EVERYTHING YOU ALWAYS WANTED TO KNOW ABOUT PANEL UNIT ROOT TESTING*

Joakim Westerlund†
University of Gothenburg
Sweden

Jörg Breitung
University of Bonn
Germany

March 22, 2010

Abstract

This paper points to some of the facts that have emerged from 20 years of research into the analysis of unit roots in panel data. Some of these are known, others are not. But they all have in common that if ignored the effects can be very serious. This is demonstrated using both simulations and theoretical reasoning.

JEL classification: C13; C33.

Keywords: Non-stationary panel data; Unit root tests; Cross-section dependence; Multi-dimensional limits.

1 Introduction

Starting with the working paper versions of Quah (1994) and Breitung and Meyer (1994) that were available already in 1989, the literature concerned with the analysis of unit roots in panel data covers more than 20 years. While during the first decade the topic was rather peripheral, it has by now become a very active research area, see Breitung and Pesaran (2008) for a recent survey of the literature. Today panel unit root tests are standard econometric tools within most fields of empirical economics, especially in macroeconomics and financial

*Previous versions of this paper were presented at the 5th Nordic Econometric Meeting in Lund and seminars in Amsterdam, Athens, Gothenburg, Maastricht, Paris and Thessaloniki. The authors would like to thank conference and seminar participants, and in particular Uwe Hassler, Jean-Pierre Urbain, Franz Palm and four anonymous referees for helpful comments and suggestions. Thank you also to the Jan Wallander and Tom Hedelius Foundation for financial support under research grant number W2006–0068:1.

†Corresponding author: Department of Economics, University of Gothenburg, P. O. Box 640, SE-405 30 Gothenburg, Sweden. Telephone: +46 31 786 5251, Fax: +46 31 786 1043, E-mail address: joakim.westerlund@economics.gu.se.

economics, and some are now available in commercial software packages such as EViews and STATA.

In the beginning when the panel unit root literature was still in its infancy econometricians tended to view the extension of the conventional unit root analysis to panel data as a rather straightforward and less exciting exercise. However, it has since then become clear that this is not the case. Indeed, subsequent work has revealed a number of fundamental differences in the way statistical inference with non-stationary data is performed and it seems fair to say that panel unit root analysis has established itself as a legitimate branch of literature.

Two of the most influential contributions to this field of research are those of Levin *et al.* (2002) and Im *et al.* (2003), henceforth LLC and IPS, respectively, who were among the first to develop so-called first-generation tests that are appropriate when the cross-sectional units are independent of each other. These works have not only been important for applied researchers, to which vast amounts of panel data are now available, but also for econometricians concerned with the development of new tests and methods for non-stationary panels. In fact, among the large number of studies that has recently been published within this field there are only a few, if any, that do not build, either directly or indirectly, on the works of LLC and IPS.

To the best of our knowledge, however, there is still no study that investigates basic principles, or facts, upon which these tests, and therefore also the whole panel unit root literature, are built. There are of course literature reviews such as those of Baltagi and Kao (2000), Banerjee (1999), Choi (2006), Bai and Kao (2006), and Breitung and Pesaran (2008), but the purpose of these is more to review, and not so much to study in detail the underlying facts. Consequently, they do not either emphasize enough the consequences of overlooking these facts. This paper goes back to the basics and shows how they make a difference. It is an in-dept review and theoretical companion to specific research papers.

In general in econometrics and statistics most methods are based on certain established facts that are assumed to be well-know. For example, time series unit root tests, like the one suggested by Dickey and Fuller (1979), henceforth denoted DF, can suffer from poor power if the autoregressive root is local-to-unity, especially in the presence of deterministic constant and trend terms. It also a well-known fact that one way to improve the power of such tests is to perform the detrending regression in a way that is efficient under the local alternative hypothesis, an idea that was first suggested by Elliott *et al.* (1996).

Panel unit root analysis also has its fair share of facts like this. But in contrast to the conventional time series analysis, which has been around for almost half a decade now, these are not very well-understood, and some are just being developed. The purpose of this paper is in part to point to these facts, in part to illustrate, both analytically and numerically, their importance, and in so doing we will make use of the LLC and IPS tests as leading examples. Some of the facts are known, others are not. But they all share the feature that if ignored the effects upon analysis can be dramatic. Proper understanding of these facts is therefore key in any research with non-stationary panel data.

To place our results in the context of the literature on dynamic panel data models in general, note that in this paper we are only concerned with those aspects that are in some sense unique to the unit root testing problem. The facts we analyze can be divided roughly into the following three broad categories: (i) asymptotic properties under the unit root null and stationary alternative hypotheses in absence of nuisance parameters, (ii) the treatment of deterministic components, and (iii) the effect of initial condition, serial correlation and cross-section dependence. These categories have of course been treated before but for different reasons than they are here.

The plan of the paper is the following. Section 2 focuses on the simplest case without any deterministic terms, short-run dynamics or cross-sectional dependence. Although admittedly very restrictive, this setup allows us to focus on some of the most primitive facts of panel unit root testing. In Section 3 we generalize the setup of Section 2 to allow for deterministic constant and trend terms. The analysis reveal that this small change has major implications for the resulting tests. Models with short-run dynamics are considered in Section 4, and in Section 5 we address the problems that arise when the cross-sectional units are no longer independent. Section 6 offers some concluding remarks.

2 The simplest case

Consider the double indexed variable y_{it}, observable for $t = 1, ..., T$ time periods and $i = 1, ..., N$ cross-sectional units. Initially we will assume that y_{it} has no deterministic part, so that

$$y_{it} = y_{it}^{s}, \tag{1}$$

where y^s_{it} is the stochastic part of y_{it}, which is assumed to evolve according to the following first-order autoregressive (AR) process:

$$y^s_{it} = \rho_i y^s_{it-1} + \varepsilon_{it}, \tag{2}$$

or, equivalently,

$$\Delta y_{it} = (\rho_i - 1) y_{it-1} + \varepsilon_{it} = \alpha_i y_{it-1} + \varepsilon_{it}. \tag{3}$$

In this section we assume that the error ε_{it} is mean zero and independent across both i and t. To make life even simpler, we assume that the errors are homoscedastic so that $E(\varepsilon^2_{it}) = \sigma^2$ for all i and t. Note that while unduely restrictive for most practical purposes, this data generating process has the advantage of being simple and illustrative.

The null hypothesis of interest is

$$H_0 : \alpha_i = 0 \text{ for all } i,$$

which corresponds to a fully non-stationary panel. As for the alternative hypothesis, we will consider two candidates, H_{1a} and H_{1b}. The first is specified as

$$H_{1a} : \alpha_i = \alpha < 0 \text{ for all } i,$$

and corresponds to a fully stationary panel with the same degree of mean reversion for all units. It is therefore quite restrictive. The second alternative is more relaxed. It reads

$$H_{1b} : \alpha_i < 0 \text{ for } i = 1, ..., N_1 \text{ with } \frac{N_1}{N} \to \delta_1 > 0 \text{ as } N_1,\ N \to \infty,$$

which corresponds to a mixed panel with δ_1 being the limiting fraction of stationary units. Note that in this formulation, there are no homogeneity restrictions with regards to the degree of mean reversion. Note also that at this point we make no assumptions concerning the remaining $N - N_1$ slopes, $\alpha_{N_1+1}, ..., \alpha_N$, which may all be zero, negative or a mixture of both. However, we do require that $\delta_1 > 0$, as otherwise the panel would escape stationarity as N_1, $N \to \infty$.

The two alternative hypotheses H_{1a} and H_{1b} are chosen to match the LLC and IPS tests, respectively, which are going to be of primary interest in this paper.

Before considering these tests, however, it is useful to introduce some notation. In particular, we define $M = \sum_{i=1}^{N} M_i$, where

$$M_i = \begin{pmatrix} M_{11i} & M_{12i} \\ M_{12i} & M_{22i} \end{pmatrix} = \sum_{t=2}^{T} \begin{pmatrix} (\Delta y_{it})^2 & y_{it-1}\Delta y_{it} \\ y_{it-1}\Delta y_{it} & y^2_{it-1} \end{pmatrix}$$

is the non-normalized moment matrix of the variables contained in the regression in (3), whose asymptotic counterpart is given by

$$M_i^\circ = \begin{pmatrix} M_{11i}^\circ & M_{12i}^\circ \\ M_{12i}^\circ & M_{22i}^\circ \end{pmatrix} = \begin{pmatrix} 1 & \int_0^1 W_i(s)dW_i(s) \\ \int_0^1 W_i(s)dW_i(s) & \int_0^1 W_i(s)^2 ds \end{pmatrix},$$

where $W_i(s)$ is a standard Brownian motion on $s \in [0,1]$. In particular, it holds that

$$\begin{pmatrix} \frac{1}{T} M_{11i} & \frac{1}{T} M_{12i} \\ \frac{1}{T} M_{12i} & \frac{1}{T^2} M_{22i} \end{pmatrix} \Rightarrow \sigma^2 M_i^\circ$$

as $T \to \infty$, where the symbol $\Rightarrow$ signifies weak convergence.

The results reported in this paper are derived using either the joint limit method wherein $N, T \to \infty$ simultaneously, or the sequential limit method wherein one of the indices is passed to infinity before the other, see Phillips and Moon (1999). In any case, since the purpose here is more to illustrate rather than to prove, details that are not essential for the understanding of the main point will be omitted. The derivations will therefore not be complete, and readers are referred to the relevant original works for a more detailed treatment.

Having introduced the main notation, we now go on to discuss the IPS and LLC tests. With no serial correlation or heteroskedasticity, and no deterministic constant or trend terms, the Levin and Lin (1992) statistic is given by

$$\tau_{LLC} = \frac{M_{12}}{\hat{\sigma}\sqrt{M_{22}}} = \hat{\alpha}\,\frac{\sqrt{M_{22}}}{\hat{\sigma}},$$

where $\hat{\sigma}^2 = \frac{1}{NT}(M_{11} - \hat{\alpha}\, M_{12})$ with $\hat{\alpha} = M_{12}/M_{22}$ being the least squares estimator of α, whose standard error is given by $\hat{\sigma}/\sqrt{M_{22}}$. Note that although in this setting the Levin and Lin (1992) statistic is the same as the LLC statistic that assumes no deterministic component and no short-run dynamics, at times it will be important to keep the distinction, as this similarity is not always going to hold when we go on to discuss more general models.

The IPS test is given by

$$\tau_{IPS} = \frac{\sqrt{N}(\overline{\tau} - E(\tau))}{\sqrt{\text{var}(\tau)}},$$

where $\overline{\tau} = \frac{1}{N}\sum_{i=1}^N \tau_i$ and τ_i is the usual DF test statistic,

$$\tau_i = \frac{M_{12i}}{\hat{\sigma}_i\sqrt{M_{22i}}} = \hat{\alpha}_i\,\frac{\sqrt{M_{22i}}}{\hat{\sigma}_i}$$

with an obvious definition of $\hat{\sigma}_i^2$ and $\hat{\alpha}_i$. It is well-known that

$$\tau_i \Rightarrow \frac{M_{12i}^\circ}{\sqrt{M_{22i}^\circ}}$$

as $T \to \infty$. The constants $E(\tau)$ and $\text{var}(\tau)$ are simply the mean and variance of this limiting distribution. Note that since M_i° is identically distributed, $E(\tau)$ and $\text{var}(\tau)$ do not need to carry an i index.

