2} \right) \sigma_{4\xi} \right)$$

such that

$$\frac{1}{N} \sum_{i=1}^{N} \text{Var} (x_i(1)^Tu_i(1)) \rightarrow T \left( \sigma^2_{\alpha} \sigma_{2\xi} + \left( \tau + \frac{T - 1}{2} \right) \sigma_{4\xi} \right) \text{ as } N \rightarrow \infty$$

In addition we have the following results concerning means

$$E \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i (\rho_N)^t (u_i (\rho_N) - u_i (1)) \right) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (1 - \rho_N) i_T t_T E (\alpha_i^2) = c T \sigma^2_{\alpha}$$

$$E \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (x_i (\rho_N) - x_i (1))^t u_i (1) \right) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma^2_{\xi} \text{tr} \{C_T (\rho) - C_T (1)\} = 0$$

For the variances we have the following

$$\text{Var} (x_i(\rho_N)^t (u_i (\rho_N) - u_i (1))) \leq O \left( \frac{1}{N} \right)$$

$$\text{Var} ((x_i (\rho_N) - x_i (1))^t u_i (1)) \leq O \left( \frac{1}{N} \right)$$

This holds since

$$\text{Var} (x_i (\rho_N)^t (u_i (\rho_N) - u_i (1))) \leq E \left( (x_i (\rho_N)^t x_i (\rho_N))^2 \right)^{1/2} E \left( ((u_i (\rho_N) - u_i (1))^t (u_i (\rho_N) - u_i (1)))^2 \right)^{1/2} = O \left( \frac{1}{N} \right)$$

where the inequality sign in the expression above follows by Schwarz’s inequality and the equality sign follows by using that $du_i (\rho_N) = u_i (\rho_N) - u_i (1)$ we have that $E \left( (du_i (\rho_N))^2 \right) = O \left( \frac{1}{N^2} \right)$ and $E \left( (x_i (\rho_N) x_i (\rho_N))^2 \right) = O \left( 1 \right).$ The result in (96) follows by using similar arguments. This means that the conditions in Part B of Lemma 2 are satisfied such that

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i (\rho_N)^t u_i (\rho_N) \stackrel{w}{\rightarrow} N \left( c T \sigma^2_{\alpha}, T \left( \sigma^2_{\alpha} \sigma_{2\xi} + \left( \tau + \frac{T - 1}{2} \right) \sigma_{4\xi} \right) \right) \text{ as } N \rightarrow \infty$$

This proves the result in (75).

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

$$x_i (\rho_N) = A_T (\rho_N) \sqrt{b} \varepsilon_{i0} \stackrel{as}{=} y_{i,-1} (\rho_N) / \sqrt{N}$$

$$u_i (\rho_N) = (1 - \rho_N) i_T \alpha_i + \varepsilon_i$$

We have that

$$E (x_i (1)^t x_i (1)) = \sigma^2_{\xi} b A_T (1)^t A_T (1) = \sigma^2_{\xi} b T$$

$$E (x_i (\rho)^t (x_i (\rho_N) - x_i (1))) = \sigma^2_{\xi} b A_T (\rho)^t (A_T (\rho_N) - A_T (1)) = O \left( \frac{1}{N} \right)$$
Lemma 4 Let Assumption 1-4 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

\[
\frac{1}{N^2} \sum_{i=1}^{N} y_{i,-1} (\rho_N)' y_{i,-1} (\rho_N) \overset{a.s.}{=} \frac{1}{N} \sum_{i=1}^{N} x_i (\rho_N)' x_i (\rho_N) \overset{P}{=} \sigma_{2e} bT \quad \text{as } N \rightarrow \infty \tag{103}
\]

This proves the result in (76). Using that $x_i (\rho)$ and $u_i (\rho)$ are independent for all values of $\rho$ we have

\[
E (x_i (1)' u_i (1)) = 0 \tag{104}
\]

\[
\text{Var} (x_i (1)' u_i (1)) = E \left( (x_i (1)' u_i (1))^2 \right) = \sigma_{1e}^2 \tag{105}
\]

\[
E (x_i (\rho_N)' (u_i (\rho_N) - u_i (1))) = 0 \tag{106}
\]

\[
E \left( (x_i (\rho_N) - x_i (\rho_N)' u_i (1)) \right) = 0 \tag{107}
\]

