

The Local Power of Some Unit Root Tests for Panel Data

Jörg Breitung
Humboldt University Berlin
Institute of Statistics and Econometrics
Spandauer Strasse 1, D-10178 Berlin, Germany

Abstract

To test the hypothesis of a difference stationary time series against a trend stationary alternative, Levin and Lin (1993) and Im, Pesaran and Shin (1997) suggest bias adjusted t -statistics. Such corrections are necessary to account for the nonzero mean of the t -statistic in the case of an OLS detrending method. In this paper the local power of panel unit root statistics against a sequence of local alternatives is studied. It is shown that the local power of the test statistics is affected by two different terms. The first term represents the asymptotic effect on the bias due the detrending method and the second term is the usual location parameter of the limiting distribution under the sequence of local alternatives. It is argued that both terms can offset each other so that the test has no power against the sequence of local alternatives. This results suggest to construct test statistics based on alternative detrending methods. We consider a class of t -statistics that do not require a bias correction. The results of a Monte Carlo experiment suggest that avoiding the bias can improve the power of the test substantially.

1 Introduction

In a panel data set, a variable y_{it} is observed for cross section units $i = 1, \dots, N$ in $t = 1, \dots, T$ time periods. A well known problem with such data is unobserved heterogeneity (e.g. Hsiao 1986 and Baltagi 1995). In a univariate time series context heterogeneity may result in individuals specific mean and short run dynamics. For illustration consider an autoregressive process of the form

$$y_{it} = \mu_i + \alpha_i y_{i,t-1} + \varepsilon_{it} , \quad (1)$$

where the error term ε_{it} is assumed to be uncorrelated across i and t . In this model individual heterogeneity is represented by the individual specific parameters μ_i, α_i and $\sigma_i^2 = E(\varepsilon_{it}^2)$. If there are no further assumptions on the parameters, then the data for each cross section unit can be analyzed separately by running N different regressions. In this case, we take no advantage from pooling the data and, thus, inference may be very inefficient. The other extreme is that we ignore a possible heterogeneity altogether and estimate a pooled regression with $\mu_1 = \dots = \mu_N, \alpha_1 = \dots = \alpha_N$ and $\sigma_1^2 = \dots = \sigma_N^2$. Of course, ignoring heterogeneity in the data may result in biased estimates (e.g. Baltagi 1995 p. 3f).

Traditional panel data analysis adopts a compromise between these two extremes and assumes that individual heterogeneity can be represented by an individual specific intercept μ_i alone. Furthermore, one often encounters additional assumptions on the individual effect μ_i , for example, that it is random and uncorrelated with the regressors. The latter model is known as “random-effects model”.

It is not surprising that early work on test for unit roots in panel data starts from the Dickey-Fuller type regression with individual specific intercept (e.g. Breitung 1992). Levin and Lin (*henceforth: LL*) (1993) and Im, Pesaran

and Shin (*henceforth: IPS*) (1997) consider more general models by allowing for individual specific short run dynamics and time trends.

It is well known that the usual dummy variable estimator (or “within-group” estimator) of dynamic models suffers from the so-called “Nickell bias” (Nickell 1981). For stationary models this bias is of order $O(T^{-1})$ and, thus the bias vanishes for $T \rightarrow \infty$. If the autoregressive model has a unit root, however, the Nickell bias does not die out as $T \rightarrow \infty$. The same is true if individual specific time trends are estimated by using the dummy-variables approach. LL (1993) construct a bias adjusted t -statistic to test the null hypothesis of a unit root process. Unfortunately, bias adjusted test statistics for the model with a constant or a time trend can suffer from a severe loss of power. For example, the power of the LL (1993) test without an intercept (and thus without the need to correct for the Nickell bias) against a stationary alternative with an autoregressive coefficient of 0.9 is virtually unity for $N = 25$ and $T = 25$. For the bias adjusted test statistic in the model with individual specific intercept (trend), the power against the same alternative drops to 0.45 (0.25). Furthermore IPS (1997) observe a serious size bias if the bias adjusted LL statistic is augmented with lagged differences.

If there is only a constant in the model, the problem is easily resolved by subtracting the first observation instead of the mean. As argued in Schmidt and Phillips (1992), the first observation is the best estimator of the constant under the hypothesis of a random walk. Furthermore, subtracting the first observation instead of the mean avoids the Nickell bias and, therefore, the test does not require a bias correction (cf Breitung and Meyer (*henceforth: BM*) (1994)).

To study the asymptotic properties we compare the local power of the bias adjusted test statistics. Our analysis demonstrates that the local power

of the test depends on two different terms. The first term represents the asymptotic effect on the bias due the detrending method and the second term is the usual location parameter of the limiting distribution under the sequence of local alternatives. It is shown that if the long-run variances are estimated consistently, both terms cancel out each other so that the test statistic is centered around zero under the local alternative. Levin and Lin (1993), suggest to estimate the long-run variances by using a nonparametric estimator computed from the first differences of the series. An attractive feature of this approach is that under the alternative the nonparametric estimator tends to zero so that a the resulting test statistic has power against the sequence of local alternatives. For the IPS test this problem is less severe. Nevertheless, the bias term also leads to a loss of (local) power.

To overcome the problem, a class of t -statistics is suggested that do not require a bias correction. These tests are based on the t -statistic from a simple least-squares regression of transformed variables and it is shown that the limiting distribution of these tests is standard normal. The results of the Monte Carlo experiments suggest that avoiding the detrending bias may improve the power of the test substantially.

The rest of the paper is organized as follows. In Section 2 the details of the test statistics are given. The local power of the tests is analyzed in Section 3. In Section 4 a class of t -statistics is suggested in order to avoid the detrending bias. Since the test are based on asymptotic properties, it is interesting to consider the relative performance of the tests in small samples, in particular, if the data are generated by a higher order autoregressive process. This problem is addressed in Section 5 by using Monte Carlo simulations. Furthermore, the actual power against a sequence of local alternatives are investigated by means of Monte Carlo simulations. Section 6 offers some

conclusions and makes suggestions for further research.

Finally, a word on the notational conventions applied in this paper. A standard Brownian motion is written as $W_i(r)$. Although there are different Brownian motions for different cross section units i , we sometimes drop the index i for convenience. This has no consequences for the final results since they depend on the *expectation* of the stochastic functionals. Furthermore, if there is no risk of misunderstanding, we drop the limits and the argument r (or dr). For example, the term $\int_0^1 rW_i(r)dr$ will be economically written as $\int rW$. A detrended Brownian motion is represented as $V(r) \equiv V = W - \int W - 12r \int rW$. As usual in this kind of literature we use $[a]$ to indicate the integer part of a .

2 The test statistics

Assume that the variable y_{it} can be represented as

$$y_{it} = \mu_i + \beta_i t + x_{it} , \quad t = 1, 2, \dots, T , \quad (2)$$

where x_{it} is generated by the autoregressive process

$$x_{it} = \sum_{k=1}^{p+1} \alpha_{ik} x_{i,t-k} + \varepsilon_{it} \quad (3)$$

and $x_{is} = 0$ for $s \leq 0$. It is assumed that ε_{it} is white noise with $E(\varepsilon_{it}^2) = \sigma_i^2$ and $E|\varepsilon_{it}|^{2+\delta} < \infty$ for all i, t and some $\delta > 0$. Furthermore ε_{it} is assumed to be independent of ε_{js} for $i \neq j$ and all t and s .

The null hypothesis is that the process is *difference stationary*, i.e.,

$$H_0 : \quad \rho_i \equiv \sum_{k=1}^{p+1} \alpha_{ik} - 1 = 0 \quad \text{for all } i = 1, \dots, N . \quad (4)$$

Under the alternative we assume that y_{it} is (trend) stationary, that is, $\rho_i < 0$ for all i .

The assumptions concerning ε_{it} ensure that there exists a functional limit theorem such that

$$T^{-1/2} \sum_{t=1}^{[rT]} \varepsilon_{it} \Rightarrow \bar{\sigma}_i W_i(r) ,$$

where $W_i(r)$ is a Brownian motion, $\bar{\sigma}_i^2 = \lim_{T \rightarrow \infty} E(T\bar{\varepsilon}_i^2)$ and $\bar{\varepsilon}_i = T^{-1} \sum_{t=1}^T \varepsilon_{it}$ (e.g. Phillips and Solo 1992). The parameter $\bar{\sigma}_i^2$ is sometimes called the “long-run variance”, since it is computed as 2π times the spectral density at frequency zero.