Fact 1: The IPS and LLC statistics are standard normally distributed as $N \to \infty$ even if T is fixed

In order to establish the asymptotic normality of τ_{LLC} and τ_{IPS} as $N \to \infty$ with T fixed we invoke two of the most important tools of the analysis of non-stationary panel data, the weak law of large numbers and the Lindeberg–Levy central limit theorem.

Consider first the LLC statistic, which can be written as

$$\tau_{LLC} = \frac{\frac{1}{T\sqrt{N}} M_{12}}{\hat{\sigma}\sqrt{\frac{1}{NT^2} M_{22}}}.$$

We begin by analyzing the denominator under H_0, which by the law of large numbers as $N \to \infty$ becomes

$$\frac{1}{NT^2} M_{22} \xrightarrow{p} \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T^2} \sum_{t=2}^{T} E(y_{it-1}^2) = \sigma^2 \frac{1}{T^2} \sum_{t=1}^{T-1} t = \sigma^2 \frac{T-1}{2T},$$

where $\xrightarrow{p}$ signifies convergence in probability. Similarly, $\frac{1}{NT} M_{12} \xrightarrow{p} 0$ and $\frac{1}{NT} M_{11} \xrightarrow{p} \sigma^2$ as $N \to \infty$, from which we deduce that $\hat{\alpha} \xrightarrow{p} 0$ and $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$.

Moreover,

$$\text{var}\left(\frac{1}{T} M_{12i}\right) = \frac{1}{T^2} \sum_{t=2}^{T} \text{var}(y_{it-1}\Delta y_{it}) = \sigma^2 \frac{1}{T^2} E(M_{22i}) = \sigma^4 \frac{T-1}{2T}.$$

Note that for a fixed T, $\sigma^4(T-1)/2T < \infty$. In view of this and the assumed independence across i, we have that by the Lindeberg–Levy central limit theorem,

$$\frac{1}{T\sqrt{N}} M_{12} = \frac{1}{T\sqrt{N}} \sum_{i=1}^{N} M_{12i} \xrightarrow{d} \sigma^2 \sqrt{\frac{T-1}{2T}} \mathcal{N}(0,1)$$

as $N \to \infty$, where $\xrightarrow{d}$ denotes convergence in distribution. Thus, by putting everything together we get

$$\tau_{LLC} = \frac{\frac{1}{T\sqrt{N}} M_{12}}{\hat{\sigma}\sqrt{\frac{1}{T^2 N} M_{22}}} \xrightarrow{d} \mathcal{N}(0,1).$$

Note that this result holds for any T. Hence, the asymptotic normality of the LLC statistic does not require $T \to \infty$. However, if individual specific parameters relating to for example deterministic terms or short-run dynamics are introduced, then this is no longer true. The reason is that consistent estimation of these parameters requires $T \to \infty$, see for example Harris and Tzavalis (1999) and LLC.

In a similar manner it can be shown that τ_{IPS} is also standard normally distributed as $N \to \infty$ with T held fixed.[1] This is done in IPS. The idea behind the proof is that since τ_i is cross-sectionally independent by assumption, the application of the Lindeberg–Levy central limit theorem requires that the variance is finite. Although this is generally far from trivial to show, if we assume that ε_{it} is normally distributed, then $\tau_i^2 = (T-1)R^2$, where R^2 is the usual R^2 statistic of the regression in (3). Clearly, since $R^2 \in [0,1]$, the variance exists for finite T and therefore

$$\tau_{IPS} \xrightarrow{d} \mathcal{N}(0,1)$$

as $N \to \infty$ by the Lindeberg–Levy central limit theorem.[2] Thus, as long as $N \to \infty$ normality of these statistics does not require passing $T \to \infty$, a fact that is oftentimes not considered in theoretical work. As a practical matter, this means that the tests can be applied even in situations in which $N >> T$.

The performance under the stationary alternative is the topic of the next section.

Fact 2: The LLC test can be more powerful than the IPS test

It has become standard to treat τ_{LLC} as a test against H_{1a} and τ_{IPS} as a test against H_{1b}. Therefore, since H_{1b} is less restrictive than H_{1a}, one might be led to believe that τ_{IPS} should dominate τ_{LLC} in terms of power, at least under the heterogeneous alternative. However, this is not true.

Consider first the case when the slope coefficient α_i is fixed under the alternative. If H_{1a} holds, then we write

$$\frac{1}{\sqrt{NT}}\tau_{LLC} = \alpha\frac{\sqrt{\frac{1}{NT}M_{22}}}{\hat{\sigma}} + (\hat{\alpha}-\alpha)\frac{\sqrt{\frac{1}{NT}M_{22}}}{\hat{\sigma}} = O_p(1) + O_p\left(\frac{1}{\sqrt{NT}}\right)O_p(1),$$

which implies that $\tau_{LLC} = O_p(\sqrt{NT})$. Similarly, if H_{1b} holds, and assuming for simplicity

[1]Instead of the Lindeberg condition the Liaponov condition is usually invoked, in which case one has to assume that the individual test statistics τ_i have more than two moments.

[2]Strictly speaking, for this result to hold we require that $E(\tau)$ and $\mathrm{var}(\tau)$ are evaluated for a finite T.

that the last $N - N_1$ units are non-stationary,

$$\begin{aligned}
\sqrt{\text{var}(\tau)}\, \tau_{IPS} &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\tau_i - E(\tau)) \\
&= \sqrt{\frac{N_1}{N}} \frac{1}{\sqrt{N_1}} \sum_{i=1}^{N_1} (\tau_i - E(\tau)) + \sqrt{1 - \frac{N_1}{N}} \frac{1}{\sqrt{N - N_1}} \sum_{i=N_1+1}^{N} (\tau_i - E(\tau)) \\
&= \sqrt{\delta_1}\, O_p(\sqrt{NT}\,) + \sqrt{1 - \delta_1}\, O_p(1),
\end{aligned}$$

where we have used that under a fixed alternative

$$\frac{1}{\sqrt{T}} E(\tau_i) = \frac{\alpha_i}{\sigma} E\left(\sqrt{\frac{1}{T} M_{22i}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) \rightarrow \frac{\alpha_i}{\sqrt{1 - \rho_i^2}} \neq E(\tau)$$

as $T \to \infty$, implying

$$\frac{1}{N_1} \sum_{i=1}^{N_1} (\tau_i - E(\tau)) \xrightarrow{p} E(\tau_i) - E(\tau) = O_p(\sqrt{T}\,)$$

so that $\frac{1}{\sqrt{N_1}} \sum_{i=1}^{N_1} (\tau_i - E(\tau)) = O_p(\sqrt{N_1 T}\,)$, which is $O_p(\sqrt{NT}\,)$ provided that $\delta_1 > 0$. It follows that $\tau_{IPS} = O_p(\sqrt{NT}\,)$.

The rate of divergence is therefore the same for both tests. Note also that the rate of divergence of τ_{IPS} is independent of the value taken by δ_1, as long as $\delta_1 > 0$. The divergence rate of this test in a panel where for example only half of the units are stationary is therefore the same as that in a panel where all units are stationary.

In other words, under a fixed alternative hypothesis and the conditions considered here little can be said about the power of the LLC and IPS tests. A more detailed analysis requires more assumptions. This is done in Bowman (1999), who studies the power of some panel unit root tests in the absence of nuisance parameters. He characterizes the class of admissible panel unit root tests, and shows that while the LLC test is admissible, and in fact uniformly most powerful against the homogenous alternative H_{1b}, the IPS test is inadmissible. The LLC test is therefore expected to be more powerful than the IPS test.

Consider next the case when α_i is local-to-unity,

$$H_{1c} : \alpha_i = \frac{c_i}{T\sqrt{N}}, \tag{4}$$

where $c_i < 0$ is a constant such that $\frac{1}{N} \sum_{i=1}^{N} c_i \to \bar{c}$ as $N \to \infty$. This corresponds to an autoregressive coefficient that approaches one with increasing values of N and T. Since $c_i < 0$ this means that α_i approaches one from below and so y_{it} is locally stationary. Note also that since c_i is not restricted to be equal across i, the mean reversion is essentially heterogenous. Alternatives of this kind has been extensively used in the time series literature to study the power of unit root test statistics. The convergence rate of the local alternative to the null hypothesis is faster in our case because the use of panel data will entail faster convergence of the pooled least squares estimator $\hat{\alpha}_i$ under H_0.

Let us assume for simplicity that $y_{i0} = 0$. Then, by Taylor expansion,

$$\begin{aligned} \frac{1}{\sigma\sqrt{T}} y_{it} &= \frac{1}{\sigma\sqrt{T}} \sum_{j=0}^{t} \rho_i^j \varepsilon_{it-j} \simeq \frac{1}{\sigma\sqrt{T}} \sum_{j=0}^{t} \varepsilon_{it-j} + \frac{c_i}{\sigma\sqrt{NT}} \sum_{j=1}^{t} \frac{j}{T} \varepsilon_{it-j} \\ &\Rightarrow W_i(s) + \frac{c_i}{\sqrt{N}} U_i(s) \end{aligned}$$

as $T \to \infty$, where $U_i(s) = \int_0^s W_i(r)dr$. Thus, by subsequently passing $N \to \infty$,

$$\frac{1}{\sigma^2 T\sqrt{N}} \sum_{i=1}^{N} \sum_{t=2}^{T} y_{it-1}\varepsilon_{it} \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{2}\right) + \bar{c} \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} E\left(\int_0^1 U_i(s)dW_i(s)\right) \sim \mathcal{N}\left(0, \frac{1}{2}\right),$$

which uses the fact that $E\big(\int_0^1 U_i(s)dW_i(s)\big) = 0$. But we also have $\frac{1}{T^2 N} M_{22} \xrightarrow{p} \frac{\sigma^2}{2}$ as $N, T \to \infty$, and therefore

$$\tau_{LLC} = \frac{1}{\hat{\sigma}\sqrt{\frac{1}{NT^2} M_{22}}} \frac{1}{T\sqrt{N}} \sum_{i=1}^{N} \sum_{t=2}^{T} \left(\frac{c_i}{T\sqrt{N}} y_{it-1}^2 + y_{it-1}\varepsilon_{it} \right) \xrightarrow{d} \mathcal{N}\left(\frac{\bar{c}}{\sqrt{2}}, 1\right).$$

Thus, in spite of the heterogeneity of this alternative, we find that the LLC test has nontrivial local power. The basic reason for this is that the power only depends on $\bar{c}$, the mean of c_i, which means that it does not matter whether c_i is equal across i. It is also interesting to note that the denominator of the LLC statistic does not contribute to the local power of the test, which stands in sharp contrast to the DF statistic, whose local power depends on both the numerator and the denominator.