Altogether this implies that

\[
\frac{1}{N} \sum_{i=1}^{N} y_{i,-1} (\rho_N)' v_i (\rho_N) \overset{a.s.}{=} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i (\rho_N)' u_i (\rho_N) \overset{w}{\rightarrow} N (0, bT \sigma_{4e}) \quad \text{as } N \rightarrow \infty \tag{108}
\]

which proves the result in (77).

\[\blacksquare\]

Proof of Proposition 2:

Part C in Lemma 2 immediately implies that $\hat{V}_{OLS} (k)$ is a consistent estimator as $N \rightarrow \infty$ of the variance in the limiting distribution of $N^k (\hat{\rho}_{OLS} - \rho_N)$ when $\rho_N$ is local-to-unity. Then using the expression for $t_{OLS}$ in (13) we have that

\[
t_{OLS} = \hat{V}_{OLS} (k)^{-\frac{1}{2}} N^k (\hat{\rho}_{OLS} - 1) = \hat{V}_{OLS} (k)^{-\frac{1}{2}} N^k (\hat{\rho}_{OLS} - \rho_N) - c\hat{V}_{OLS} (k)^{-\frac{1}{2}} \tag{109}
\]

Proposition 2 now follows by the results already obtained.

\[\blacksquare\]

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

Using the expressions for $y_{i,-1}$ and $v_i$ given in (44)-(45) we have that

\[
\hat{y}_{i,-1} = y_{i,-1} - \iota_T y_{0} = (A_T (\rho) - \iota_T) \sqrt{\tau} \varepsilon_{i0} + C_T (\rho) \varepsilon_i \tag{110}
\]

\[
\hat{v}_i = v_i + (\rho - 1) \iota_T y_{i0} = (\rho - 1) \sqrt{\tau} \varepsilon_{i0} + \varepsilon_i \tag{111}
\]

Using the equation in (17) we have

\[
\sqrt{N} (\hat{\rho}_0 - \rho) = \left( \frac{1}{N} \sum_{i=1}^{N} \hat{y}_{i,-1} \hat{y}_{i,-1} \right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \hat{y}_{i,-1} \hat{v}_i \tag{112}
\]

Proposition 3 now follows by the results in Lemma 4 below.

\[\textbf{Lemma 4} \quad \text{Let Assumption 1-4 be satisfied. When } \rho_N = 1 - c/\sqrt{N} \text{ and } \tau_N = b\sqrt{N} + o(\sqrt{N}) \text{ for } b,c \geq 0, \text{ then the following results hold}
\]

\[
\frac{1}{N} \sum_{i=1}^{N} \hat{y}_{i,-1}' \hat{y}_{i,-1} \overset{P}{\rightarrow} \sigma_{2e} T (T - 1) \quad \text{as } N \rightarrow \infty \tag{113}
\]

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \hat{y}_{i,-1}' \hat{v}_i \overset{w}{\rightarrow} N \left( c^2 b \sigma_{2e} T (T - 1), \sigma_{4e} T (T - 1) \right) \quad \text{as } N \rightarrow \infty \tag{114}
\]
Proof of Lemma 4:

In the following we will use that

\[ A_T (\rho_N) - A_T (1) = (\rho_N - 1) \tilde{A}_T + o \left( \frac{1}{\sqrt{N}} \right) \]  \hspace{1cm} (115)

\[ T \frac{\tilde{A}_T}{T-1} \]  \hspace{1cm} (116)

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

We use the following specifications

\[ x_i (\rho) = \tilde{y}_{i-1} (\rho) = (A_T (\rho) - \nu_T) \sqrt{T} \varepsilon_{i0} + C_T (\rho) \varepsilon_i \]  \hspace{1cm} (117)

\[ u_i (\rho) = \tilde{v}_i (\rho) = (\rho - 1) \sqrt{T} \nu_T \varepsilon_{i0} + \varepsilon_i \]  \hspace{1cm} (118)

such that

\[ x_i (\rho) - x_i (1) = (A_T (\rho) - \nu_T) \sqrt{T} \varepsilon_{i0} + (C_T (\rho) - C_T (1)) \varepsilon_i \]  \hspace{1cm} (119)

\[ u_i (\rho) - u_i (1) = (\rho - 1) \sqrt{T} \nu_T \varepsilon_{i0} \]  \hspace{1cm} (120)