LL (1993) suggest a test procedure against the alternative $\rho_1 = \dots = \rho_N < 0$. Let e_{it} ($v_{i,t-1}$) denote the residuals from a regression of Δy_{it} ($y_{i,t-1}$) on $1, t, \Delta y_{i,t-1}, \dots, \Delta y_{i,t-p}$. Furthermore, let $\tilde{e}_{it} = e_{it}/\sigma_i$ and $\tilde{v}_{it} = v_{it}/\sigma_i$, where in practice σ_i^2 is estimated using the residuals e_{it} . The LL test is based on the bias adjusted t -statistic for $\rho = 0$ in the regression:

$$\tilde{e}_{it} = \rho \tilde{v}_{i,t-1} + \eta_{it} .$$

LL (1993) show that under the null hypothesis, the ordinary t -statistic tends to minus infinity if a constant or a time trend is included in the model. Therefore, they suggest a bias adjusted test statistic given by

$$\lambda_{LL} = \frac{\sum_{i=1}^N \sum_{t=1}^T [\tilde{e}_{it} \tilde{v}_{i,t-1} - (\bar{\sigma}_i/\sigma_i) a_T]}{b_T \sqrt{\sum_{i=1}^N \sum_{t=1}^T \tilde{v}_{i,t-1}^2}} , \quad (5)$$

where a_T and b_T are the small sample analogs of

$$a_\infty = E \int V dV \quad (6)$$

$$b_\infty^2 = \frac{\text{var} [\int V dV]}{E \int V^2} \quad (7)$$

and $V \equiv V(r)$ is a detrended Brownian motion. LL (1993) suggest to use a nonparametric estimator for $\bar{\sigma}_i^2$ based on on the first differences of the data.

IPS (1997) relax the assumption of a common parameter ρ under the alternative. Accordingly, model (2) is estimated for each cross section unit separately, yielding an individual specific Dickey-Fuller t -statistic τ_i . The IPS statistic is given by:

$$\lambda_{IPS} = N^{-1/2} \sum_{i=1}^N [\tau_i - m_T] / \omega_T ,$$

where τ_i is the usual augmented Dickey-Fuller t -statistic for cross section unit i , and m_T, ω_T^2 are small sample analogs of

$$m_\infty = E \left[\frac{\int V dV}{\sqrt{\int V^2}} \right] \quad (8)$$

$$\omega_\infty^2 = \text{var} \left[\frac{\int V dV}{\sqrt{\int V^2}} \right] . \quad (9)$$

If the model does not include a trend term or if it is assumed that the cross section units possess a common trend such that $\beta = \beta_1 = \dots = \beta_N$, the test procedure of Breitung and Meyer (1994) can be used. Here we generalize their procedure by allowing for individual specific short run dynamics. To remove the individual constants, y_{i0} is subtracted so that $E(y_{it} - y_{i0}) = \beta t$. The advantage of this approach to remove the individual effect is that under the null hypothesis $\rho = 0$ the regressor $(y_{i,t-1} - y_{i0})$ is uncorrelated with the error ε_{it} and, therefore, the t -statistic has a standard normal limiting distribution. As the LL test, this test can be performed in two steps. First, we compute u_{it} and $w_{i,t-1}$ as the residuals from regressions of Δy_{it} and $y_{i,t-1}$ on the lagged differences $\Delta y_{i,t-1}, \dots, \Delta y_{i,t-p}$ but not on the deterministic terms. To account for possible heteroscedasticity we standardize the series as $\tilde{u}_{it} = u_{it}/\sigma_i$ and $\tilde{w}_{it} = w_{it}/\sigma_i$, where in practice an estimate of the variance of u_{it} is used.

The test statistic is the t -statistic for $\rho = 0$ in the regression

$$\tilde{u}_{it} = \rho(\tilde{w}_{i,t-1} - \tilde{w}_{i0}) + \beta t + \nu_{it} \quad (10)$$

yielding

$$\lambda_{BM} = \frac{\sum_{i=1}^N \sum_{t=1}^T (\tilde{w}_{i,t-1} - \tilde{w}_{i0}) \tilde{u}_{it}}{\sqrt{\sum_{i=1}^N \sum_{t=1}^T (\tilde{w}_{i,t-1} - \tilde{w}_{i0})^2}} \quad (11)$$

To analyze the asymptotic behavior of the tests, it is important to specify the relationship between N and T (cf Phillips and Moon 1999). For our analysis it is convenient to apply sequential limits denoted by $(T, N \rightarrow \infty)_{seq}$, wherein $T \rightarrow \infty$ is followed by $N \rightarrow \infty$. Although such an asymptotic framework is more restrictive than using a joint limit and requires moment conditions that are difficult to verify (cf IPS 1997), we follow Kao (1999), Moon and Phillips (1999) and others and apply a sequential limit. Whether our results continue to hold for a joint limit theory is an interesting problem for future research.

Theorem 1: *Assume that $\{y_{it}\}$ is a collection of N independent difference stationary time series generated as in (2) with $\rho = 0$ and $\beta_1 = \dots = \beta_N$. For $T \rightarrow \infty$ and $N \rightarrow \infty$, the statistic λ_{BM} defined in (11) has a standard normal limiting distribution.*

In the proof of the theorem it is shown that under the null hypothesis the test statistic can be represented as

$$\lambda_{BM} = z + O_p(N^{-1/2}) + O_p(T^{-1/2}) , \quad (12)$$

where z is a standard normally distributed random variable. The term $O_p(T^{-1/2})$ is due to the estimation of individual specific short run dynamics.

If we are willing to assume that the autoregressive parameters are identical for all cross section units, that is, $\alpha_{ik} = \alpha_{jk}$ for all $i \neq j$, then the last term in (12) cancels and a standard normal limiting distribution is obtained for *fixed* T and $N \rightarrow \infty$ (BM 1994).

3 Local Power

In this section we study the local power of alternative test procedures. Since the test of Theorem 1 is designed to test against an alternative with a common time trend, it is expected that the test lacks power whenever the time series are stationary around *individual specific* time trends. To investigate this problem we consider a sequence of local alternatives given by

$$y_{it} = \mu_i + \beta_i t + x_{it} , \quad (13)$$

where

$$x_{it} = \left(1 - \frac{c}{T\sqrt{N}}\right) x_{i,t-1} + \varepsilon_{it} , \quad c > 0 \quad (14)$$

and β_i is a random variable with

$$\beta_i = \beta + \eta_i ,$$

and the assumptions:

$$E(\eta_i) = 0 , \quad E(\eta_i \varepsilon_{it}) = 0 , \quad E(\eta_i^2) = \sigma_\eta^2 .$$

This random effects specification is used to analyze the asymptotic properties of the test considered in Theorem 1. For all other tests the assumption on the drift parameter is irrelevant.

We will further assume that the initial value of y_{i0} is fixed or stochastic with a finite variance. When the initial conditions are allowed to go into the

remote past, the initial condition plays a role in the limiting distribution of the process (e.g. Phillips and Lee 1996). In what follows, however, we will neglect such complications in order to keep the analysis reasonably simple.

In the following Lemma we state the important fact that under the local alternative the limiting process of x_{it} is the same as under the null hypothesis.

Lemma 1: *Under the local alternative given in (13) – (14) and a sequential limit $(T, N \rightarrow \infty)_{seq}$ we have*

$$T^{-1/2}x_{i,[rT]} \Rightarrow W_i(r), \quad 0 \leq c < \infty.$$

This is an important difference to the asymptotic theory in the usual time series context, where under the local alternative the limiting process is an Ornstein-Uhlenbeck process (cf Phillips 1987).

To investigate the asymptotic properties of the tests we first state a useful lemma.

Lemma 2: *Let $W = W(r)$ denote a standard Brownian motion and $V \equiv V(r)$ is the detrended Brownian motion. Then,*

- (i) $E \int W^2 = 1/2$
- (ii) $E \left(\int W dW \int W^2 \right) = 1/3$
- (iii) $E \left(\int W^2 \right)^2 = 7/12$
- (iv) $E \int V^2 = 1/15$
- (v) $E \left(W(1) \int W \right) = 1/2.$

The probability limits of the tests depend on the parameters σ_i and $\bar{\sigma}_i$. First, we consider the theoretical value of $\bar{\sigma}_i^2$ under the local alternative.

Lemma 3: *Under the local alternative (13) – (14) we have*

$$\bar{\sigma}_i^2 = \lim_{T \rightarrow \infty} E(T^{-1} x_{iT}^2) = \sigma_i^2 .$$

In what follows we derive the result assuming that $\bar{\sigma}_i^2$ is estimated consistently for all values of $c \geq 0$.

The test considered in Theorem 1 is derived under the assumption that all cross section units have the same linear trend. Accordingly, the asymptotic properties of the test crucially depend on the variance of the drift parameter β_i . The following theorem presents the local power for $\sigma_\eta^2 = 0$ and shows that for $\sigma_\eta^2 > 0$ the test is inconsistent.

Theorem 2: *Under the sequence of local alternatives given in (13) – (14) with $\sigma_\eta^2 = 0$ the t -statistic for $\rho = 0$ in (10) is asymptotically distributed as $N(-c/\sqrt{2}, 1)$. For $\sigma_\eta^2 > 0$ the test statistic tends to $+\infty$ at the rate \sqrt{T} and, thus, the test is inconsistent.*

Therefore, for this test statistic it is important that there is a common trend for all cross section units. If the slope of the time trend varies across individuals the test suffers from a lack of power.