Let us now consider the local power of the IPS statistic. Using Taylor expansion and then inserting

$$\begin{aligned}
\frac{1}{\sigma^2 T}\sum_{t=2}^{T} y_{it-1}\Delta y_{it} &\Rightarrow \int_0^1 \left(W_i(r)+\frac{c_i}{\sqrt{N}}U_i(r)\right)\left(dW_i(r)+\frac{c_i}{\sqrt{N}}W_i(r)\right)dr \\
&= \int_0^1 W_i(r)dW_i(r)+\frac{c_i}{\sqrt{N}}\left(\int_0^1 U_i(r)dW_i(r)+\int_0^1 W_i(r)^2dr\right)+O_p\left(\frac{1}{N}\right) \\
&= M^\circ_{12i}+\frac{c_i}{\sqrt{N}}(R_{1i}+M^\circ_{22i})+O_p\left(\frac{1}{N}\right), \\
\frac{1}{\sigma^2 T^2}\sum_{t=2}^{T} y_{it-1}^2 &\Rightarrow \int_0^1 \left(W_i(r)+\frac{c_i}{\sqrt{N}}U_i(r)\right)^2 \\
&= \int_0^1 W_i(r)^2dr+\frac{2c_i}{\sqrt{N}}\int_0^1 W_i(r)U_i(r)dr+O_p\left(\frac{1}{N}\right) \\
&= M^\circ_{22i}+\frac{2c_i}{\sqrt{N}}R_{2i}+O_p\left(\frac{1}{N}\right),
\end{aligned}$$

we obtain

$$\tau_i \Rightarrow \frac{M^\circ_{12i}}{\sqrt{M^\circ_{22i}}}+\frac{c_i}{\sqrt{N}}\left(\sqrt{M^\circ_{22i}}+\frac{R_{1i}}{\sqrt{M^\circ_{22i}}}-\frac{M^\circ_{12i}R_{2i}}{(M^\circ_{22i})^{3/2}}\right)+O_p\left(\frac{1}{N}\right).$$

It follows that as N, $T \to \infty$,

$$\tau_{IPS} \xrightarrow{d} \mathcal{N}(0,1)+\frac{\bar{c}}{\sqrt{\mathrm{var}(\tau)}}E\left(\sqrt{M^\circ_{22i}}+\frac{R_{1i}}{\sqrt{M^\circ_{22i}}}-\frac{M^\circ_{12i}R_{2i}}{(M^\circ_{22i})^{3/2}}\right),$$

see Harris *et al.* (2010).

Using simulations where the Brownian motion $W_i(r)$ is approximated by a random walk of length $T = 1,000$ we find

$$E\left(\sqrt{M^\circ_{22i}}+\frac{R_{1i}}{\sqrt{M^\circ_{22i}}}-\frac{M^\circ_{12i}R_{2i}}{(M^\circ_{22i})^{3/2}}\right) = 0.6221 - 0.0794 + 0.0382 = 0.581.$$

Since $0.581/\sqrt{\mathrm{var}(\tau)} = 0.581/0.985 \cong 0.6 < 1/\sqrt{2} = 0.707$ it follows that the local power of the IPS test is always smaller than that of the LLC test, which corroborates the results of Bowman (1999).

To illustrate these findings a small simulation experiment was conducted using (1), (2) and (4) with $\varepsilon_{it} \sim \mathcal{N}(0,1)$ and $y_{i0} = 0$ to generate the data. Two specifications are considered. In the first, $c_i = c$ for all i, suggesting a completely homogenous AR parameter, while in the second, $c_i \sim U(2c, 0)$. Hence, $\mathrm{var}(c_i) = c^2/3 > 0$ whenever $c < 0$ and so the individual AR coefficients are no longer restricted to be equal. However, the mean is still c, just as in the first specification. The empirical rejection frequencies are based on 5,000 replications and the 5% critical value.[3] The results are summarized in Table 1. We see that in agreement with the theoretical results, τ_{LLC} is uniformly more powerful than τ_{IPS}. We also see that the actual power corresponds roughly to the asymptotic power, at least for large samples and small values of c.

These results suggest that if it is power that we are concerned about, then the LLC test is the one to use.

Table 1: Power against different local alternatives.

		$c_i = c$		$c_i \sim U(2c, 0)$	
T, N	c	LLC	IPS	LLC	IPS
20	-1	15.4	12.5	15.3	12.4
	-2	33.5	24.0	31.3	23.3
	-5	90.4	73.8	76.6	67.9
50	-1	16.6	13.1	16.2	12.8
	-2	36.6	26.8	33.9	26.1
	-5	93.7	80.4	85.0	75.9
100	-1	16.8	13.3	16.5	13.3
	-2	38.7	26.8	36.7	26.1
	-5	94.8	83.5	89.8	79.9
Asymptotic	-1	17.4	14.8	17.4	14.8
	-2	40.9	32.8	40.9	32.8
	-5	97.1	91.23	97.1	91.2

Notes: The table reports the 5% rejection frequencies when the AR parameter is set to $\alpha_i = c_i/T\sqrt{N}$.

3 Models with deterministic terms

Fact 3: Deterministic components need not be treated as in the DF approach

In the presence of deterministic constant and trend terms, LLC and IPS suggest following the DF proposal of using least squares demeaning. One might therefore think that this is also the simplest way to handle such terms. But this is not true.

[3] From now on all simulations will be conducted at the 5% level using 5,000 replications. Also, in order to reduce the effect of the initial condition, the first 100 observations of each cross-sectional unit will henceforth be disregarded.

Consider the model

$$y_{it} = \mu_i + y_{it}^s, \tag{5}$$

where the constant μ_i now represents the deterministic part of y_{it}, while y_{it}^s again represents the stochastic part. As usual, the allowance for deterministic terms of this kind makes it necessary to appropriately augment the regression in (3). Let us therefore introduce x_{it} to denote a generic vector containing all regressors other than y_{it-1} with γ_i being the associated vector of slope coefficients. In the current case with a constant this yields

$$\Delta y_{it} = \alpha_i y_{it-1} - \alpha_i \mu_i + \varepsilon_{it} = \alpha_i y_{it-1} + \gamma_i x_{it} + \varepsilon_{it}, \tag{6}$$

where $\gamma_i = -\alpha_i \mu_i$ and $x_{it} = 1$ for all i and t. The matrix of sample moments is augmented accordingly as

$$M_i = \begin{pmatrix} M_{11i} & M_{12i} & M_{13i} \\ M_{12i} & M_{22i} & M_{23i} \\ M_{13i}' & M_{23i}' & M_{33i} \end{pmatrix} = \sum_{t=2}^{T} \begin{pmatrix} (\Delta y_{it})^2 & y_{it-1}\Delta y_{it} & \Delta y_{it} x_{it}' \\ y_{it-1}\Delta y_{it} & y_{it-1}^2 & y_{it-1} x_{it}' \\ x_{it}\Delta y_{it} & x_{it} y_{it-1} & x_{it} x_{it}' \end{pmatrix}$$

with x_{it} ordered last. Moreover, since the focus here is on α_i and not on γ_i, the analysis will be carried out in two steps, where the first involves projecting Δy_{it} and y_{it-1} upon x_{it}. The second step is then to test for a unit root in the resulting projection errors, whose sum of squares can be written in terms of the partitions of M_i as

$$M_{abi}^p = M_{abi} - M_{a3i} M_{33i}^{-1} M_{3bi}.$$

The corresponding limiting sum of squared projection errors is defined as

$$M_{abi}^{\circ p} = M_{abi}^{\circ} - M_{a3i}^{\circ} (M_{33i}^{\circ})^{-1} M_{3bi}^{\circ}$$

with an obvious definition of M_{abi}°.

Also, except for M^p, to simplify the notation let us from now on suppress any dependence upon p. For example, we write $\hat{\sigma}^2 = \frac{1}{NT}(M_{11}^p - \hat{\alpha}\, M_{12}^p)$ and $\hat{\alpha} = M_{12}^p / M_{22}^p$, which are the same definitions as in Section 2 but with the elements of M^p in place of the corresponding elements of M.

Consider now the DF approach of using least squares demeaning, in which case

$$M_{abi}^p = M_{abi} - \frac{1}{T} M_{a3i} M_{3bi},$$

so that for example $M_{12i}^{p} = \sum_{t=2}^{T}(y_{it-1} - \overline{y}_{i,-1})\Delta y_{it}$, where $\overline{y}_{i,-1} = \frac{1}{T}\sum_{t=2}^{T} y_{it-1}$. The limiting version of this quantity is given by $M_{12i}^{\circ p} = \int_0^1 (W_i(s) - \overline{W}_i)dW_i(s)$, where $\overline{W}_i = \int_0^1 W_i(s)ds$. Thus, since $E(M_{12i}^{\circ p}) = -1/2$ under H_0, we have that in the sequential limit, where $T \to \infty$ is followed by $N \to \infty$

$$\frac{1}{TN} M_{12}^{p} \xrightarrow{p} \sigma^2 E(M_{12i}^{\circ p}) = -\frac{\sigma^2}{2}.$$

Since $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$ and $\frac{1}{T^2}E(M_{22i}^{p}) \to \frac{\sigma^2}{6}$ we have that as $N, T \to \infty$

$$\frac{1}{\sqrt{N}} \tau_{LLC} = \frac{\frac{1}{TN} M_{12}^{p}}{\hat{\sigma}\sqrt{\frac{1}{T^2 N} M_{22}^{p}}} \xrightarrow{p} -\frac{\sqrt{6}}{2}$$

and by using $T^{-2}\text{var}(M_{12i}^{p}) \to \sigma^4/12$:

$$\frac{\sqrt{12N}}{\sigma^2}\left(\frac{1}{TN} M_{12}^{p} + \frac{\sigma^2}{2}\right) \xrightarrow{d} \mathcal{N}(0,1).$$

It follows that

$$\tau_{LLC}^{c} = \frac{\sqrt{2N}\left(\frac{1}{TN} M_{12}^{p} + \frac{\hat{\sigma}^2}{2}\right)}{\hat{\sigma}\sqrt{\frac{1}{NT^2} M_{22}^{p}}} \xrightarrow{d} \mathcal{N}(0,1).$$

This is the bias-adjusted LLC statistic, which has been superscripted by c to indicate that it is robust to the presence of the constant in the model. The point here is that least squares demeaning is not enough to get rid off the effect of μ_i. There is also a bias that needs to be accounted for, which complicates the testing considerably. This is the so-called Nickell bias first analyzed in the context of a fixed T by Nickell (1981), which has been further analyzed by Hahn and Kuersteiner (2002) in cases when $\frac{N}{T}$ is assumed to converge to a constant.