It follows immediately that

\[ \frac{1}{N} \sum_{i=1}^{N} E \left( x_i (1)' x_i (1) \right) = \frac{1}{N} \sum_{i=1}^{N} \sigma_{\varepsilon_i}^2 \text{tr} \left\{ C_T (1)' C_T (1) \right\} \rightarrow \sigma_{\varepsilon_i}^2 \frac{T(T-1)}{2} \]  \hspace{1cm} as \( N \rightarrow \infty \)  \hspace{1cm} (121)

We also have

\[ E \left( (x_i (\rho_N) - x_i (1))' x_i (\rho_N) \right) = \sigma_{\varepsilon_i}^2 \left( \tau_N (A_T (\rho_N) - \nu_T)' (A_T (\rho_N) - \nu_T) + \text{tr} \left\{ (C_T (\rho_N) - C_T (1))' C_T (\rho_N) \right\} \right) \]

\[ = O \left( \frac{1}{\sqrt{N}} \right) \]  \hspace{1cm} (122)

This holds since \( \tau_N (A_T (\rho_N) - \nu_T)' (A_T (\rho_N) - \nu_T) = c^2 b/\sqrt{N} \tilde{A}_T A_T + o \left( \frac{1}{\sqrt{N}} \right) = O \left( \frac{1}{\sqrt{N}} \right) \). Altogether this gives the result in (113) according to Part A of Lemma 2.

In order to prove the result in (114) we will show that the conditions in Part B of Lemma 2 are satisfied. We have that

\[ E \left( x_i (1)' u_i (1) \right) = \sigma_{\varepsilon_i}^2 \text{tr} \left\{ C_T (1)' \right\} = 0 \]  \hspace{1cm} (123)

\[ \text{Var} \left( x_i (1)' u_i (1) \right) = \sigma_{\varepsilon_i}^2 \frac{T(T-1)}{2} \]  \hspace{1cm} (124)

\[ E \left( (x_i (\rho_N)' (u_i (\rho_N) - u_i (1))) \right) = (\rho_N - 1) \tau_N \sigma_{\varepsilon_i}^2 \nu_T (A_T (\rho_N) - A_T (1)) \]  \hspace{1cm} (125)

\[ E \left( ((x_i (\rho_N) - x_i (1))' u_i (1)) \right) = 0 \]  \hspace{1cm} (126)

\[ \text{Var} \left( x_i (\rho_N)' (u_i (\rho_N) - u_i (1)) \right) \leq O \left( \frac{1}{\sqrt{N}} \right) \]  \hspace{1cm} (127)

\[ \text{Var} \left( ((x_i (\rho_N) - x_i (1))' u_i (1)) \right) \leq O \left( \frac{1}{\sqrt{N}} \right) \]  \hspace{1cm} (128)
such that
\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} E \left( x_i (\rho_N)' (u_i (\rho_N) - u_i (1)) \right) \rightarrow c^2 b \frac{T (T - 1)}{2} \sigma_{2e} \text{ as } N \rightarrow \infty
\] (129)
where we have used that \( u_T' (A_T (\rho_N) - A_T (1)) = -c u_T' A_T / \sqrt{N} + o \left( \frac{1}{\sqrt{N}} \right) \) and \((\rho_N - 1) \tau_N = -cb + o \left( \frac{1}{\sqrt{N}} \right).\) Altogether by Part B of Lemma 2 this proves the result in (114). \( \blacksquare \)

**Proof of Proposition 4:**
Part C in Lemma 2 immediately implies that \( \hat{V}_0 \) is a consistent estimator as \( N \rightarrow \infty \) of the variance in the limiting distribution of \( \sqrt{N} (\hat{\rho}_0 - \rho_N) \) when \( \rho_N \) is local-to-unity. Then using the expressions for \( t_0 \) and \( \bar{t}_0 \) we have that
\[
t_0 = \hat{V}_0^{-\frac{1}{2}} \sqrt{N} (\hat{\rho}_0 - 1) = \hat{V}_0^{-\frac{1}{2}} \sqrt{N} (\hat{\rho}_0 - \rho_N) - c \hat{V}_0^{-\frac{1}{2}}
\] (130)
\[
\bar{t}_0 = \sqrt{\frac{T (T - 1)}{2}} \sqrt{N} (\hat{\rho}_0 - 1) = \sqrt{\frac{T (T - 1)}{2}} \sqrt{N} (\hat{\rho}_0 - \rho_N) - c \sqrt{\frac{T (T - 1)}{2}}
\] (131)
Proposition 4 now follows by the results already obtained. \( \blacksquare \)