Next we consider the bias corrected test statistics. Under the local alternative the bias adjusted (BA) statistic converges to the limit

$$\lambda_{BA}^*(c) = \lim_{N, T \rightarrow \infty} \sqrt{N} \frac{\left[E \left(N^{-1} T^{-1} \sum_{i=1}^N \sum_{t=1}^T \tilde{e}_{it} \tilde{v}_{i,t-1} \right) - N^{-1} \sum_{i=1}^N (\bar{\sigma}_i / \sigma_i) a_\infty \right]}{b_\infty \sqrt{E \left(N^{-1} T^{-2} \sum_{i=1}^N \sum_{t=1}^T \tilde{v}_{i,t-1}^2 \right)}} .$$

Notice that numerator and denominator have been normalized so that both converge to a fixed limit.

Since

$$\tilde{\varepsilon}_{it}\tilde{v}_{i,t-1} = [\sigma_i^{-1}\varepsilon_{it} - c/(T\sqrt{N})\tilde{v}_{i,t-1}]\tilde{v}_{i,t-1}$$

the limit can be written as

$$\lambda_{BA}^*(c) = \frac{\lim_{N,T \rightarrow \infty} \sqrt{N} \left[N^{-1} \sum_{i=1}^N E(\Gamma_{Ti}) - a_\infty \right]}{b_\infty \sqrt{E \int V^2}} - \frac{c\sqrt{E \int V^2}}{b_\infty} \quad (15)$$

where we use $\bar{\sigma}_i/\sigma = 1$ under the local alternative and

$$\Gamma_{Ti} = T^{-1} \sum_{t=1}^T \sigma_i^{-1} \varepsilon_{it} \tilde{v}_{i,t-1} .$$

It turns out that the limit of the bias adjusted statistic depends on two different terms on the right hand side of (15). The first term is due to the detrending method represented by the statistic Γ_{Ti} . The second term is proportional to $\sqrt{E \int V^2}$ and is similar to the usual location parameter in the asymptotic distribution under the null hypothesis. For example, in the simple regression model $y_t = x_t\beta + u_t$ with stationary variables, the location parameter is proportional to $\sqrt{E(x_t^2)}$.

It is important to notice that the expectation of Γ_{Ti} enters the test statistic with the factor \sqrt{N} and, therefore, for the asymptotic analysis the expectation must be determined with an accuracy up to $O(N^{-1/2})$. The following Lemma provides an approximation of this expectation that is sufficient for our purpose.

Lemma 4: *Under the local alternative given in (13) – (14) the asymptotic expectation of Γ_{Ti} is given by*

$$\lim_{T \rightarrow \infty} E(\Gamma_{Ti}) = (1/15)c/\sqrt{N} - 0.5 + o(N^{-1/2}) .$$

Since the result of Lemma 4 is crucial for the local power of the bias adjusted test, the accuracy of the approximation is investigated in a Monte Carlo experiments. First, we generate 10,000 realizations of Γ_{Ti} by letting $T = 200$, $c = 5$ and repeat the experiment with various values for N .¹ If Lemma 4 is correct, a regression of the sample means of Γ_{Ti} on c/\sqrt{N} and a constant should yield an estimate for the intercept close to -0.5 and a slope of roughly $1/15=0.067$. Using $N \in \{30, 35, 40, \dots, 500\}$ the following regression function was obtained for the 71 realizations:

$$E(\Gamma_{Ti}) \simeq \begin{array}{r} -0.495 \\ (0.00060) \end{array} + \begin{array}{r} 0.0629 \\ (0.0016) \end{array} c/\sqrt{N}, \quad (16)$$

where standard errors are given in parentheses. The estimated slope coefficient is slightly smaller than 0.067 and a test of the hypothesis that the true value is $1/15$ yields a t -statistic of -2.44 . Although, the t -statistic is significant with respect to a significance level of 0.05, the approximation in Lemma 4 seems to perform fairly well in finite samples.

Now we present the limiting distribution of the bias adjusted test statistic.

Theorem 3: *Consider a sequence of local alternatives given in (13) – (14). If the estimator for $\bar{\sigma}_i$ converges weakly to σ_i , the bias adjusted test statistic is asymptotically distributed as $\mathcal{N}(0, 1)$.*

It turns out that the bias adjusted test can fail to have power against the sequence of local alternatives. This finding suggests that the power may be improved by a modification that avoids the bias correction altogether. Such

¹We repeated the experiment for different values of c and T . The results turns out to be fairly robust.

a modified test procedure is suggested in Section 4.

It is important to notice that the test suggested by LL (1993) test employs a nonparametric estimator that converge to zero for a stationary alternative. Therefore, the unit root tests are inconsistent if the long-run variance is estimated by using the differences of the time series (cf Phillips and Ouliaris 1990, Theorem 5.3). Phillips and Perron (1988) suggest to estimate $\bar{\sigma}_i^2$ by using the residuals of the autoregression instead of using Δy_{it} . However, if the residuals are used to estimate $\bar{\sigma}_i^2$, then the estimated long-run variance converge to σ_i^2 (Phillips and Perron 1988) and Theorem 3 implies that such a test has no power against the local alternative.

Finally the local power of the IPS test is investigated. As in the case of the bias adjusted statistic considered above, the probability limit of the test statistic depends on two terms. The first term is due to the detrending method and depends on

$$\Gamma_{Ti}^* = \frac{\sum_{t=1}^T \sigma_i^{-1} \varepsilon_{it} \tilde{v}_{i,t-1}}{\sqrt{\sum_{t=1}^T \tilde{v}_{i,t-1}^2}} .$$

Since this statistic is a ratio of correlated random variables, the analytic evaluation of this bias is very complicated. To obtain a suitable approximation we apply a similar simulation technique that was also used to check the reliability of Lemma 4. Using the same setup as before the following approximation is found for the expectation of Γ_{Ti}^* :

$$E(\Gamma_{Ti}^*) \simeq \begin{matrix} -2.151 & + & 0.212 c/\sqrt{N} \\ (0.0030) & & (0.0077) \end{matrix} \quad (17)$$

This approximation can be used to compute the limiting distribution of the IPS test given in

Theorem 4: For a sequence of local alternative given in (13) – (14) the LL test is asymptotically distributed as $\mathcal{N}(\theta_c^{IPS}, 1)$, where

$$\theta_c^{IPS} = \frac{c}{\omega_\infty} \left[\lim_{T \rightarrow \infty} \frac{\partial E(\Gamma_{Ti}^*)}{\partial(c/\sqrt{N})} \Big|_{c=0} - E\sqrt{\int V^2} \right].$$

Again we find that the local power depends on two terms. Our Monte Carlo experiment suggests that the derivative of $E(\Gamma_{Ti}^*)$ is positive so that the detrending bias implies a substantial loss of power.

Using 10,000 Monte Carlo replications, the expression $E\left(\sqrt{\int V^2}\right)$ is estimated as 0.243. It is interesting to compare our results for the IPS statistic to the local power of the statistic λ_{BM} that is considered in Theorem 2. Since $2^{-1/2} = 0.707$ is larger than $(0.212 - 0.243)/\omega_{100} = 0.0401$, where $\omega_{100} = \sqrt{0.597}$ is taken from IPS (1997), it is seen that the local power of the statistic λ_{BM} is much higher than the local power of the IPS test. However, this is no surprise, since the former statistic is based on much more restrictive assumptions (common time trend, homogeneous alternatives) than the IPS statistic.

4 Test statistics without bias adjustment

From the local power analysis we found that bias corrections used for the LL and IPS tests may imply a severe loss of power. It is therefore desirable to avoid the bias term when constructing the t -statistics. For the case that the model includes only a constant, such an unbiased statistic is easily obtained by subtracting the first observation instead of the individual mean. This is the approach used in BM (1994). However, it was shown in the previous section that the test procedure may become inconsistent against a sequence

of local alternatives. In this section we consider a class of test statistics that do not involve a bias term.²

To facilitate the exposition we will assume that the data is generated by an AR(1) process and, thus, no augmentation with lagged differences is needed. For higher order processes, Δy_{it} and $y_{i,t-1}$ are replaced by the residuals from regressions of Δy_{it} and $y_{i,t-1}$ on $\Delta y_{i,t-1}, \dots, \Delta y_{i,t-p}$. Furthermore, we define the $T \times 1$ vectors $\mathbf{y}_i = [\Delta y_{i1}, \dots, \Delta y_{iT}]'$ and $\mathbf{x}_i = [y_{i0}, \dots, y_{i,T-1}]'$. In order to construct an unbiased test statistic we use the transformed vectors $\mathbf{y}_i^* = \mathbf{A}\mathbf{y}_i = [y_{i1}^*, \dots, y_{iT}^*]'$ and $\mathbf{x}_i^* = \mathbf{B}\mathbf{x}_i = [x_{i1}^*, \dots, x_{iT}^*]'$ such that

$$E(\mathbf{y}_i^* | \mathbf{x}_i^*) = 0 \quad (18)$$

for all i . Imposing further assumptions to rule out degenerate cases it is possible to show that a t -statistic based on the transformed variables has a standard normal limiting distribution.