As mentioned in Section 2, as soon as one moves away from the simplest case with no deterministic components and no short-run dynamics, the statistic proposed in Levin and Lin (1992) need not be the same as the one in LLC. In the current setting Levin and Lin (1992) suggest using

$$\tau_{LL}^{c} = \frac{\sqrt{5}}{2}\tau_{LLC} + \sqrt{\frac{15N}{8}},$$

which is even more complicated than τ_{LLC}^{c}, as now it is not only the bias of the numerator but the bias of the whole test statistic that is subtracted. To appreciate the effect of this change let us begin by expanding τ_{LL}^{c} as

$$\frac{2}{\sqrt{5}}\,\tau_{LL}^{c} = \tau_{LLC} + \sqrt{\frac{3N}{2}} = \sqrt{N}\frac{\frac{1}{NT}M_{12}^{p}}{\hat{\sigma}\sqrt{\frac{1}{NT^2}M_{22}^{p}}} + \sqrt{N}\frac{\frac{1}{2}\,\sigma^2}{\sigma\sqrt{\frac{\sigma^2}{6}}}$$

$$= \frac{\sqrt{N}\left(\frac{1}{NT}M_{12}^{p} + \frac{1}{2}\,\sigma^2\right)}{\hat{\sigma}\sqrt{\frac{1}{NT^2}M_{22}^{p}}} - \frac{1}{2}\,\sigma^2\sqrt{N}\left(\frac{1}{\hat{\sigma}\sqrt{\frac{1}{NT^2}M_{22}^{p}}} - \frac{1}{\sigma\sqrt{\frac{\sigma^2}{6}}}\right),$$

which, by Taylor expansion of the second term, yields

$$\frac{2}{\sqrt{5}}\,\tau_{LL}^{c} \simeq \frac{\sqrt{N}\left(\frac{1}{NT}M_{12}^{p} + \frac{1}{2}\,\sigma^2\right)}{\sqrt{\hat{\sigma}^2\frac{1}{NT^2}M_{22}^{p}}} + \hat{\sigma}^2\sqrt{\frac{27}{2\sigma^8}}\,\sqrt{N}\left(\frac{1}{NT^2}M_{22}^{p} - \frac{\sigma^2}{6}\right)$$

$$+ \frac{1}{NT^2}M_{22}^{p}\sqrt{\frac{27}{72\sigma^8}}\,\sqrt{N}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{d} \frac{\frac{\sigma^2}{\sqrt{12}}Z_1}{\sigma\sqrt{\frac{\sigma^2}{6}}} + \sigma^2\sqrt{\frac{27}{2\sigma^8}}\,\frac{\sigma^2}{\sqrt{45}}Z_2,$$

where Z_1 and Z_2 are two independent standard normal random variables. In order to obtain this results, we have used that $\frac{1}{T^2}\,\mathrm{var}(M_{22i}^{p}) \to \frac{\sigma^4}{45}$ and $\sqrt{N}(\hat{\sigma}^2 - \sigma^2) = o_p(1)$, see Lemma 2 of Moon and Phillips (2004). It follows that

$$\frac{2}{\sqrt{5}}\,\tau_{LL}^{c} \xrightarrow{d} \left(\frac{1}{\sqrt{2}} + \sqrt{\frac{3}{10}}\right)\mathcal{N}(0,1) \sim \frac{2}{\sqrt{5}}\,\mathcal{N}(0,1),$$

or $\tau_{LL}^{c} \xrightarrow{d} \mathcal{N}(0,1)$.

Thus, although the end result is the same as for τ_{LLC}^{c}, the route to normality is more complicated, and involves additional approximations, which is suggestive of poor small-sample properties. On the other hand, the bias-adjustment of LLC requires estimation of σ^2, which obviously increases the variability of their test.

The relationship between the two statistics is easily seen by noting that

$$\frac{2}{\sqrt{5}}\,\tau_{LL}^{c} = \tau_{LLC} + \sqrt{\frac{3N}{2}} = \tau_{LLC} + \frac{1}{2}\,\sqrt{N}\,\frac{\sigma}{\sqrt{\frac{\sigma^2}{6}}}$$

$$= \tau_{LLC} + \frac{1}{2}\,\sqrt{N}\,\operatorname*{plim}_{N,T\to\infty}\frac{\hat{\sigma}}{\sqrt{\frac{1}{NT^2}M_{22}^{p}}}$$

which is asymptotically equivalent to

$$\frac{1}{\sqrt{2}}\tau_{LLC}^{c} = \tau_{LLC} + \frac{\sqrt{N}}{2}\frac{\hat{\sigma}}{\sqrt{\frac{1}{NT^2}M_{22}^{p}}}.$$

The demeaning not only complicates the route to normality but also impact the local power of the tests. Consider τ^c_{LLC}. From Moon and Perron (2008) we have that under H_{1c},

$$\frac{1}{\sigma\sqrt{T}}(y_{it} - \overline{y}_i) \;\Rightarrow\; W_i(s) - \overline{W}_i + \frac{c_i}{\sqrt{N}}\,(U_i(s) - \overline{U}_i) + O_p\left(\frac{1}{N^{3/4}}\right)$$

as $T \to \infty$, which implies

$$\frac{1}{\sigma^2 T} M^p_{12i} \;\Rightarrow\; M^{\circ p}_{12i} + \frac{c_i}{\sqrt{N}}\left(M^{\circ p}_{22i} + \int_0^1 (U_i(r) - \overline{U}_i) dW_i(r)\right) + O_p\left(\frac{1}{N^{3/4}}\right).$$

Using $E(M^{\circ p}_{12i}) = \sigma^2/2$, $E(M^{\circ p}_{22i}) = \sigma^2/6$, $\mathrm{var}(M^{\circ p}_{12i}) = \sigma^2/12$ and

$$E\left(\int_0^1 (U_i(r) - \overline{U}_i) dW_i(r)\right) \;=\; -E(W_i(1)\overline{U}_i) \;=\; -\frac{1}{6}$$

it is possible to show that as $N,\, T \to \infty$

$$\begin{aligned}\frac{\sqrt{12N}}{\sigma^2}\left(\frac{1}{NT} M^p_{12} + \frac{\sigma^2}{2}\right) &\;\xrightarrow{d}\; \mathcal{N}(0,1) + \sqrt{12}\,\overline{c}\,E\left(M^{\circ p}_{22i} + \int_0^1 (U_i(r) - \overline{U}_i) dW_i(r)\right) \\ &\;\sim\; \mathcal{N}(0,1).\end{aligned}$$

Hence, under the typical local alternative given by (4) the limiting distribution of the numerator of τ^c_{LLC} does not depend on c_i. For the denominator we have as $T \to \infty$

$$\frac{1}{\sigma^2 T^2} M^p_{22i} \;\Rightarrow\; M^{\circ p}_{22i} + \frac{2c_i}{\sqrt{N}} \int_0^1 (W_i(r) - \overline{W}_i)(U_i(r) - \overline{U}_i) dr + O_p\left(\frac{1}{N^{3/4}}\right),$$

suggesting that if we consider in addition $N \to \infty$

$$\frac{1}{NT^2}\, M^p_{22} \;\xrightarrow{p}\; \sigma^2\, E(M^{\circ p}_{22i}) \;=\; \frac{\sigma^2}{6},$$

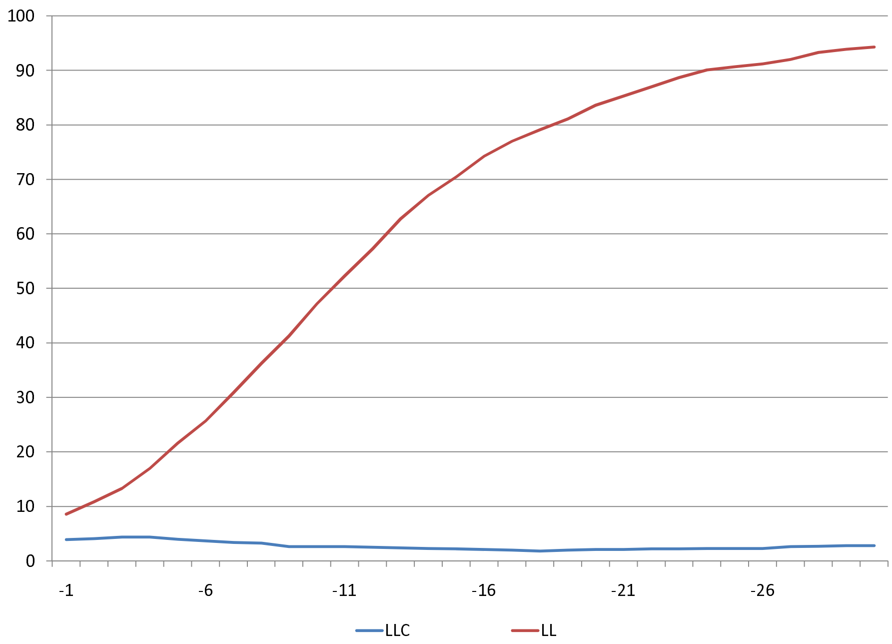
from which it follows that

$$\tau^c_{LLC} \;=\; \frac{\sqrt{2N}\left(\frac{1}{TN}\, M^p_{12} + \frac{\hat{\sigma}^2}{2}\right)}{\hat{\sigma}\sqrt{\frac{1}{NT^2} M^p_{22}}} \;\xrightarrow{d}\; \mathcal{N}(0,1).$$

In other words, τ^c_{LLC} does not have any power against H_{1c}. This is illustrated in Figure 1, which plots the local power as a function of c when the data are generated from (1), (2) and (4) with $c_i \sim U(2c, 0)$. As in Table 1, the results are based on 5,000 replications and the 5% critical value. Note in particular that the power function of τ^c_{LLC} flat, thereby corroborating the theoretical results. We also see that the power function of τ^c_{LL} is strictly increasing in c, suggesting that unlike τ^c_{LLC}, τ^c_{LL} does have power against H_{1c}.

This loss of power is rather unexpected because while it is fairly well-known that the presence of individual specific trends can be problematic in this regard, an issue that will be discussed in detail in the next section, here we only consider the case with individual specific intercepts. The reason is the absence of serial correlation, or rather the absence of correction thereof, which turns out to be rather important for a number of reasons, see Section 4 for a more detailed analysis.

Figure 1: Local power of τ^c_{LL} and τ^c_{LLC} as a function of c.



The point here is that these complications are all due to the fact that the constant is removed by least squares demeaning. Thus, in order to avoid bias, one needs to consider alternatives to least squares demeaning. For example, Breitung and Meyer (1994) suggest using the initial value y_{i0} as an estimator of the intercept. This suggests running a regression of Δy_{it} on $y^*_{it-1} = y_{it-1} - y_{i0}$. To analyze the effect of this modification, note that under the null hypothesis

$$E(\Delta y_{it} y^*_{it-1}) = E\big(\Delta y_{it}\,(y_{it-1} - y_{i0})\big) = E\left(\varepsilon_{it} \sum_{s=1}^{t-1} \varepsilon_{is}\right) = 0.$$

In other words, using y_{i0} as an estimator of the intercept removes the bias. In fact, it is not difficult to show that as $N, T \to \infty$

$$\tau^c_{BM} = \frac{\sum_{i=1}^N \sum_{t=2}^T y^*_{it-1} \Delta y_{it}}{\hat{\sigma}\sqrt{\sum_{i=1}^N \sum_{t=2}^T (y^*_{it-1})^2}} \xrightarrow{d} \mathcal{N}(0,1),$$

where τ^c_{BM} is the Breitung and Meyer (1994) statistic.

Interestingly, as pointed out by Phillips and Schmidt (1992), y_{i0} is also the maximum likelihood estimator of the intercept under H_0, which has been shown to lead to significant power gains when compared to least squares demeaning, see Madsen (2010). In fact, it is not difficult to see that under H_{1c},

$$\tau^c_{BM} \xrightarrow{d} \mathcal{N}\left(\frac{\bar{c}}{\sqrt{2}}, 1\right)$$

as $N, T \to \infty$, which is the same results we obtained earlier for the LLC statistic in the model without any deterministic terms.

To examine the extent of these gains in small samples, Table 2 reports some results based on data generated from (1), (2) and (4) with $c_i \sim U(2c, 0)$. Consistent with the results of Madsen (2010) we see that the tests based on removing the initial condition are almost uniformly more powerful than those based on least squares demeaning. We also see that this increase in power comes at no cost in terms of size accuracy. Note that if the demeaning is accomplished by subtracting the initial observation the LLC and Levin and Lin (1992) tests are the same and hence we only report the results of the latter test.[4]

Another possibility is to demean y_{it-1} recursively as $y_{it-1} - \frac{1}{t-1}\sum_{s=1}^{t-1} y_{is}$, which can be used instead of y^*_{it-1} to produce yet another unbiased and standard normally distributed test statistic.