**A.4 Proofs of the propositions in Section 3.3: Harris-Tzavalis**
Using the expressions for \( y_{i,-1} \) and \( v_i \) given in (44)-(45) and that \( Q_{T' T} = 0 \) we have
\[
Q_T y_{i,-1} = Q_T C_T (\rho) \varepsilon_i + Q_T A_T (\rho) \sqrt{\tau} \varepsilon_{i,0} \quad (132)
\]
\[
Q_T v_i = Q_T \varepsilon_i \quad (133)
\]
where \( Q_T = I_T - \frac{1}{2} \nu_T' \nu_T \) is symmetric and idempotent. Using the expression for \( \hat{\rho}_{WG} \) in (25) we have
\[
\sqrt{N} \left( \hat{\rho}_{WG} - \rho + \frac{3}{T + 1} \right) = \left( \frac{1}{N} \sum_{i=1}^{N} y_{i,-1} Q_T y_{i,-1} \right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( y_{i,-1} Q_T \varepsilon_i + \frac{3}{T + 1} y_{i,-1} Q_T y_{i,-1} \right) \quad (134)
\]
Proposition 5 now follows by the results in Lemma 5 below.

**Lemma 5** Let Assumption 1-4 be satisfied. When \( \rho_N = 1 - c / \sqrt{N} \) and \( \tau_N = b \sqrt{N} + o \left( \sqrt{N} \right) \) for \( b, c \geq 0, \) then the following results hold
\[
\frac{1}{N} \sum_{i=1}^{N} y_{i,-1} Q_T y_{i,-1} \xrightarrow{p} \sigma_{2e} \frac{(T - 1) (T + 1)}{6} \text{ as } N \rightarrow \infty
\] (135)
\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( y_{i,-1} Q_T \varepsilon_i + \frac{3}{T + 1} y_{i,-1} Q_T y_{i,-1} \right) \xrightarrow{w} N (-c \sigma_{2e} (a_1 - c b a_2), g_1 m_4 + g_2 \sigma_{4e}) \text{ as } N \rightarrow \infty
\] (136)
where
\[
a_1 = \frac{(T - 1) (T - 2)}{12} \quad a_2 = \frac{T (T - 1)}{4}
\] (137)
and
\[
g_1 = \frac{(T - 1) (T - 2) (2 T - 1)}{15 T (T + 1)} \quad g_2 = \frac{(T - 1) (17 T^3 - 44 T^2 + 77 T - 24)}{60 T (T + 1)}
\] (138)
Proof of Lemma 5:

The following results will be used below

\[ Q_T A_T (1) = Q_T t_T = 0 \]  
\[ Q_T A_T (\rho) = Q_T (A_T (\rho) - A_T (1)) \]  
\[ \text{tr} \{ C_T (1)' Q_T C_T (1) \} = \frac{(T - 1)(T + 1)}{6} \]  
\[ \text{tr} \{ C_T (1)' t_T' C_T (1) \} = \frac{T (T - 1)(2T - 1)}{6} \]  
\[ \text{tr} \{ C_T (1)' Q_T \} = - \frac{T - 1}{2} \]  

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

\[
\tilde{C}_T = \begin{bmatrix}
0 & 0 & \cdots & 0 & 0 \\
0 & 0 & \cdots & \vdots & \\
1 & \ddots & \ddots & \vdots & \\
\vdots & \ddots & \ddots & 0 & \\
T - 2 & \cdots & 1 & 0 & 0
\end{bmatrix} \quad \tilde{A}_T = \begin{bmatrix}
0 \\
1 \\
2 \\
\vdots \\
T - 1
\end{bmatrix}
\]  

(144)

We will also use the following results

\[
(A_T (\rho_N) - A_T (1))' Q_T A_T (\rho_N) = (1 - \rho_N)^2 \tilde{A}_T' Q_T \tilde{A}_T + o(1/N)
\]  
\[ \text{tr} \{ C_T (1)' \tilde{C}_T \} = \frac{T (T - 1)(T - 2)}{6} \]  
\[ \tilde{A}_T' C_T (1)' \tilde{C}_T = \frac{T (T - 1)(T - 2)(3T - 1)}{24} \]  
\[ \tilde{A}_T' C_T (1)' \tilde{C}_T t_T = \frac{T (T - 1)(T - 2)}{6} \]  
\[ \tilde{A}_T' Q_T \tilde{A}_T = \tilde{A}_T' \tilde{A}_T - \frac{1}{T} (\tilde{A}_T' t_T)^2 = \frac{T (T - 1)(T + 1)}{12} \]  