Theorem 5: *Let Δy_{it} be white noise with $E(\Delta y_{it}) = \beta_i$, $E(\Delta y_{it} - \beta_i)^2 = \sigma_i^2 > 0$ and $E(\Delta y_{it} - \beta_i)^4 < \infty$. Under the assumption (18) and*

$$\begin{aligned} \lim_{T \rightarrow \infty} E(T^{-1} \mathbf{y}_i^{*'} \mathbf{y}_i^*) &> 0 \\ \lim_{T \rightarrow \infty} E(T^{-1} \mathbf{x}_i^{*'} \mathbf{A}' \mathbf{A} \mathbf{x}_i^*) &> 0 \end{aligned}$$

the statistic

$$\lambda_{UB} = \frac{\sum_{i=1}^N \sigma_i^{-2} \mathbf{y}_i^{*'} \mathbf{x}_i^*}{\sqrt{\sum_{i=1}^N \sigma_i^{-2} \mathbf{x}_i^{*'} \mathbf{A}' \mathbf{A} \mathbf{x}_i^*}}$$

has a standard normal limiting distribution as $(N, T \rightarrow \infty)_{seq}$.

²Another possibility is to use alternative estimation methods like the Generalized Methods of Moments (GMM). Breitung (1997) apply second differences and obtains a unit root test without bias adjustment by using an appropriate GMM estimator.

It is important to notice that for this result the lagged levels need not be trend-adjusted. However, if y_{it} has a (individual specific) time trend under the alternative, then a test statistic without adjusting for the trend can be shown to converge to zero in probability and, thus, fails to be consistent.

A simple way to satisfy the assumptions (18) is to use an upper triangular matrix \mathbf{A} , where the elements of each row sum to zero. In other words, only the present and future observations are used to transform the differences Δy_{it} . A well known example for such a transformation is the Helmert transformation given by

$$y_{it}^* = s_t \left[\Delta y_{it} - \frac{1}{T-t} (\Delta y_{i,t+1} + \dots + \Delta y_{iT}) \right], \quad t = 1, 2, \dots, T-1 \quad (19)$$

where $s_t^2 = (T-t)/(T-t+1)$. This transformation is also used in Arellano and Bover (1995), for example. An important property of this transformation is that whenever Δy_{it} is a white noise process with constant variance, then the same is true for y_{it}^* . Obviously, if y_{it} is a random walk with (individual specific) time trend, then y_{it}^* has a zero mean and is uncorrelated with $y_{i,t-1}$.

Next we consider the local power of the class of unbiased test statistic given in Theorem 5. As argued before, it is natural to demean Δy_{it} by using the “forward filter” given in (19). In this case the matrix \mathbf{A} is an upper triangular matrix. A natural way to adjust the lagged levels for the trend is to use a “backward filter” based on the past observations of the process. In this case, the assumption in (18) is obviously fulfilled. However, the following theorem shows that the resulting test statistic has an expectation of zero under the local alternative.

Theorem 6: *Let \mathbf{A} be an upper triangular matrix with the property that the elements of each row sums to zero and \mathbf{B} is a lower triangular matrix. For a sequence of local alternative given in (13) – (14) the statistic λ_{UB}*

given in Theorem 5 is asymptotically normally distributed with expectation

$$\lim_{T, N \rightarrow \infty} E(\lambda_{UB}) = 0.$$

This result suggests that the matrix \mathbf{A} (resp. \mathbf{B}) should be different from an upper (lower) diagonal matrix. The following test procedure is based on an unbiased test statistic with a forward filter given in (19) but using a matrix \mathbf{B} that is different from a lower triangular matrix:

$$x_{it}^* = y_{i,t-1} - \frac{t-1}{T} y_{iT} . \quad (20)$$

Notice that $T^{-1} y_{iT} = T^{-1} \sum_{t=1}^T \Delta y_{it}$ is an estimator of β_i and, thus, the transformed variable x_{it}^* is adjusted for a time trend. It is easy to verify that in this case $E(\mathbf{y}_i^* | \mathbf{x}_i^*) = 0$ and, thus, Theorem 5 states that under the null hypothesis this test has a standard normal limiting distribution.

Although the asymptotic null distribution of the test statistic does not depend on individual specific constants³ the power of the test depends on the individual specific constants (cf BM 1994). To obtain a test statistic with a power that does not depend on the individual specific constants we subtract the first observation. Furthermore, the series are adjusted for short run dynamics and individual specific variances by replacing y_{it}^* and x_{it}^* by the standardized residuals from the regression of y_{it}^* and x_{it}^* on $y_{i,t-1}^*, \dots, y_{i,t-p}^*$. Let \tilde{y}_{it}^* and \tilde{x}_{it}^* denote the resulting residuals that were standardized using the estimated standard deviations from the residuals corresponding to y_{it}^* . The test statistic is the usual t -statistic for $\phi = 0$ from the regression

$$\tilde{y}_{it}^* = \phi(\tilde{x}_{it}^* - \tilde{x}_{i1}^*) + e_{it}^* \quad t = 2, 3, \dots, T-1 . \quad (21)$$

The resulting test statistic is called the UB statistic in what follows. Since the transformation matrix \mathbf{A} corresponding to the Helmert transformation (19)

³The reason is that the term $x_i^{*'} A' A x_i^*$ is $O_p(T^2)$ and, hence, the individual specific constants do not affect the asymptotic null distribution.

satisfies $\mathbf{A}'\mathbf{A} = \mathbf{I}$ so that according to Theorem 5, the t -statistic is asymptotically standard normal.

To compute the local power function of this test statistic we need an approximation for

$$E(\xi_{Ti}) = E \left[T^{-1} \sum_{t=1}^T y_{it}^* (x_{it}^* - x_{i1}^*) \right]$$

that is accurate up to $O(N^{-1/2})$. Such an approximation is again obtained as for the LL and IPS statistic by the estimated regression function:

$$E(\xi_{Ti}^*) \simeq \begin{array}{cc} -0.0104 & - 0.0407 c/\sqrt{N} \\ (0.0021) & (0.0104) \end{array} .$$

Since the test statistic is constructed to have an expectation of zero under the null hypothesis, we expect to find a constant close to zero. The estimated constant is indeed quite small but nevertheless significant. The slope coefficient is significantly negative so that the test seem to have a local power larger than the size. The following theorem present further details on the local power of the UB statistic.

Theorem 7: *For a sequence of local alternative given in (13) – (14) the UB test is asymptotically distributed as $\mathcal{N}(\theta_c^{UB}, 1)$, where*

$$\theta_c^{UB} = c \sqrt{6} \left[\lim_{T \rightarrow \infty} \frac{\partial E(\xi_{Ti})}{\partial (c/\sqrt{N})} \Big|_{c=0} \right] .$$

It is interesting to compare the local powers of the IPS and the UB test. Since $\sqrt{6} \cdot 0.0407 > 0.0401$, the UB statistic has a location parameter which is more than twice as large in absolute value compared to the IPS statistic. Again, however, we emphasize that this comparison is unfair, because the IPS test is more general than the UB test as it allows for a heterogeneous autoregressive parameter under the alternative.

5 Small sample properties

The asymptotic properties of the tests do not depend on number of lagged differences that are used to account for higher order autoregressive models. However, as noted by IPS (1997) for a small number of time periods T , the null distribution may be substantially affected by the augmentation lag. They therefore present tables for the mean and the variance of τ_i that depend on the type of deterministic (constant/trend), the number of time periods T and the augmentation lag p . For example, if there is a trend in the model and $T = 25$, IPS (1997) report a mean of -2.167 if no lagged difference is included, whereas the mean is -1.968 , if the regression is augmented by four lagged differences. This may not seem to matter very much but it should be noted that the mean between the expectations of the IPS statistics with $p = 4$ and $p = 0$ is $\sqrt{N}(-1.968 + 2.167) = 0.199\sqrt{N}$. Thus, for moderate or large N the test may be very conservative if the dependence on the augmentation lag is ignored.

From the usual Dickey-Fuller test for aggregated time series it is known that the power of the test deteriorates substantially with an increasing augmentation lag. It is therefore expected that also the power of panel unit root tests are affected by the choice of the augmentation lag.

To study the robustness of the size and power of the tests considered in the previous sections we generate time series according to the process

$$x_{it} = \alpha x_{i,t-1} + \varepsilon_{it} \tag{22}$$

and $y_{it} = \mu_i + \beta_i t + x_{it}$. The initial values of the process are set equal to zero. The errors are i.i.d. with $\varepsilon_{it} \sim N(0, 1)$. Since all tests are invariant to the parameters μ_i and β_i , these parameters are set equal to zero. For the bias and variance corrections of the LL and IPS tests the tabulated values in LL

Table 1: Empirical size and power for $T = 30$ and $N = 20$

	LL	IPS	UB	LL	IPS	UB
α	$p = 0$			$p = 1$		
1.00	0.026	0.067	0.067	0.003	0.060	0.061
0.95	0.058	0.109	0.133	0.011	0.092	0.109
0.90	0.176	0.229	0.358	0.038	0.162	0.291
0.80	0.809	0.769	0.916	0.242	0.555	0.829
α	$p = 2$			$p = 4$		
1.00	0.000	0.044	0.060	0.000	0.038	0.065
0.95	0.000	0.063	0.105	0.000	0.053	0.089
0.90	0.000	0.116	0.249	0.000	0.078	0.189
0.80	0.000	0.355	0.756	0.000	0.153	0.624

Note: Empirical sizes computed from 5000 Monte Carlo replications of model (22). p denotes the number of lagged differences. The nominal size is 0.05.