These alternative demeaning approaches can be seen as special cases of a more general class of test statistics. In particular, letting $\Delta y_i = (\Delta y_{i2}, \ldots, \Delta y_{iT})'$ and $y_{i,-1} = (y_{i1}, \ldots, y_{iT-1})'$, these statistics can be written as

$$\frac{\sum_{i=1}^N (\Delta y_i)' C y_{i,-1}}{\hat{\sigma}\sqrt{\sum_{i=1}^N y'_{i,-1} C' C y_{i,-1}}}.$$

The matrix C has the property that $C\iota_{T-1} = 0$, where ι_{T-1} is a $(T-1) \times 1$ vector of ones. Therefore, pre-multiplying $y_{i,-1}$ by C eliminates the individual specific constant. The statistic has expectation zero if

$$E\big((\Delta y_i)' C y_{i,-1}\big) = \sigma^2 \, tr(CD) = 0,$$

[4]The performance of the LLC test under least squares demeaning can be inferred from Figure 1.

Table 2: Size and local power for different demeaning procedures.

			LL		IPS	
c	N	T	LS	ML	LS	ML
0	10	50	7.1	7.1	7.3	5.8
	20	50	6.8	6.9	7.4	6.3
	10	100	6.4	7.5	4.8	5.4
	20	100	6.6	7.0	5.3	5.8
-1	10	50	10.1	11.9	9.4	8.5
	20	50	9.3	12.9	9.7	10.3
	10	100	8.6	13.5	6.7	10.0
	20	100	9.6	13.2	8.2	11.2
-2	10	50	13.2	18.5	10.5	12.5
	20	50	12.5	20.1	11.6	14.4
	10	100	10.9	18.0	7.0	14.4
	20	100	11.7	19.9	10.0	16.2
-5	10	50	25.9	41.0	19.8	31.2
	20	50	23.7	46.1	20.9	37.3
	10	100	21.6	39.8	15.3	33.3
	20	100	24.0	43.3	17.3	36.8

Notes: The table reports the 5% rejection frequencies when the AR parameter is set to $\alpha_i = c_i/T\sqrt{N}$, where $c_i \sim U(2c, 0)$. LS and ML refer to demeaning by least squares and maximum likelihood, respectively, where the latter is based on removing the first observation from y_{it}. LL refer to the Levin and Lin (1992) test

where D is a matrix with elements $d_{jk} = 1$ if $j < k$ and $d_{jk} = 0$ for $j \geq k$. Note that in the case of least squares demeaning, $C = I_{T-1} - \frac{1}{T-1}\iota_{T-1}\iota'_{T-1}$, where I_{T-1} is the identity matrix. Since in this case $tr(CD) \neq 0$, bias correction is needed.

The same principle can be used to construct bias-corrected statistics in models with trends, an issue to be discussed in the next section.

Fact 4: Incidental trends reduce the local power of the LLC test

Suppose now that instead of (5) we have

$$y_{it} = \mu_i + \beta_i t + y_{it}^s, \tag{7}$$

where $\beta_i t$ is a unit specific trend term, giving

$$\begin{aligned}\Delta y_{it} &= -\alpha_i \mu_i + (\alpha_i + 1)\beta_i - \alpha_i \beta_i t + \alpha_i y_{it-1} + \varepsilon_{it} \\ &= \alpha_i y_{it-1} + \gamma_i' x_{it} + \varepsilon_{it} \qquad (8)\end{aligned}$$

with $x_{it} = \begin{pmatrix} 1, & t \end{pmatrix}'$.

The incidental trends problem refers to the need of having to estimate an increasing number of trend coefficients $\beta_1, ..., \beta_N$ as $N \to \infty$, which reduces the discriminatory power against H_0, see Moon and Phillips (1999).[5] In particular, as we will now demonstrate the presence of trends even has an order effect on the neighborhoods around the unit root null for which asymptotic power is non-negligible.

As Moon *et al.* (2007) show in the case with incidental trends the LLC statistic is asymptotically equivalent to

$$\tau_{LLC}^t = \frac{193}{112} \tau_{LLC} + \sqrt{\frac{252}{772}} \frac{10}{T} \frac{\sqrt{M_{22}^p}}{\hat{\sigma}},$$

where the superscript t indicates invariance with respect to the trend, while τ_{LLC} is now the LLC statistic based on the detrended data. Moon and Perron (2004) consider another statistic, which in the present setting may be written as

$$\tau_{MP}^t = \tau_{LLC} + \frac{NT}{2} \frac{\hat{\sigma}}{\sqrt{M_{22}^p}}.$$

It follows that

$$\sqrt{\frac{193}{112}} \tau_{LLC}^t = \tau_{MP}^t + \frac{15}{2T} \frac{\left(M_{22}^p - \frac{1}{15}\hat{\sigma}^2\right)}{\hat{\sigma}\sqrt{M_{22}^p}},$$

suggesting that τ_{LLC}^t will inherit some of the asymptotic properties of τ_{MP}^t. In particular, from Theorem 4 of Moon and Perron (2004) we know that τ_{MP}^t has power within $\frac{1}{N^{1/6}T}$ neighborhoods of H_0, but not for any higher powers of N and T. In particular, τ_{MP}^t has no power against H_{1c} when the neighborhood is of order $\frac{1}{T\sqrt{N}}$. The above relationship suggests that τ_{LLC}^t should have the same property. Thus, just as in the case of an intercept, we see that the presence of the trend leads to a loss of power. This is illustrated in Table 3, which plots

[5] Moon and Phillips (1999) show that the maximum likelihood estimator of the local-to-unity parameter in the stochastic trend is inconsistent. They call this phenomenon, which arises because of the presence of an infinite number of nuisance parameters, an incidental trend problem because it is analogous to the well-known incidental parameter problem in dynamic panels when T is fixed

the local power of the LLC and IPS tests for some different values of c when α_i is generated according to (4) with $c_i \sim U(0, 2c)$. In accordance with the theoretical results we see that the power can be very low and practically nonexistent in many cases if there is a trend in the model.

Table 3: Local power in the presence of incidental trends.

c	N	T	LLC	IPS
-1	10	50	9.9	10.9
	20	50	10.6	12.2
	10	100	6.9	7.7
	20	100	8.2	8.4
-4	10	50	13.4	12.4
	20	50	12.6	14.3
	10	100	10.2	9.4
	20	100	10.0	10.4
-8	10	50	22.8	22.1
	20	50	20.0	20.5
	10	100	16.9	14.6
	20	100	14.9	14.9

Notes: The table reports the 5% rejection frequencies when the AR parameter is set to $\alpha_i = c_i / T\sqrt{N}$, where $c_i \sim U(2c, 0)$.

In view of the previous fact one might think that this loss of power is due to the fact that the detrending is carried out using least squares. However, this is not true. Take as an example the study of Breitung (2000), who proposes a generalized version of the demeaning by initial value procedure discussed in the previous section. Specifically, using y_{i0} and $\frac{1}{T}\sum_{t=2}^{T} \Delta y_{it} = \frac{1}{T}(y_{iT} - y_{i0})$ as estimators of the constant and trend, respectively, Breitung (2000) proposes replacing (8) with a regression of Δy_{it}^* on y_{it-1}^*, where $y_{it}^* = y_{it} - y_{i0} - \frac{1}{T}(y_{iT} - y_{i0})t$ and

$$\Delta y_{it}^* = s_t \left(\Delta y_{it} - \frac{1}{T-t}(y_{iT} - y_{it}) \right)$$

with $s_t^2 = (T-t)/(T-t+1)$. The effect of this is easily seen by noting that

$$\begin{aligned} E(\Delta y_{it}^* y_{it-1}^*) &= s_t\, E\left(\left(\Delta y_{it} - \frac{1}{T-t}(y_{iT} - y_{it})\right)\left(y_{it} - y_{i0} - \frac{1}{T}(y_{iT} - y_{i0})\right)\right) \\ &= s_t\, E\left(\left(\Delta y_{it}^s - \frac{1}{T-t}(y_{iT}^s - y_{it}^s)\right)\left(y_{it}^s - y_{i0}^s - \frac{1}{T}(y_{iT}^s - y_{i0}^s)\right)\right) \\ &= s_t\, E\left(\left(\varepsilon_{it} - \frac{1}{T-t}(y_{iT}^s - y_{it}^s)\right)\left(y_{it-1}^s - \frac{t-1}{T} y_{iT}^s\right)\right) \\ &= s_t\left(\frac{t-1}{T}\sigma^2 - \frac{(t-1)(T-t)}{(T-t)T}\sigma^2\right) \;=\; 0, \end{aligned}$$

showing that the bias has been successfully eliminated.

However, as Moon *et al.* (2006) show, just as with τ_{MP}^t and τ_{LLC}^t, the Breitung (2000) test has no power in neighborhoods that shrinks to zero at a faster rate than $\frac{1}{N^{1/4}T}$. The reduced power effect in the presence of trends is therefore not specific to τ_{MP}^t and τ_{LLC}^t but is a general property of this type of tests. In fact, as Ploberger and Phillips (2002) show, the panel unit root test that maximizes the average local power has nontrivial power in local neighborhoods that shrink at the same rate, $\frac{1}{N^{1/4}T}$.

Fact 5: The initial condition may affect the asymptotic properties of the tests

The power of panel unit root tests is usually evaluated while assuming that all N units are initiated at zero. Although this is a convenient assumption that simplifies the theoretical considerations, as Moon *et al.* (2007) points out it is very unrealistic and, as we will see, by no means innocuous. Suppose for example that y_{it} is generated according to (5) with a constant and where y_{it}^s is as in (2). But suppose now that instead of setting y_{i0}^s to zero, we follow the convention in the time series context and consider the initial condition as a draw from the unconditional distribution under the stationary alternative, see for example Phillips and Sul (2003). That is, we set

$$y_{i0}^s \;=\; \frac{\sigma}{\sqrt{1-\rho_i^2}}\eta_i,$$

where η_i is independent and identically distributed with mean $\overline{\eta}$ and variance σ_η^2. Note that for $\rho_i < 1$ and $\eta_i \sim N(0,1)$ this initial condition implies that y_{it}^s is stationary.