(148) (149)

Using the equation in (144) we have the following

\[
\sqrt{N} \left( \rho_{WG} - \rho + \frac{3}{T + 1} \right) = \left( \frac{1}{N} \sum_{i=1}^{N} x_i (\rho)' x_i (\rho) \right)^{-1} \left( \frac{1}{N} \sum_{i=1}^{N} x_i (\rho)' u_i (\rho) \right)
\]  

(150)

with

\[
x_i (\rho) = Q_T y_{i,-1} (\rho) = Q_T A_T (\rho) \sqrt{T} \varepsilon_{i0} + Q_T C_T (\rho) \varepsilon_i
\]  
\[
u_i (\rho) = Q_T \varepsilon_i + \frac{3}{T + 1} Q_T y_{i,-1} (\rho) = Q_T \varepsilon_i + \frac{3}{T + 1} (Q_T A_T (\rho) \sqrt{T} \varepsilon_{i0} + Q_T C_T (\rho) \varepsilon_i)
\]  

(151) (152)

We have that

\[
x_i (1) = Q_T C_T (1) \varepsilon_i
\]  
\[
u_i (1) = Q_T \varepsilon_i + \frac{3}{T + 1} Q_T C_T (1) \varepsilon_i
\]  
\[
x_i (\rho) - x_i (1) = Q_T (A_T (\rho) - t_T) \sqrt{T} \varepsilon_{i0} + Q_T (C_T (\rho) - C_T (1)) \varepsilon_i
\]  
\[
u_i (\rho) - u_i (1) = \frac{3}{T + 1} (Q_T (A_T (\rho) - t_T) \sqrt{T} \varepsilon_{i0} + Q_T (C_T (\rho) - C_T (1)) \varepsilon_i)
\]  

(153) (154) (155) (156)
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

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

(157)

and also

$$E ( (x_i(\rho_N) - x_i (1))' x_i (\rho_N) )$$

$$= \sigma_{iz}^2 \text{tr} \{ (C_T(\rho_N) - C_T (1))' Q_T C_T (\rho_N) \} + \tau_N (A_T(\rho_N) - A_T (1))' Q_T A_T (\rho_N)$$

$$= \mathcal{O} \left( \frac{1}{\sqrt{N}} \right)$$

(158)

This holds since $\tau_N = b\sqrt{N} + o(\sqrt{N})$ in combination with $(A_T(\rho_N) - A_T (1))' Q_T A_T (\rho_N) = \mathcal{O}(1/N)$. Altogether this gives the result in (135) according to Part A of Lemma 2.

We prove (136) by showing that the conditions in Part B of Lemma 2 are satisfied. We have the following

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

(159)

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

$$E \left( (\varepsilon'_i C_T (1)' Q_T \varepsilon_i)^2 \right) = \frac{2(T - 1)(T - 1)}{6T} E (\varepsilon_i^4) + \frac{(T - 1)(2T^2 - 4T + 3)}{6T} \sigma_{iz}^4$$

$$E \left( (\varepsilon'_i C_T (1)' Q_T C_T (1) \varepsilon_i)^2 \right) = \frac{(T^2 - 1)(T^2 + 1)}{30T} E (\varepsilon_i^4) + \frac{(T^2 - 1)(T^2 + 1)(T - 2)}{20T} \sigma_{iz}^4$$

$$E (\varepsilon'_i C_T (1)' Q_T \varepsilon_i \varepsilon'_i C_T (1)' Q_T C_T (1) \varepsilon_i) = -\frac{(T^2 - 1)}{12} E (\varepsilon_i^4) - \frac{(T^2 - 1)(T - 2)}{12} \sigma_{iz}^4$$

This gives that

$$\text{Var} (x_i (1)' u_i (1)) = E \left( (\varepsilon'_i C_T (1)' Q_T \varepsilon_i)^2 \right) + \left( \frac{3}{T - 1} \right)^2 E \left( (\varepsilon'_i C_T (1)' Q_T C_T (1) \varepsilon_i)^2 \right)$$

$$+ \frac{6}{T + 1} E (\varepsilon'_i C_T (1)' Q_T \varepsilon_i \varepsilon'_i C_T (1)' Q_T C_T (1) \varepsilon_i)$$

$$= g_1 E (\varepsilon_i^4) + g_2 \sigma_{iz}^4$$

(160)

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

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

(161)