(1993) and IPS (1997) are used. To represent a typical regional panel data set, we let $T = 30$ (years) and $N = 20$ (countries). All rejection frequencies are computed from 1000 realizations with a nominal significance level of 0.05.

It is important to notice that the autoregressive coefficient is identical for all cross section units. Thus, our simulation design favors the LL and the UB statistic that assume a homogenous autoregressive parameter under the alternative.

Table 1 presents the rejection frequencies for the different tests. For $p > 0$ the LL test turns out to be quite conservative. This was also observed by IPS (1997) and, therefore, the values for the mean and variance of this test should also be tabulated for different augmentation lags. With respect to the power of the test it turns out that for $p = 0$ the power of the LL and IPS test are roughly similar. For $p > 0$ the IPS test tends to be more powerful than the LL test, at least if the critical values of the LL test are not adjusted for different augmentation lags.

Table 2: Power against local alternatives

		LL	LL*	IPS	UB
N	T	N and $T \rightarrow \infty$			
25	25	0.378	0.064	0.384	0.668
50	50	0.269	0.056	0.300	0.660
70	70	0.210	0.033	0.296	0.608
100	100	0.170	0.050	0.261	0.579
		T fixed, $N \rightarrow \infty$			
50	25	0.235	0.038	0.342	0.575
70	25	0.156	0.038	0.313	0.535
100	25	0.090	0.028	0.273	0.450
		N fixed, $T \rightarrow \infty$			
25	50	0.415	0.061	0.419	0.724
25	70	0.378	0.020	0.421	0.742
25	100	0.298	0.028	0.402	0.783

Note: This table reports the rejection rates computed from 5000 replications of model (22) with $\alpha = 1 - 20/(T\sqrt{N})$. The significance level is 0.05. The statistic LL* is similar to the LL test but using the residuals from the autoregression to estimate σ_i^* . For this test the values for the expectation and variances are computed by Monte Carlo simulations.

The UB statistic suggested in Section 4 appears to be substantially more powerful than the LL and IPS tests. Furthermore the size and power of the UB test is much more robust with respect to the augmentation lag. Notice that for the UB test no tables are required for different values of p and T .

Next a simulation experiment is performed to assess the validity of the theoretical results for the actual power of the test. For this purpose we set $\alpha = 1 - 20/(T\sqrt{N})$. If the test does not have power against such alternative, we expect that the power of the test tends to the size as $N \rightarrow \infty$ and $T \rightarrow \infty$. In our Monte Carlo comparison we also include a variant of the LL test that estimate the long-run variances by using the residuals instead of the first difference of the process. As shown in Section 3 such a test has a local power

equal to the size. The critical values for this test are computed by Monte Carlo simulation. The resulting test is denoted by LL^* .

Table 2 presents the outcome of such a Monte Carlo experiment. As predicted by Theorem 3, the power of the LL^* test is close the size for all N and T . All other tests appear to converge to a limit larger than than the size, where the limiting power of the UB test is roughly twice as large as the limiting power of the IPS test.

The results of the Monte Carlo experiment can be compared to the results of our theoretical analysis. From Theorem 4 it is expected that the IPS test has a power of $\Phi(z_c)$, where $\Phi(\cdot)$ denotes the c.d.f. of the standard normal distribution and

$$z_c = -1.65 + \frac{c}{\omega_T}(0.031 \pm 0.0077) ,$$

where ± 0.0077 indicates plus/minus one standard deviation of the regression estimate in (17). From $\alpha = 0.9$ we obtain $c = 21.9$ so that Theorem 4 predicts a power in the interval $[0.167, 0.282]$. The empirical power for $N = 100$ and $T = 100$ is 0.261, which corresponds well to the results of Section 3. An analog calculation using the results for the UB statistic yields an (fairly wide) interval of $[0.490, 0.862]$. The empirical power presented in Table 2 fall into this interval so that the simulation results are in line with the findings in Section 3.

Finally it is interesting to note that the power of the tests appears to deteriorate with fixed T and increasing N . For the LL test the local power seem to tend slowly to the size as T is fixed and $T \rightarrow \infty$.

6 Conclusion

In this paper we have considered the local power of some well known tests and a new test for unit roots in panel data. We found that the LL and IPS tests

suffer from a severe loss of power if individual specific trends are included. Therefore, a class of test statistic is suggested that does not employ a bias adjustment and it is found that the power of this test is substantially higher than the LL and the IPS tests. Furthermore, it turns out that the LL test is very sensitive to the augmentation lag and the short run parameter. It is therefore recommended to apply tables for the mean and variance that take into account the lag-augmentation of the test.

The results further indicate that the power of the tests is very sensitive to the specification of the deterministic terms. If there is only a constant or a joint linear trend, then subtracting the first observation yields a very powerful test. Including individual specific trends when it is unnecessary leads to a dramatic loss of power. Hence, in practice it is desirable to have a test for a common deterministic trend against the alternative of individual specific time trends.

As pointed out by a referee, there are other detrending methods that may be used to construct an improved test procedure. A natural candidate is the “quasi difference” detrending suggested by Elliot, Rothenberg and Stock (1996) (see also Phillips and Xiao 1998). Unfortunately, it can be shown that a t -statistic computed from quasi differenced data also suffers from a (Nickell type) bias so that again a bias correction is required to obtain a reasonable test procedure. Nevertheless, a test procedure based on quasi differences may perform better than test procedures with OLS detrending. In this paper, our strategy is to avoid the bias term altogether. The comparison of our approach to a test procedure based on quasi differences is left for future research.

Acknowledgement

The research for this paper was carried out within Sonderforschungsbereich 373 at the Humboldt University Berlin and the METEOR research project “Dynamic and Nonstationary Panels: Theoretical and Empirical Issues”. I thank Carsten Trenkler and two referees for their helpful comments and suggestions.

Appendix

Proof of Theorem 1:

For convenience we will assume that σ_i^2 is known and without effect on the main result we set $\sigma_i^2 = 1$. Let $z_{it} = [\Delta x_{i,t-1}, \dots, \Delta x_{i,t-p}]'$ and a superscript “+” indicates that the respective series is adjusted for a common time trend. That is, \tilde{w}_{it}^+ is the residual from a pooled regression of \tilde{w}_{it} on a time trend. Then,

$$\tilde{w}_{i,t-1}^+ \tilde{u}_{it}^+ = w_{i,t-1} u_{it}^+ = (x_{i,t-1} - \tilde{c}'_i z_{it})(\varepsilon_{it} - \tilde{b}t),$$

where \tilde{b} and \tilde{c}_i denote the least-square estimates from the regressions

$$\begin{aligned} \varepsilon_{it} &= bt + \nu_{it} \\ x_{i,t-1} &= c'_i z_{it} + \zeta_{it}. \end{aligned}$$

For these estimates we have

$$\begin{aligned} \tilde{b} &= O_p(T^{-3/2}N^{-1/2}) \\ \tilde{c}_i &= O_p(1). \end{aligned}$$

It follows that

$$\sum_{i=1}^N \sum_{t=1}^T (x_{i,t-1} - \tilde{c}'_i z_{it}) \tilde{b}t = O_p(T).$$

Furthermore, since z_{it} is uncorrelated with ε_{it} we have

$$\sum_{i=1}^N \sum_{t=1}^T \tilde{c}'_i z_{it} (\varepsilon_{it} - \tilde{b}t) = O_p(T^{1/2}N^{1/2})$$

so that

$$\sum_{i=1}^N \sum_{t=1}^T (\tilde{w}_{i,t-1}^+ - \tilde{w}_{i0}^+) \tilde{u}_{it}^+ = \sum_{i=1}^N \sum_{t=1}^T x_{i,t-1} \varepsilon_{it} + O_p(T) + O_p(T^{1/2}N^{1/2})$$

Furthermore,

$$\sum_{i=1}^N \sum_{t=1}^T (w_{i,t-1}^+ - w_{i0}^+)^2 = O_p(T^2 N)$$

so that

$$\lambda_{BM} = \frac{\sum_{i=1}^N \sum_{t=1}^T (w_{i,t-1}^+ - w_{i0}^+) u_{it}^+}{\sqrt{\sum_{i=1}^N \sum_{t=1}^T (w_{i,t-1}^+ - w_{i0}^+)^2}} = z + O_p(T^{-1/2}) + O_p(N^{-1/2}),$$

where z is a standard normally distributed random variable.

Proof of Lemma 1:

From Phillips (1987) it follows that for a local alternative with $e^{-c/T\sqrt{N}} \approx 1 - c/(T\sqrt{N})$ the limiting process is given by

$$\begin{aligned} J_i^c(r) &= W_i(r) + (c/\sqrt{N}) \int_0^r e^{r-s} c W_i(s) ds && (\text{as } T \rightarrow \infty) \\ &= W_i(r) && (\text{as } N \rightarrow \infty). \end{aligned}$$

Hence, the limiting process is the same as for the case $c = 0$.