Similar to what we had before when $y_{i0}^s = 0$, Harris *et al.* (2009) show that under H_{1c}, as

$T \to \infty$

$$\frac{1}{\sigma\sqrt{T}}(y_{it} - \overline{y}_i) \Rightarrow \frac{\eta_i}{N^{1/4}}\sqrt{\frac{-c_i}{2}}\left(r - \frac{1}{2}\right) + W_i(s) - \overline{W}_i + \frac{c_i}{\sqrt{N}}(U_i(s) - \overline{U}_i) + O_p\left(\frac{1}{N^{3/4}}\right),$$

where $U_i(s)$ is again the integral of $W_i(r)$ over $[0, s]$, implying

$$\frac{1}{\sigma^2 T} M_{12i}^p \Rightarrow M_{12i}^{\circ p} + \frac{c_i}{\sqrt{N}}\left(M_{22i}^{\circ p} + \int_0^1 (U_i(r) - \overline{U}_i)dW_i(r)\right) - \frac{\eta_i}{N^{1/4}}\sqrt{\frac{-c_i}{2}}\int_0^1 \left(r - \frac{1}{2}\right)dW_i(r) + O_p\left(\frac{1}{N^{3/4}}\right).$$

It follows that as N, $T \to \infty$

$$\frac{\sqrt{12N}}{\sigma^2}\left(\frac{1}{NT}M_{12}^p + \frac{\sigma^2}{2}\right) \xrightarrow{d} \mathcal{N}(0,1) + \sqrt{12}\,\overline{c}\,E\left(M_{22i}^{\circ p} + \int_0^1 (U_i(r) - \overline{U}_i)dW_i(r)\right) - \operatorname*{plim}_{N\to\infty}\frac{1}{N^{3/4}}\sum_{i=1}^{N}\eta_i\sqrt{\frac{-c_i}{2}}\int_0^1\left(r - \frac{1}{2}\right)dW_i(r).$$

But $E(dW_i(r)) = 0$, suggesting that the last term on the right-hand side is $O_p(1/N^{1/4})$. Thus, since the second term is zero,

$$\frac{\sqrt{12N}}{\sigma^2}\left(\frac{1}{NT}M_{12}^p + \frac{\sigma^2}{2}\right) \xrightarrow{d} \mathcal{N}(0,1),$$

which is the same result as the one we obtained when $y_{i0}^s = 0$. By further using $\frac{1}{T^2N}M_{22}^p \xrightarrow{p} \sigma^2/6$ it follows that $\tau_{LLC}^c \xrightarrow{d} \mathcal{N}(0,1)$. Thus, just as before the asymptotic distribution of τ_{LLC}^c is independent of both $\overline{c}$ and y_{i0}^s.

The analysis for τ_{LL}^c is more tedious, but the qualitative effect is easily appreciated by focusing on $\sqrt{N}(\frac{1}{NT^2}M_{22}^p - \frac{\sigma^2}{6})$. Remember that from before that

$$\frac{2}{\sqrt{5}}\tau_{LL}^c = \frac{\sqrt{N}\left(\frac{1}{NT}M_{12}^p + \frac{1}{2}\sigma^2\right)}{\sqrt{\hat{\sigma}^2\frac{1}{NT^2}M_{22}^p}} + \hat{\sigma}^2\sqrt{\frac{27}{2\sigma^8}}\sqrt{N}\left(\frac{1}{NT^2}M_{22}^p - \frac{\sigma^2}{6}\right) + O_p(1),$$

where the $O_p(1)$ term is a remainder comprised of $\sqrt{N}$ times the error of a Taylor expansion of the inverse square root of $\frac{1}{NT^2}M_{22}^p$ around $\sigma^2/6$ plus another term that is asymptotically negligible. We already know that the first term on the right-hand side is independent of $\overline{c}$ and y_{i0}^s. Any dependence must therefore come through the second term, which then also determines the $O_p(1)$ term. By using the same steps as before,

$$\frac{1}{\sigma^2T^2}M_{22i}^p \Rightarrow M_{22i}^{\circ p} + \frac{2c_i}{\sqrt{N}}\int_0^1 (U_i(r) - \overline{U}_i)(W_i(r) - \overline{W}_i)dr + \frac{\eta_i^2 c_i}{2\sqrt{N}}\int_0^1\left(r - \frac{1}{2}\right)^2 dr + \frac{\eta_i\sqrt{-2c_i}}{N^{1/4}}\int_0^1\left(r - \frac{1}{2}\right)(W_i(r) - \overline{W}_i)dr + O_p\left(\frac{1}{N^{3/4}}\right)$$

as $T \to \infty$, and therefore

$$\sqrt{N}\left(\frac{1}{NT^2}M_{22}^p - \frac{\sigma^2}{6}\right) \overset{d}{\to} \frac{\sigma^2}{\sqrt{45}}\mathcal{N}(0,1) + 2\overline{c}E\left(\int_0^1 (U_i(r) - \overline{U}_i)(W_i(r) - \overline{W}_i)dr\right)$$
$$+ \frac{E(\eta_i^2)\overline{c}}{2}\int_0^1 \left(r - \frac{1}{2}\right)^2 dr$$
$$+ \operatorname*{plim}_{N\to\infty}\frac{1}{N^{3/4}}\sum_{i=1}^N \eta_i\sqrt{-2c_i}\int_0^1 \left(r - \frac{1}{2}\right)(W_i(r) - \overline{W}_i)dr$$

as $T \to \infty$ and then $N \to \infty$. But $E(W_i(r) - \overline{W}_i) = 0$ and $\int_0^1 (r - 1/2)^2 dr = 1/12$, which means that while the fourth term is eliminated, the third term is not, and therefore the dependence upon η_i remains. Thus, unlike τ_{LLC}^c, the asymptotic distribution of τ_{LL}^c does depend on the initial condition.

The same is true for the IPS test. In fact, as Harris *et al.* (2009) show,

$$\tau_{IPS}^c \overset{d}{\to} \mathcal{N}(0,1) + \overline{c}\left(0.282 - 0.135(\overline{\eta}^2 + \sigma_\eta^2)\right),$$

which shows that the local power of the IPS test is decreasing in $\overline{\eta}^2$ and σ_η^2. In particular, note that if the initial condition is large enough so that $0.282 > 0.135(\overline{\eta}^2 + \sigma_\eta^2)$, then this test is no longer unbiased. This is illustrated in Figures 2 and 3, which plot the local power of the IPS test for different combinations of N and $\overline{\eta}$ when $\sigma_\eta^2 = 0$ and α_i is generated as in (4) with $c_i = -10$ for all i. We see that if there is a constant present then the power is decreasing in both N and $\overline{\eta}$, while if there is no deterministic component then the power is very close to one.

4 Models with short-run dynamics

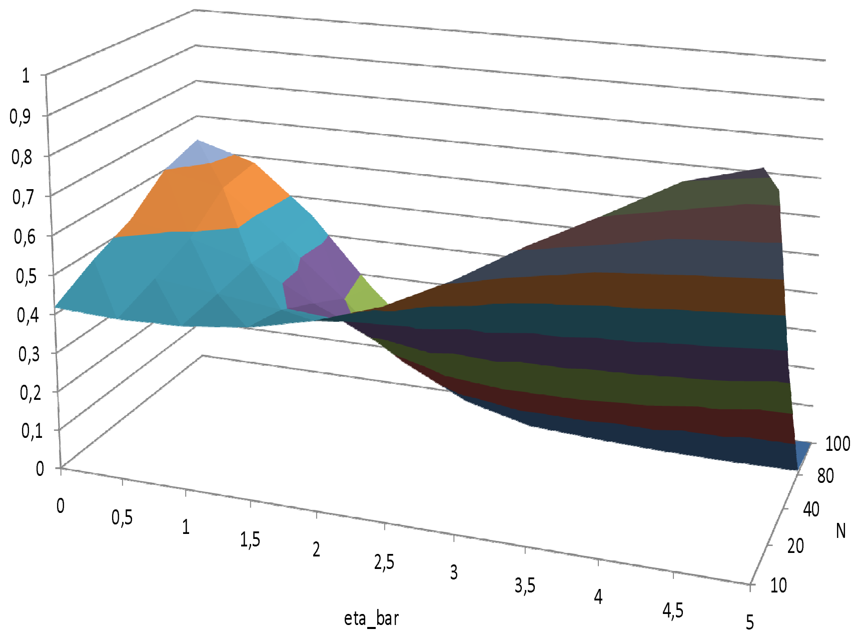
Fact 6: Lag augmentation does not remove the effects of serial correlation

Suppose that (1) holds so that y_{it} is purely stochastic, but that the error ε_{it} in (2) is no longer independent across t. In particular, suppose that ε_{it} follows a stationary and invertible AR process of order ℓ,

$$\phi_i(L)\varepsilon_{it} = \left(1 - \sum_{j=1}^{\ell}\phi_{ij}L^j\right)\varepsilon_{it} = \varepsilon_{it} - \sum_{j=1}^{\ell}\phi_{ij}\varepsilon_{it-j} = e_{it}, \quad (9)$$

where L is the lag operator and e_{it} is a mean zero error that has variance σ^2 for all i but is otherwise independent across both i and t. Note that the long-run variance of ε_{it}, henceforth

Figure 2: Local power of τ^c_{IPS} for different initial values.



denoted ω_i^2, can be expressed in terms of the lag polynomial $\phi_i(L)$ as

$$\omega_i^2 = \frac{\sigma^2}{\phi_i(1)^2}.$$

Under H_0, (1), (2) and (9) can be combined to obtain the following augmented DF (ADF) regression:

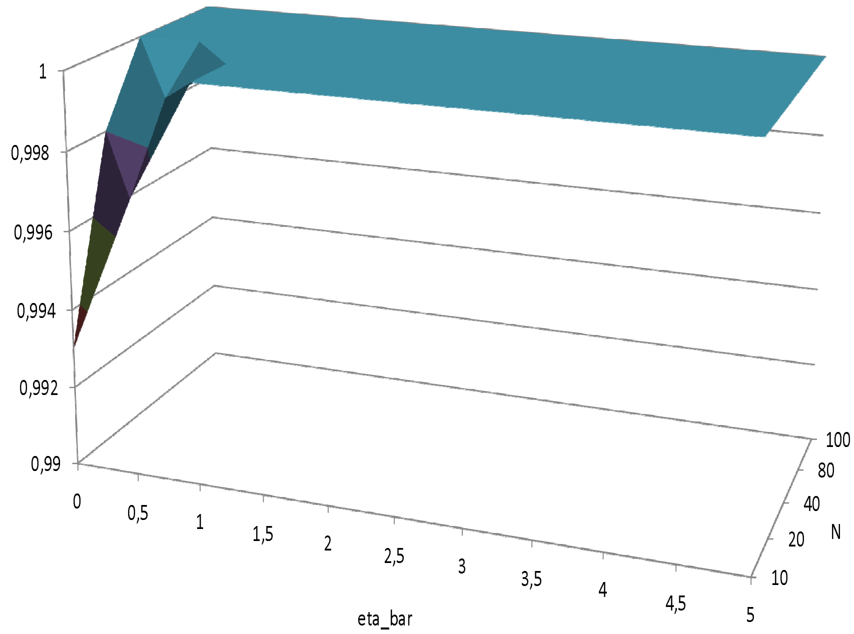
$$\Delta y_{it} = \alpha_i y_{it-1} + \sum_{j=1}^{\ell} \phi_{ij} \Delta y_{it-j} + e_{it} = \alpha_i y_{it-1} + \gamma_i' x_{it} + e_{it}, \qquad (10)$$

where $x_{it} = (\ \Delta y_{it-1}, \ \ldots, \ \Delta y_{it-\ell}\ )'$ is now the vector of lagged differences with $\gamma_i = (\ \phi_{i1}, \ \ldots, \ \phi_{i\ell}\ )'$ being the associated vector of lag coefficients. This gives rise to the ADF test statistic,

$$\tau_i = \frac{M^p_{12i}}{\hat{\sigma}_i \sqrt{M^p_{22i}}} = \hat{\alpha}_i \frac{\sqrt{M^p_{22i}}}{\hat{\sigma}_i},$$

where $\hat{\sigma}_i^2 = \frac{1}{T}(M^p_{11i} - \hat{\alpha}_i M^p_{12i})$ and $\hat{\alpha}_i = M^p_{12i}/M^p_{22i}$, which again suppresses the dependence upon ℓ. Note also that in this setup M^p_{abi} takes the projection onto the lags of Δy_{it} rather than onto a vector of deterministic components as in Section 3.