For the first mean term we have that

$$E ( (x_i (\rho_N) - x_i (1))' u_i (1) ) = \sigma_{iz}^2 \text{tr} \left\{ \left( I_T + \frac{3}{T + 1} C_T (1) \right)' Q_T (C_T (\rho_N) - C_T (1)) \right\}$$

$$= \sigma_{iz}^2 \text{tr} \left\{ \left( I_T + \frac{3}{T + 1} C_T (1) \right)' Q_T \hat{C}_T \right\} (\rho_N - 1) + o \left( \frac{1}{\sqrt{N}} \right)$$

30
such that as $N \to \infty$
\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} E \left( (x_i (\rho_N) - x_i (1))' u_i (1) \right) = -c \sigma_2 \left( \text{tr} \left\{ Q_T \tilde{C}_T \right\} + \frac{3}{T+1} \text{tr} \left\{ C_T (1)' Q_T \tilde{C}_T \right\} \right)
\] (162)

For the second mean term we have that
\[
E \left( (x_i (\rho_N)' (u_i (\rho_N) - u_i (1))) \right) = \frac{\sigma_N^2}{3} \left( \text{tr} \left\{ C_T (\rho_N)' Q_T (C_T (\rho_N) - C_T (1)) \right\} + \tau_N A_T (\rho_N)' Q_T (A_T (\rho_N) - A_T (1)) \right) + \frac{3}{T+1} \left( \text{tr} \left\{ C_T (1)' Q_T \tilde{C}_T \right\} (\rho_N - 1) + \tau_N (\rho_N - 1)^2 \tilde{A}_T Q_T \tilde{A}_T \right) + o \left( \frac{1}{\sqrt{N}} \right)
\]
such that as $N \to \infty$
\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} E \left( (x_i (\rho_N)' (u_i (\rho_N) - u_i (1))) \right) \to -c \sigma_2 \frac{3}{T+1} \left( \text{tr} \left\{ C_T (1)' Q_T \tilde{C}_T \right\} - c b \tilde{A}_T Q_T \tilde{A}_T \right)
\] (163)

For the variance terms we have that
\[
\text{Var} \left( (x_i (\rho_N) - x_i (1))' u_i (1) \right) \leq O \left( \frac{1}{\sqrt{N}} \right)
\] (164)
\[
\text{Var} \left( (x_i (\rho_N)' (u_i (\rho_N) - u_i (1))) \right) \leq O \left( \frac{1}{\sqrt{N}} \right)
\] (165)

Altogether this gives the result in (136) according to Part B of Lemma 2 since
\[
-c \sigma_2 \left( \text{tr} \left\{ Q_T \tilde{C}_T \right\} + \frac{6}{T+1} \text{tr} \left\{ C_T (1)' Q_T \tilde{C}_T \right\} - c b \frac{3}{T+1} \tilde{A}_T Q_T \tilde{A}_T \right)
\] = \[-c \sigma_2 \left( \frac{(T-1)(T-2)}{6} + \frac{6(T-1)(T-2)}{24} - c b \frac{T(T-1)}{4} \right)
\] = \[-c \sigma_2 \left( \frac{(T-1)(T-2)}{12} - c b \frac{T(T-1)}{4} \right)
\] (166)

**Proof of Proposition 6:**

Part C in Lemma 2 immediately implies that $\tilde{V}_{WG}$ 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 $\rho_N$ is local-to-unity. Then using the expressions for $t_{WG}$ and $\bar{t}_{WG}$ we have that
\[
t_{WG} = \tilde{V}_{WG}^{-\frac{1}{2}} \sqrt{N} \left( \hat{\rho}_{WG} - 1 + \frac{3}{T+1} \right) = \tilde{V}_{WG}^{-\frac{1}{2}} \sqrt{N} \left( \hat{\rho}_{WG} - \rho_N + \frac{3}{T+1} \right) - c \tilde{V}_{WG}^{-\frac{1}{2}}
\] (167)
\[
\bar{t}_{WG} = \tilde{V}_{WG}^{-\frac{1}{2}} \sqrt{N} \left( \hat{\rho}_{WG} - 1 + \frac{3}{T+1} \right) = \tilde{V}_{WG}^{-\frac{1}{2}} \sqrt{N} \left( \hat{\rho}_{WG} - \rho_N + \frac{3}{T+1} \right) - c \tilde{V}_{WG}^{-\frac{1}{2}}
\] (168)

Proposition 6 now follows by the results already obtained. □

31
References


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.


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