Proof of Lemma 2:

(i) The expression $E \int W^2$ is equal to the limit of

$$\begin{aligned} T^{-2} \sum_{t=1}^T E \left(\sum_{j=1}^t \varepsilon_j^* \right)^2 &= T^{-2} \sum_{t=1}^T t \\ &= 1/2 + o(1), \end{aligned}$$

where ε_t^* is a standard normally distributed white noise process.

(ii) The finite sample analog of $\int W dW$ is

$$T^{-1} \sum_{t=1}^T (\varepsilon_1^* + \cdots + \varepsilon_{t-1}^*) \varepsilon_t^* \tag{A.1}$$

and for $\int W^2$ the finite sample analog is

$$T^{-2} \sum_{t=1}^T (\varepsilon_1^* + \cdots + \varepsilon_t^*)(\varepsilon_1^* + \cdots + \varepsilon_t^*) . \quad (\text{A.2})$$

In the latter expression the cross products $(\varepsilon_1^* \varepsilon_t^*), \dots, (\varepsilon_{t-1}^* \varepsilon_t^*)$ enter with a factor 2. These cross products also occur in (A.1). All other terms in (A.2) are uncorrelated with (A.1). It follows that

$$\begin{aligned} & E \left\{ T^{-3} \left[\sum_{t=2}^T \left(\sum_{j=1}^{t-1} \varepsilon_j^* \right) \varepsilon_t^* \right] \left[\sum_{t=1}^T \left(\sum_{j=1}^t \varepsilon_j^* \right) \left(\sum_{j=1}^t \varepsilon_j^* \right) \right] \right\} \\ &= 2T^{-3} \sum_{t=2}^T \sum_{j=1}^t j + o(1) \\ &= 2T^{-3} \sum_{t=2}^T t^2/2 + o(1) \\ &= 1/3 + o(1) . \end{aligned}$$

(iii) Let $S_t = \sum_{j=1}^t \varepsilon_j^*$. Then, the expression $E(\int W^2)^2$ is equal to the limit

$$\lim_{T \rightarrow \infty} E \left\{ T^{-4} \left[\sum_{t=1}^T S_t^2 \right]^2 \right\} .$$

For $t \leq r$ we have for $k \geq 0$:

$$\begin{aligned} E(S_t^2 S_{t+k}^2) &= E \left[\sum_{i=1}^t \sum_{j=1}^t \varepsilon_i^* \varepsilon_j^* \right] \left[\sum_{i=1}^{t+k} \sum_{j=1}^{t+k} \varepsilon_i^* \varepsilon_j^* \right] \\ &= 2t^2 + t(t+k) \end{aligned}$$

where the first term on the right hand side results from the products $(\varepsilon_i^* \varepsilon_j^*)^2$ and $(\varepsilon_i^* \varepsilon_j^*)(\varepsilon_j^* \varepsilon_i^*)^2$ for $i, j = 1, \dots, t$ and the second term is due to the products $(\varepsilon_i^*)^2 (\varepsilon_j^*)^2$ for $i = 1, \dots, t$ and $j = 1, \dots, t+k$.

From

$$\left[\sum_{t=1}^T S_t^2 \right]^2 = \sum_{t=1}^T (S_t^2)^2 + 2 \sum_{t=1}^{T-1} S_t^2 S_{t+1}^2 + 2 \sum_{t=1}^{T-2} S_t^2 S_{t+2}^2 + \dots$$

it follows that

$$\begin{aligned} T^{-4} E \left[\sum_{t=1}^T S_t^2 \right]^2 &= 2 \int_0^1 \int_0^{1-s} 3r^2 + rs + o(1) \\ &= (1/2) + (1/12) + o(1) = 7/12 + o(1) . \end{aligned}$$

(iv) We have

$$\int V^2 = \int W^2 - \left(\int W \right)^2 - 12 \left[\int (r - 1/2) W \right]^2 .$$

First consider for $k \geq 0$

$$E \left[S_t \sum_{j=1}^T S_j \right] = tT - \sum_{j=1}^{t-1} j$$

and, thus, $E(\int W)^2$ is identical to the limit

$$\begin{aligned} \lim_{T \rightarrow \infty} E \left[T^{-2} \left(\sum_{t=1}^T S_t^2 \right)^2 \right] &= \int_0^1 r dr - \int_0^1 \int_0^s r dr ds \\ &= 1/3 . \end{aligned}$$

Similarly, it can be shown that

$$\begin{aligned} \lim_{T \rightarrow \infty} 12 E \left[T^{-5} \left(\sum_{t=1}^T (t - T/2) S_t^2 \right)^2 \right] \\ &= 12 \int_0^1 (r - 1/2) r dr - 12 \int_0^1 (s - 1/2) \int_0^s (r - 1/2)(r - s) dr ds \\ &= 1/10 . \end{aligned}$$

With these results we get

$$\int V^2 = \frac{1}{2} - \frac{1}{3} - \frac{1}{10} = \frac{1}{15}.$$

(v) From the finite sample analog of $W(1) \int W$ we conclude

$$\begin{aligned} T^{-2} E \left\{ \left[\sum_{t=1}^T \varepsilon_t^* \right] \left[\sum_{t=1}^T \sum_{j=1}^t \varepsilon_j^* \right] \right\} &= T^{-2} \sum_{t=1}^T \sum_{j=t}^T j \\ &= T(T+1)/2T^2 \\ &= 1/2 + o(1) \end{aligned}$$

and, thus, $E[W(1) \int W] = 1/2$.

Proof of Theorem 2:

(1) We first consider the case $\sigma_\eta^2 = 0$. Without affecting the asymptotic results, we assume that σ_i^2 is known. However, since the result depends on the individual specific variances, we do not set σ_i^2 to unity. Furthermore, under the local alternative we have $\bar{\sigma}_i/\sigma_i \rightarrow 1$.

Using the fact that the estimate of β is super-consistent with $\tilde{\beta} - \beta = O_p(T^{-3/2}N^{-1/2})$ we have under the sequence of local alternatives given in (13) – (14)

$$\begin{aligned} T^{-1} \sum_{t=1}^T (\tilde{w}_{i,t-1} - \tilde{w}_{i0}) \tilde{u}_{it} &= T^{-1} \sum_{t=1}^T \sigma_i^{-2} [x_{i,t-1} - x_{i0}] [\varepsilon_{it} - c T^{-1} N^{-1/2} x_{i,t-1}] + o_p(1) \\ &= T^{-1} \sum_{t=1}^T \sigma_i^{-2} x_{i,t-1} \varepsilon_{it} - c \sigma_i^{-2} N^{-1/2} T^{-2} \sum_{t=1}^T x_{i,t-1}^2 + o_p(1) \end{aligned}$$

so that by using $T^{-1/2} x_{i,[rT]} \Rightarrow \sigma_i W(r)$ we get

$$T^{-1} \sum_{t=1}^T (\tilde{w}_{i,t-1} - \tilde{w}_{i0}) \tilde{u}_{it} \Rightarrow \int W dW - (c/\sqrt{N}) \int W^2.$$

Using $E(\int W dW) = 0$ we obtain

$$\lim_{N,T \rightarrow \infty} E \left[N^{-1/2} T^{-1} \sum_{i=1}^N \sum_{t=1}^T (\tilde{w}_{i,t-1} - \tilde{w}_{i0}) \tilde{u}_{it} \right] = -c E \left[\int W^2 \right].$$

Furthermore, we have

$$\lim_{N,T \rightarrow \infty} E \left[N^{-1} T^{-2} \sum_{i=1}^N \sum_{t=1}^T (\tilde{w}_{i,t-1} - \tilde{w}_{i0})^2 \right] = E \left[\int W^2 \right]$$

so that by using Lemma 2 (i)

$$\lim_{N,T \rightarrow \infty} E[\lambda_{BM}] = -c \sqrt{E \left[\int W^2 \right]} = -c/\sqrt{2}.$$

For the variance we obtain

$$\text{var}[\lambda_{BM}] = 1 - \frac{2N^{-1/2} c E(\int W dW \int W^2) - N^{-1} c^2 [E(\int W^2)^2 - (E \int W^2)^2]}{E \int W^2} + o(1),$$

where we make use of the fact that $E \int W dW = 0$. It follows that $\lim_{N,T \rightarrow \infty} \text{var}(\lambda_{BM}) = 1$.