Figure 3: Local power of τ_{IPS} for different initial values.



Under H_0, using M^e_{12i} and M^e_{13i} to denote M_{12i} and M_{13i} with e_{it} in place of Δy_{it},

$$\begin{aligned}\frac{1}{T}M^{ep}_{12i} &= \frac{1}{T}M^e_{12i} - \frac{1}{T}M^e_{13i}\,M^{-1}_{33i}\,M_{32i} = \frac{1}{T}M^e_{12i} + O_p\left(\frac{1}{\sqrt{T}}\right), \\ \frac{1}{T^2}M^p_{22i} &= \frac{1}{T^2}M_{22i} - \frac{1}{T^2}M_{23i}\,M^{-1}_{33i}\,M_{32i} = \frac{1}{T^2}M_{22i} + O_p\left(\frac{1}{T}\right).\end{aligned}$$

These results, together with $\frac{1}{T}M^p_{11i} \stackrel{p}{\to} \sigma^2$ and

$$\begin{pmatrix} \frac{1}{T}\,M^e_{12i} \\ \frac{1}{T^2}\,M_{22i} \end{pmatrix} \Rightarrow \begin{pmatrix} \sigma\omega_i\,M^\circ_{12i} \\ \omega_i^2\,M^\circ_{22i} \end{pmatrix}$$

imply that

$$\tau_i = \frac{M^p_{12i}}{\hat{\sigma}_i\sqrt{M^p_{22i}}} \Rightarrow \frac{M^\circ_{12i}}{\sqrt{M^\circ_{22i}}}.$$

Thus, the asymptotic distribution of τ_i is not affected by the presence of short-run dynamics, suggesting that the distribution of the IPS statistic should be unaffected too, see Section 4 of IPS. In other words, with this test lag augmentation successfully removes the short-run dynamics of the panel. However, this is not true for the LLC statistic. Note first that if there are no deterministic components, then as N, $T \to \infty$

$$\frac{1}{T\sqrt{N}} \sum_{i=1}^{N} M_{12i}^{p} \overset{d}{\to} \frac{\sigma\sqrt{\overline{\omega^2}}}{\sqrt{2}} \mathcal{N}(0,1),$$

where $\overline{\omega^2} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \omega_i^2$. Furthermore,

$$\frac{1}{NT^2} M_{22}^{p} \overset{p}{\to} \frac{\overline{\omega^2}}{2},$$

and therefore,

$$\tau_{LLC} = \frac{M_{12}^{p}}{\hat{\sigma}\sqrt{M_{22}^{p}}} \overset{d}{\to} \mathcal{N}(0,1).$$

Thus, in this case lag augmentation removes the effect of the short-run dynamics also for the LLC statistic. However, the situation changes dramatically if the model includes a constant or a linear time trend. Consider for example the case with short-run dynamics and a constant, in which we let $x_{it} = \begin{pmatrix} 1, & \Delta y_{it-1}, & \ldots, & \Delta y_{it-\ell} \end{pmatrix}'$ and re-define M_{12i}^{p} and M_{22i}^{p} accordingly. This yields

$$\lim_{N,T\to\infty} E\left(\frac{1}{TN} M_{12}^{p}\right) = \sigma\overline{\omega} \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} E(M_{12i}^{\circ}) \to -\frac{\sigma\overline{\omega}}{2},$$

where $\overline{\omega} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \omega_i$. It follows that in this case lag augmentation does not remove short-run parameters from the mean of the statistic. To cope with this problem LLC propose a bias and serial correlation corrected version of τ_{LLC}, which we again denote by τ_{LLC}^{c}.[6] The problem is that since the mean of τ_{LLC} depends on both σ^2 and ω_i^2, for the bias-correction to work these parameters have to be estimated consistently, an issue that will be considered in more detail below. It follows that in the presence of deterministic terms lag augmentation alone is not enough to remove the short-run parameters from the asymptotic distribution of the LLC statistic.

For the estimation of ω_i^2 LLC propose using

$$\hat{\omega}_i^2 = \frac{1}{T} \sum_{t=2}^{T} (\Delta y_{it})^2 + \frac{2}{T} \sum_{j=1}^{q-1} \left(1 - \frac{j}{q}\right) \sum_{t=j+1}^{T} \Delta y_{it} \Delta y_{it-j},$$

which is the conventional Newey and West (1994) long-run variance estimator. It is important to note that by using Δy_{it} this estimator is in fact imposing H_0. Thus, if H_0 holds then we

[6] Note that because σ^2 is assumed to be equal across i, there is no need for any correction for heteroskedasticity as in LLC.

have from Andrews (1991) that $\hat{\omega}_i^2 \xrightarrow{p} \omega_i^2$ as $T \to \infty$ with $q \to \infty$ and $\frac{q}{T} \to 0$. This indicates that the following bias-corrected statistic can be used

$$\tau_{LLC}^c \;=\; \sqrt{2}\,\tau_{LLC} + \frac{NT}{\sqrt{2}} \frac{\hat{\omega}}{\sqrt{M_{22}^p}}$$

with $\hat{\omega} = \frac{1}{N}\sum_{i=1}^N \hat{\omega}_i$, whose asymptotic distribution is given by

$$\tau_{LLC}^c \;=\; \frac{\sqrt{2N}\left(\frac{1}{TN} M_{12}^p + \frac{1}{2}\,\hat{\sigma}\,\hat{\omega}\right)}{\hat{\sigma}\sqrt{\frac{1}{NT^2} M_{22}^p}} \;\xrightarrow{d}\; \frac{\frac{\sigma\sqrt{\overline{\omega^2}}}{\sqrt{6}}\,\mathcal{N}(0,1)}{\sigma\sqrt{\frac{\overline{\omega^2}}{6}}} \;\sim\; \mathcal{N}(0,1),$$

see Theorem 5 of LLC. However, although theoretically not an issue as long as $q \to \infty$ and $\frac{q}{T} \to 0$, in practice the optimal truncation lag q unknown. There is also the problem that $\hat{\omega}_i^2$ tends to zero under H_{1a}, an issue that we will discuss in more detail in the next section.

To sidestep these difficulties, Breitung and Das (2005) propose to pre-whiten the variables. Their idea is as follows. Under H_0 we have

$$\Delta y_{it} \;=\; \sum_{j=1}^{\ell} \phi_{ij}\Delta y_{it-j} + e_{it}, \tag{11}$$

or, in terms of levels,

$$y_{it} \;=\; \sum_{s=1}^{t} \Delta y_{is} \;=\; \sum_{j=1}^{\ell} \phi_{ij} \sum_{s=1}^{t} \Delta y_{is-j} + \sum_{s=1}^{t} e_{is} \;=\; \sum_{j=1}^{\ell} \phi_{ij} y_{it-j} + y_{it}^s,$$

where y_{it}^s is as in (2) but with H_0 imposed. It is a random walk with serially uncorrelated increments e_{it}. For simplicity we also assume that $y_{i0} = 0$. Thus, in contrast to $\frac{1}{\sqrt{T}}\, y_{it}$, whose long-run variance is given by ω_i^2, the long-run variance of $\frac{1}{\sqrt{T}}\, y_{it}^s$ is just σ^2. For the estimation of the lag coefficients ϕ_{ij}, Breitung and Das (2005) recommend using the above regression in first differences with H_0 imposed. For a fixed ℓ this yields

$$\frac{1}{\sqrt{T}}\, y_{it}^* \;=\; \frac{1}{\sqrt{T}}\, y_{it}^s - \sum_{j=1}^{\ell} (\hat{\phi}_{ij} - \phi_{ij})\, \frac{1}{\sqrt{T}}\, y_{it-j} \;=\; \frac{1}{\sqrt{T}}\, y_{it}^s + O_p\left(\frac{1}{\sqrt{T}}\right) O_p(1),$$

where $\hat{\phi}_{ij}$ is the least squares estimate of ϕ_{ij} in (11), which means that $\sqrt{T}(\hat{\phi}_{ij} - \phi_{ij}) = O_p(1)$. Similarly,

$$\Delta y_{it}^* \;=\; e_{it} - \sum_{j=1}^{\ell} (\hat{\phi}_{ij} - \phi_{ij})\Delta y_{it-j} \;=\; e_{it} + O_p\left(\frac{1}{\sqrt{T}}\right).$$

Thus, replacing Δy_{it} and y_{it-1} by Δy_{it}^* and y_{it-1}^*, respectively, eliminates the effects of the serial correlation without requiring any estimation of ω_i^2.

Table 4: Size for different corrections for short-run dynamics.

		$\phi_1 = 0.5$				$\phi_1 = -0.5$			
		LLC		IPS		LLC		IPS	
N	T	Aug	Pre	Aug	Pre	Aug	Pre	Aug	Pre
					No lags				
10	20	7.5	7.5	3.8	3.8	13.4	13.4	21.7	21.7
20	20	5.2	5.2	3.1	3.1	22.4	22.4	31.7	31.7
40	20	2.7	2.7	0.6	0.6	45.1	45.1	54.4	54.4
10	50	4.8	4.8	1.7	1.7	29.2	29.2	42.0	42.0
20	50	2.8	2.8	0.4	0.4	50.0	50.0	67.5	67.5
40	50	0.7	0.7	0.1	0.1	82.9	82.9	90.6	90.6
					The true number of lags				
10	20	8.9	7.1	5.6	3.5	8.4	7.6	4.5	4.0
20	20	8.9	5.8	3.5	2.7	7.6	7.2	3.2	3.0
40	20	8.2	6.1	2.0	0.9	6.0	5.7	1.7	1.3
10	50	7.5	6.7	3.6	2.6	6.5	6.0	3.5	3.4
20	50	6.9	5.5	3.5	2.8	6.2	5.8	3.6	3.5
40	50	6.1	5.3	2.5	2.0	5.2	5.2	2.8	2.7
					The Schwarz Bayesian information criterion				
10	20	10.0	7.3	6.0	3.2	8.1	8.0	4.8	4.2
20	20	9.3	6.8	4.3	2.8	7.9	7.1	3.5	3.0
40	20	8.0	5.9	2.1	0.9	6.0	5.3	2.1	1.5
10	50	7.5	6.7	3.6	2.6	6.4	6.1	3.4	3.4
20	50	6.9	5.6	3.6	2.8	6.3	5.8	3.6	3.4
40	50	5.9	5.2	2.5	1.7	5.2	5.3	2.9	2.7

Notes: The table reports the 5% rejection frequencies under H_0. ϕ_1 refers to the first-order AR serial correlation parameter. Aug and Pre refer to least squares augmentation and pre-whitening, respectively.

In Table 4 we compare the size accuracy of the IPS and LLC tests using both least squares lag augmentation and pre-whitening. The data are generated from (1), (2) and (9) with $\ell = 1$, which makes ϕ_1, the assumed homogenous first-order AR coefficient, an interesting nuisance parameter to study. For the choice of lag length we consider three alternatives. The first is to ignore the serial correlation and to set $\ell = 0$, while the second is to set ℓ equal to its true value. The third alternative is to chose ℓ in a data-dependent fashion by using the Schwarz Bayesian information criterion with a maximum of five lags.