For $\sigma_\eta^2 \neq 0$ we have

$$\begin{aligned} T^{-2} \sum_{t=1}^T (\tilde{w}_{i,t-1} - \tilde{w}_{i0}) \tilde{u}_{it} &= T^{-2} \sum_{t=1}^T \sigma_i^{-2} [x_{i,t-1} - x_{i0} + \eta_i(t-1)] \\ &\quad \times \left[\varepsilon_{it} \sum_{t=1}^T -(c/T\sqrt{N}) x_{i,t-1} + \eta_i \right] + o_p(1) \\ &= \sigma_i^{-2} \eta_i^2 (T-1)T/2T^2 + o_p(1) = \sigma_i^{-2} \eta_i^2 / 2 + o_p(1) \end{aligned}$$

and

$$N^{-1} T^{-2} \sum_{i=1}^N \sum_{t=1}^T (\tilde{w}_{i,t-1} - \tilde{w}_{i0}) \tilde{u}_{it} \Rightarrow \psi_0 \sigma_\eta^2 / 2,$$

where $\psi_0 = N^{-1} \sum_{i=1}^N \sigma_i^{-2}$. For the denominator we obtain

$$\begin{aligned} N^{-1} T^{-3} \sum_{i=1}^N \sum_{t=1}^T (\tilde{w}_{i,t-1} - \tilde{w}_{i0})^2 &= T^{-3} \sum_{t=1}^T \sigma_i^{-2} \sigma_\eta^2 (t-1)^2 + o_p(1) \\ &= \psi_0 \sigma_\eta^2 / 3 + o_p(1) \end{aligned}$$

and, thus

$$\frac{\sum_{t=1}^N \sum_{t=1}^T (\tilde{w}_{i,t-1} - \tilde{w}_{i0}) \tilde{u}_{it}}{\sqrt{\sum_{i=1}^N \sum_{t=1}^T (\tilde{w}_{i,t-1} - \tilde{w}_{i0})^2}} = \sqrt{3\psi_0 \sigma_\eta^2 T / 2} + o_p(T^{1/2}).$$

It follows that λ_{BM} tends to infinity with \sqrt{N} .

Proof of Lemma 3:

Under the local alternative we have

$$\Delta x_{it} = -\frac{c}{T\sqrt{N}} x_{i,t-1} + \varepsilon_{it}$$

so that the k 'th autocovariance of Δx_t results as

$$\begin{aligned} \gamma_{ki} &= \lim_{T \rightarrow \infty} E \left[T^{-1} \sum_{t=k+1}^T \Delta x_{it} \Delta x_{i,t-k} \right] \\ &= \lim_{T \rightarrow \infty} E \left[T^{-1} \sum_{t=k+1}^T \left(-\frac{c}{T\sqrt{N}} x_{i,t-1} + \varepsilon_{it} \right) \left(-\frac{c}{T\sqrt{N}} x_{i,t-k-1} + \varepsilon_{i,t-k} \right) \right]. \end{aligned}$$

Since the variance and covariance is an monotonic increasing function of the autoregressive parameter we have

$$\begin{aligned} E \left[T^{-1} \sum_{t=k+1}^T \frac{c^2}{T^2 N} x_{i,t-1} x_{i,t-k-1} \right] &\leq T^{-1} \sum_{t=k+1}^T \frac{c^2 \sigma_i^2 t}{T^2 N} \\ &< \frac{(T+1)^2 c^2 \sigma_i^2}{2T^3 N} \\ &\rightarrow 0 \quad \text{for } T \rightarrow \infty \text{ or } N \rightarrow \infty, \end{aligned}$$

$$\begin{aligned} E \left[T^{-1} \sum_{t=k+1}^T \frac{c^2}{T\sqrt{N}} x_{i,t-1} \varepsilon_{i,t-k} \right] &< \frac{T c \sigma_i^2}{T^2 \sqrt{N}} \\ &\rightarrow 0 \quad \text{for } T \rightarrow \infty \text{ or } N \rightarrow \infty \end{aligned}$$

and

$$E \left[T^{-1} \sum_{t=k+1}^T \varepsilon_{it} \varepsilon_{i,t-k} \right] \longrightarrow \begin{cases} \sigma_i^2 & \text{for } k = 0 \\ 0 & \text{for } k \neq 0 \end{cases}$$

for $T \rightarrow \infty$. Collecting all results we obtain

$$\bar{\sigma}_{ki}^2 = \gamma_{0i} + 2 \sum_{k=1}^{\infty} \gamma_{ki} = \sigma_i^2 .$$

Proof of Lemma 4:

First, assume that $c = 0$ and, thus, the null hypothesis is true. From the definition of $V(r)$ we have

$$\lim_{T \rightarrow \infty} E(\Gamma_{Ti}) = -E \left[W(1) \int W \right] - E \left\{ 12 \left[\int (r - 1/2) W \right] \left[\int (r - 1/2) dW \right] \right\} .$$

Using Lemma 2 (v), the first term on the right hand side is equal to 1/2. For the second term we obtain from the finite sample analog

$$\begin{aligned} & 12T^{-4} E \left\{ \left[\sum_{t=1}^T (t - T/2) \left(\sum_{j=1}^t \varepsilon_t^* \right) \right] \left[\sum_{t=1}^T (t - T/2) \varepsilon_t^* \right] \right\} \\ &= 12T^{-4} \sum_{t=1}^T (t - T/2) \sum_{j=t}^T (j - T/2) + o(1) \\ &= \int_0^1 (s - 1/2) \int_s^1 (r - 1/2) dr ds + o(1) \\ &= o(1) , \end{aligned}$$

where $\varepsilon_t^* \sim i.i.\mathcal{N}(0, 1)$.

To analyze this expression under the local alternative, let $\alpha = 1 - c/(T\sqrt{N})$. Then, we obtain for the above expressions result as

$$\begin{aligned} & T^{-2} E \left\{ \left[\sum_{t=1}^T \sum_{j=0}^{t-1} \alpha^j \varepsilon_{t-j}^* \right] \left[\sum_{t=1}^T \varepsilon_t^* \right] \right\} \\ &= T^{-2} \sum_{t=1}^T \sum_{j=t}^T \alpha^{j-t} \end{aligned}$$

and

$$\begin{aligned} & 12T^{-4} E \left\{ \left[\sum_{t=1}^T (t - T/2) \left(\sum_{j=0}^{t-1} \alpha^j \varepsilon_{t-j}^* \right) \right] \left[\sum_{t=1}^T (t - T/2) \varepsilon_t^* \right] \right\} \\ &= -12T^{-4} \sum_{t=1}^T (t - T/2) \sum_{j=t}^T (j - T/2) \alpha^{j-t} . \end{aligned}$$

A Taylor approximation of α^{j-t} around unity gives

$$\alpha^{j-t} = 1 - (j - t)c/(T\sqrt{N}) + o(N^{-1/2}) .$$

We therefore have

$$\begin{aligned} T^{-2} \sum_{t=1}^T \sum_{j=t}^T \alpha^{j-t} &= (1/2) - T \left[\int_0^1 \int_s^1 (r - s) dr ds \right] c/(T\sqrt{N}) + o(N^{-1/2}) \\ &= (1/2) - c/(6\sqrt{N}) + o(N^{-1/2}) \end{aligned}$$

and

$$\begin{aligned} & 12T^{-4} \sum_{t=1}^T (t - T/2) \sum_{j=t}^T (j - T/2) \alpha^{j-t} \\ &= -T \left[\int_0^1 (s - 1/2) \int_s^1 (r - 1/2)(r - s) dr ds \right] c/(T\sqrt{N}) + o(N^{-1/2}) \\ &= 0.1 c/\sqrt{N} + o(N^{-1/2}) . \end{aligned}$$

Collecting these results gives

$$\begin{aligned} E(\Gamma_{Ti}) &= -1/2 + \left[\frac{1}{6} - \frac{1}{10} \right] c/\sqrt{N} + o(N^{-1/2}) \\ &= -(1/2) + (1/15)c/\sqrt{N} + o(N^{-1/2}) . \end{aligned}$$

Proof of Theorem 3:

Using (15), $\bar{\sigma}_i = \sigma_i$ and Lemma 4 we obtain

$$\lambda_{BA}^*(c) = \frac{c/15}{b_\infty \sqrt{E \int V^2}} - \frac{c\sqrt{E \int V^2}}{b_\infty}$$

$$\begin{aligned}
&= \frac{c\sqrt{1/15}}{b_\infty} - \frac{c\sqrt{1/15}}{b_\infty} \\
&= 0.
\end{aligned}$$

For the variance we have

$$\text{var}(\lambda_{BA}) = \frac{\text{var}(\int V dV) - 2cN^{-1/2}\text{cov}(\int V dV, \int V^2) + c^2N^{-1}\text{var}(\int V^2)}{b_\infty^2 E(\int V^2)} + o(1)$$

so that $\lim_{T, N \rightarrow \infty} \text{var}(\lambda_{BA}) = 1$.

Proof of Theorem 4:

Under the local alternative the expectation of Γ_{Ti}^* depends on $c/(T\sqrt{N})$. If we take the limit according to $T \rightarrow \infty$, the limit of Γ_{Ti}^* still depends on c/\sqrt{N} . Assuming that $\lim_{T \rightarrow \infty} E(\Gamma_{Ti}^*)$ is a differentiable function with respect to c/\sqrt{N} , it admits the expansion

$$\lim_{T \rightarrow \infty} E(\Gamma_{Ti}^*) = m_\infty + \lim_{T \rightarrow \infty} \left. \frac{\partial E(\Gamma_{Ti}^*)}{\partial(c/\sqrt{N})} \right|_{c=0} \frac{c}{\sqrt{N}} + o(N^{-1/2}).$$

We therefore have under the local alternative

$$\lambda_{IPS}^*(c) = \lim_{T \rightarrow \infty} E(\lambda_{IPS}) = \frac{c}{\omega_\infty} \left[\lim_{T \rightarrow \infty} \left. \frac{\partial E(\Gamma_{Ti}^*)}{\partial(c/\sqrt{N})} \right|_{c=0} - E\left(\sqrt{\int V^2}\right) \right].$$

Furthermore:

$$\lim_{T \rightarrow \infty} \text{var}(\tau_i) = \omega_\infty^2 - N^{-1/2} 2c \text{cov}\left(\frac{\int V dV}{\sqrt{\int V^2}}, \sqrt{\int V^2}\right) + N^{-1} c^2 \text{var}\left(\sqrt{\int V^2}\right)$$

so that $\lim_{T, N \rightarrow \infty} \text{var}(\lambda_{IPS}) = 1$.

Proof of Theorem 5:

From assumption (18) and the law of iterated expectations we have

$$\begin{aligned}
E(\mathbf{y}_i^* \mathbf{x}_i^*) &= 0 \\
\text{var}(\mathbf{y}_i^* \mathbf{x}_i^*) &= \sigma_i^2 E(\mathbf{x}_i^* \mathbf{A}' \mathbf{A} \mathbf{x}_i^*).
\end{aligned}$$

For some fixed T the central limit theorem implies as $N \rightarrow \infty$

$$N^{-1/2} \sum_{i=1}^N \mathbf{y}_i^* \mathbf{x}_i^* \Rightarrow \mathcal{N} \left(0, N^{-1} \sum_{i=1}^N \sigma_i^2 E(\mathbf{x}_i^* \mathbf{A}' \mathbf{A} \mathbf{x}_i^*) \right).$$

Furthermore, as $T \rightarrow \infty$ we have

$$N^{-1/2} T^{-2} \sum_{i=1}^N \sum_{t=1}^T \mathbf{x}_i^* \mathbf{A}' \mathbf{A} \mathbf{x}_i^* = \sum_{i=1}^N \sigma_i^2 E(T^{-2} \mathbf{x}_i^* \mathbf{A}' \mathbf{A} \mathbf{x}_i^*)$$

and, thus, the t -statistic

$$\lambda_{BA}^* = \frac{\sum_{i=1}^N \sigma_i^{-2} \mathbf{y}_i^* \mathbf{x}_i^*}{\sqrt{\sum_{i=1}^N \sigma_i^{-2} \mathbf{x}_i^* \mathbf{A}' \mathbf{A} \mathbf{x}_i^*}}$$

is asymptotically standard normal as $(N, T \rightarrow \infty)_{seq}$.

Proof of Theorem 6:

Under the local alternative we have

$$\begin{aligned} E(x_i' \mathbf{B}' \mathbf{A} y_i) &= E[\text{tr}(\mathbf{B}' \mathbf{A} y_i x_i')] \\ &= -\frac{c}{T\sqrt{N}} \text{tr}[\mathbf{B}' \mathbf{A} E(x_i x_i')] \\ &= -\frac{c \sigma_i}{T\sqrt{N}} \text{tr}[\mathbf{B}' \mathbf{A} \mathbf{H}], \end{aligned}$$

where $\mathbf{H} = \min(j, k)_{j,k=1,\dots,T-1}$. It is easy to verify that since $\mathbf{B}' \mathbf{A}$ is a upper triangular matrix we have $\mathbf{B}' \mathbf{A} \mathbf{H} = 0$ and, therefore, the expectation is zero for any T and N .

Proof of Theorem 7

First consider the expectation of

$$T^{-2} \sum_{t=1}^{T-1} x_{it}^* = T^{-2} \sum_{t=1}^{T-1} [y_{i,t-1}^2 - 2(t-1)y_{i,t-1}y_{iT}/T + (t-1)^2 y_{iT}^2/T^2]$$

$$= \int W^2 - 2W(1) \int rW + W(1)^2 \int r^2 dr + o_p(1) .$$

From Lemma 2 (i) it follows that $\int W^2 = 1/2$. From the finite sample analog of $W(1) \int rW$ we obtain

$$\begin{aligned} E \left[\sum_{t=1}^T \varepsilon_t^* \right] \left[\sum_{t=1}^T t \left(\sum_{j=1}^t \varepsilon_j^* \right) \right] &= T^{-2} \sum_{t=1}^T t^2 \\ &= 1/3 + o(1) \end{aligned}$$

so that $E[W(1) \int rW] = 1/3$.

Furthermore we have $E[W(1)^2 \int r^2 dr] = 1/3$ so that

$$\lim_{T \rightarrow \infty} E \left[T^{-2} \sum_{t=1}^{T-1} x_{it}^{*2} \right] = (1/2) - (2/3) + (1/3) = 1/6 .$$

Assuming that $\lim_{T \rightarrow \infty} E(\xi_{Ti})$ is differentiable with respect to (c/\sqrt{N}) , we obtain from the Taylor expansion

$$\lim_{T \rightarrow \infty} E(\xi_{Ti}) = \lim_{T \rightarrow \infty} \left. \frac{\partial E(\xi_{Ti})}{\partial (c/\sqrt{N})} \right|_{c=0} \frac{c}{\sqrt{N}} + o(N^{-1/2}) .$$

It follows that

$$\lambda_{UB}^*(c) = \lim_{T \rightarrow \infty} E(\lambda_{UB}) = c \sqrt{6} .$$

As for the case of the LL and IPS statistics it is straightforward to show that

$$\lim_{T, N \rightarrow \infty} \text{var}(\lambda_{UB}) = 1 .$$

References

- Arellano M., Bover O.** (1995) Another Look at the Instrumental-Variable Estimation of Error-Components Models, *Journal of Econometrics*, 68, 29–51.
- Baltagi, B.H** (1995), *Econometric Analysis of Panel Data*, Wiley, Chichester.
- Breitung, J.** (1992), *Dynamische Modelle für die Paneldatenanalyse*, PhD dissertation (in German), Haag + Herchen, Frankfurt.
- Breitung, J.** (1997), Testing for Unit Roots in Panel Data using a GMM approach, *Statistical Papers*, 38, 253–269.
- Breitung, J. and Meyer, W.** (1994), Testing for Unit Roots in Panel Data: Are wages on different bargaining levels cointegrated?, *Applied Economics*, 26, 353–361.
- Dickey, D.A. and Fuller, W.A.** (1979), Distribution of the Estimates for Autoregressive Time Series With a Unit Root, *Journal of the American Statistical Association*, 74, 427–431.
- Elliot, G., Rothenberg, T.J. and Stock, J.H.** (1996), Efficient Tests for an Autoregressive Unit Root, *Econometrica*, 813–836.
- Hsiao, C.** (1986), *Analysis of Panel Data*, Cambridge University Press, Cambridge.
- Im, K.S., Pesaran, M.H., Shin, Y.** (1997), Testing for Unit Roots in Heterogenous Panels, University of Cambridge, DAE Working Paper No 9526, revised version.

- Kao, C.** (1999), Spurious Regression and Residual-based Tests for Cointegration in Panel Data, *Journal of Econometrics*, 90, 1–44.
- Levin, A. and Lin, C.-F.** (1993), Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties, University of California San Diego, Unpublished Working Paper.
- Moon, H.R. and Phillips, P.C.B.** (1999), Estimation of Autoregressive Roots near Unity using Panel Data, Yale University, mimeo.
- Nickell, S.** (1981), Biases in Dynamic Models with Fixed Effects, *Econometrica*, 49, 1417–1426.
- Phillips, P.C.B.** (1987), Towards a Unified Asymptotic Theory of Autoregression, *Biometrika*, 74: 535–48.
- Phillips, P.C.B. and Lee, C.C.** (1996), Efficiency Gains from Quasi-Differencing under Nonstationarity, in: Robinson, P.M. and Rosenblatt, M. (eds.) *Essays in Memory of E.J. Hannan*, 300–314.
- Phillips, P.C.B. and Moon, H.R.** (1999), Linear Regression Limit Theory for Nonstationary Panel Data, forthcoming in: *Econometrica*.
- Phillips, P.C.B. and Ouliaris, S.** (1990), Asymptotic Properties of Residual Based Tests for Cointegration, *Econometrica*, 58, 165–193.
- Phillips, P.C.B. and Perron, P.** (1988), Testing for a Unit Root in Time Series Regression, *Biometrika*, 75, 335–346.
- Phillips, P.C.B. and Solo, V.** (1992), Asymptotics for Linear Processes, *Annals of Statistics* 20, 971–1001.

Phillips, P.C.B. and Xiao, Z. (1998), A Primer on Unit Root Testing, Journal of Economic Surveys, 12, 423–467.

Schmidt, P. and Phillips, P.C.B. (1992), LM Test for a Unit Root in the Presence of Deterministic Trends, Oxford Bulletin of Economics and Statistics 54, 257–287.