The first thing to notice is the size distortions that result from ignoring the serial correlation, especially when $\phi_1 = -0.5$, and the effectiveness by which they are removed in the two correction procedures. Note also that there are basically no differences in the results depending on whether ℓ is treated as known or not. Similarly, unreported simulation results suggest that there no difference in power. The point here is that since the performance is so similar, why not use pre-whitening, which leads to a much simpler test?

Fact 7: The consistency of the LLC test depends on the long-run variance estimator

As long as that $q \to \infty$ such that $\frac{q}{T} \to 0$, we have that under H_0, $\hat{\omega}_i^2 - \omega_i^2 = o_p(1)$. Thus, provided that H_0 holds, $\hat{\omega}_i$ is consistent for ω_i, which as we have seen is a requirement for τ_{LLC}^c to be asymptotically normal. The problem is under the fixed alternative hypothesis H_{1a} y_{it} is stationary, and therefore Δy_{it} is over-differenced with a zero long-run variance. Thus, in contrast to what happens under H_0, in this case $\hat{\omega}_i^2$ does not converge to ω_i^2 but in fact goes to zero suggesting that $\hat{\omega}_i$ should go to zero too. In fact, as Westerlund (2010) shows, if $q \to \infty$ with $N, T \to \infty$ and $\frac{q}{T} \to 0$,

$$\hat{\omega} = \frac{1}{N}\sum_{i=1}^{N}\hat{\omega}_i = O_p\left(\frac{1}{\sqrt{q}}\right).$$

From Section 2 we know that $\tau_{LLC} = O_p(\sqrt{NT})$ under H_{1a}, suggesting that τ_{LLC}^c is of the same order. Therefore, to determine the effect of the inconsistency of $\hat{\omega}$ on

$$\tau_{LLC}^c = \sqrt{2}\,\tau_{LLC} + \frac{NT}{\sqrt{2}}\frac{\hat{\omega}}{\sqrt{M_{22}^p}},$$

we only need to consider the second term, the bias, which under H_{1a} can be written as

$$\frac{NT}{\sqrt{2}}\frac{\hat{\omega}}{\sqrt{M_{22}^p}} = \frac{\sqrt{NT}}{\sqrt{2}}\hat{\omega}\frac{1}{\sqrt{\frac{1}{NT}M_{22}^p}} = \sqrt{NT}\,O_p\left(\frac{1}{\sqrt{q}}\right)O_p(1).$$

In order to appreciate the implications of this last result, suppose that q is set independent of T, so that the bias term is $O_p(\sqrt{NT})$. The problem here is that while $\tau_{LLC} \to -\infty$, the bias term is diverging at the same rate but in the opposite direction, which means that τ_{LLC}^c need not be a consistent test. The only way to prevent this from happening is therefore to set q as a function of T, ensuring that the order of the bias term is lower than $O_p(\sqrt{NT})$, and hence that $\tau_{LLC}^c \to -\infty$. Thus, the practice of using the data to select q is not expected to work here. A similar problem arises when trying to use the residual $\hat{\varepsilon}_{it} = \Delta y_{it} - \hat{\alpha}_i y_{it-1}$ instead of Δy_{it}, which would otherwise seem like a natural thing to do in view of the inconsistency of

$\hat{\omega}_i^2$ under the stationary alternative. However, this means that under the alternative $\hat{\omega}_i^2 \xrightarrow{p} \omega_i^2 > 0$, which again leads to consistency problems.

To illustrate these results Table 5 reports the size-adjusted power of τ_{LLC}^c for three different bandwidth selection rules. The first is the automatic rule of Newey and West (1994), while the other two are deterministic, and involve setting q either equal to $4(T/100)^{2/9}$ as suggested by Newey and West (1994) or equal to $3.21T^{1/3}$ as in LLC. The data are generated from (2), (4) and (5) with $c_i \sim U(0, 2c)$. In agreement with the results of Westerlund (2010) we see that the power can be very low and practically nonexisting unless $b = 3.21T^{1/3}$, which is also the most generous rule considered. For example, if $T = 100$, then $4(T/100)^{2/9} = 4$ while $3.21T^{1/3} \simeq 15$, an increase by almost a factor of four.[7]

Table 5: Size-adjusted power of the LLC test for different bandwidths.

			Bandwidth selection rule		
c	N	T	NW	$4(T/100)^{2/9}$	$3.21T^{1/3}$
−5	10	100	3.9	3.9	6.2
	20	100	3.6	3.5	5.0
	10	200	1.9	1.6	3.5
	20	200	2.5	2.5	2.8
−10	10	100	3.9	3.6	10.8
	20	100	3.7	3.7	7.9
	10	200	2.1	1.3	5.6
	20	200	2.7	1.8	3.9
−20	10	100	7.2	6.0	38.5
	20	100	6.8	6.8	29.7
	10	200	5.2	1.6	20.6
	20	200	3.4	2.0	11.6
−40	10	100	40.4	39.7	86.0
	20	100	27.3	28.3	85.2
	10	200	31.4	8.9	73.3
	20	200	18.5	4.2	65.2

Notes: The table reports the 5% size-adjusted rejection frequencies when the AR parameter is set to $\alpha_i = c_i/T\sqrt{N}$, where $c_i \sim U(0, 2c)$. NW refers to the automatic bandwidth selection rule of Newey and West (1994).

[7] See Westerlund (2010) for some simulation results of the size of the test.

5 Cross section dependence

Fact 8: Cross-section dependence leads to deceptive inference

We consider two types of dependence, weak and strong. The first type refers to a situation in which all the eigenvalues of the covariance matrix of y_{it} are bounded as $N \to \infty$, which rules out the presence of unobserved common factors, but allows the cross-sectional units to be for example spatially correlated. This is the case considered by O'Connell (1998) in his pioneering work on the effects of cross-section dependence on panel unit root tests. The second type of dependence refers to a situation when at least one eigenvalue diverges with N, which arises when cross-section dependence is due to common factors.

Suppose as in the previous section that (3) holds so that

$$\begin{pmatrix} \Delta y_{1t} \\ \vdots \\ \Delta y_{Nt} \end{pmatrix} = \begin{pmatrix} \alpha_1 & & 0 \\ & \ddots & \\ 0 & & \alpha_N \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ \vdots \\ y_{Nt-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{Nt} \end{pmatrix}$$

or, in matrix format,

$$\Delta y_t \; = \; \Pi y_{t-1} + \varepsilon_t.$$

However, instead of looking at the case when ε_{it} is dependent across t, we now consider the case when it is dependent across i.

In particular, let us begin by assuming that all eigenvalues of the covariance matrix

$$\Omega \; = \; \text{cov}(\varepsilon_t) \; = \; E(\varepsilon_t \varepsilon_t')$$

are bounded as $N \to \infty$, which means that the dependence is of the weak form. By the spectral decomposition, $\Omega = \Omega^{1/2}(\Omega^{1/2})' = V\Lambda V'$, where Λ is the diagonal matrix with the ordered eigenvalues $\lambda_1 \geq \cdots \geq \lambda_N$ along the diagonal, while V is the matrix of orthonormal eigenvectors. Let $y_t^* = \begin{pmatrix} y_{1t}^*, & \ldots, & y_{Nt}^* \end{pmatrix}' = \Omega^{-1/2} y_t$, which under H_0 is nothing but a vector of uncorrelated random walks.

The above assumptions imply that M_{12} can be written as

$$M_{12} \; = \; \sum_{t=2}^{T} y_{t-1}' \Delta y_t \; = \; \sum_{t=2}^{T} (y_{t-1}^*)' \Lambda \Delta y_t^* \; = \; \sum_{i=1}^{N} \lambda_i M_{12i}^*,$$

where $M_{12i}^* = \sum_{t=2}^{T} y_{it-1}^* \Delta y_{it}^*$. Let $\overline{\lambda} = \lim\limits_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \lambda_i$ with a similar definition of $\overline{\lambda^2}$. By using the results of the previous sections,

$$\frac{1}{NT} M_{12} \; \xrightarrow{p} \; \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \lambda_i \frac{1}{T} E(M_{12i}^*) \; = \; \overline{\lambda} \frac{1}{T} E(M_{12i}^*) \; = \; 0$$

such that $\frac{1}{T\sqrt{N}} M_{12} \overset{d}{\to} \frac{\sqrt{\overline{\lambda^2}}}{\sqrt{2}} \mathcal{N}(0,1)$ as N, $T \to \infty$. Similarly,

$$\frac{1}{NT^2} M_{22} = \frac{1}{NT^2} \sum_{t=2}^{T} y'_{t-1} y_{t-1} = \frac{1}{NT^2} \sum_{i=1}^{N} \lambda_i M^*_{22i} \overset{p}{\to} \overline{\lambda} \frac{1}{T^2} E(M^*_{22i}) \to \frac{\overline{\lambda}}{2},$$

where the limit is taken as $N \to \infty$ followed by $T \to \infty$. This result, together with $\hat{\sigma}^2 \overset{p}{\to} \overline{\lambda}$, suggest that as N, $T \to \infty$

$$\tau_{LLC} = \frac{\frac{1}{T\sqrt{N}} M_{12}}{\hat{\sigma}\sqrt{\frac{1}{NT^2} M_{22}}} \overset{d}{\to} \frac{\frac{\sqrt{\overline{\lambda^2}}}{\sqrt{2}} \mathcal{N}(0,1)}{\sqrt{\frac{\overline{\lambda}}{2}}} \sim \frac{\sqrt{\overline{\lambda^2}}}{\overline{\lambda}} \mathcal{N}(0,1),$$

which summarizes Theorem 1 of Breitung and Das (2005). In other words, if the dependence is weak then τ_{LLC} is still asymptotically normal. However, as long as $\lambda_i \neq \lambda_j$ for at least some $i \neq j$, $\frac{\sqrt{\overline{\lambda^2}}}{\overline{\lambda}} > 1$ and so the variance will tend to increase with deceptive inference as a result. A similar result applies to τ_{IPS}. That is, the IPS test will also tend to be misleading in the presence of weak cross-section dependence.

In order to analyze the effects of strong dependence suppose that the error ε_{it} in (3) has the following common factor representation:

$$\varepsilon_{it} = \theta_i g_t + \epsilon_{it}. \tag{12}$$

Assuming that ϵ_{it} and g_t are white noise processes with unit variance, the covariance matrix of ε_{it} can be written as

$$\Omega = \Theta\Theta' + I_N,$$

where $\Theta = \begin{pmatrix} \theta_1, & \ldots, & \theta_N \end{pmatrix}'$. The main difference here in comparison to the case with weak dependence is the presence of g_t, which suggests that the largest eigenvalue of Ω is no longer bounded but is in fact $O(N)$, while the remaining eigenvalues are $O(1)$. Intuitively, the information regarding the common component Θg_t accumulates as we sum up the observations across i and therefore the largest eigenvalue will increase with N. It follows that

$$\frac{1}{NT} M_{12} = \frac{\lambda_1}{NT} \sum_{t=2}^{T} y^*_{1t-1} \Delta y^*_{1t} + O_p\left(\frac{1}{\sqrt{N}}\right),$$

$$\frac{1}{NT^2} M_{22} = \frac{\lambda_1}{NT^2} \sum_{t=2}^{T} (y^*_{1t-1})^2 + O_p\left(\frac{1}{\sqrt{N}}\right)$$

and therefore,

$$\frac{1}{\sqrt{N}} \tau_{LLC} \Rightarrow \frac{\int_0^1 W(s) dW(s)}{\sqrt{\int_0^1 W(s)^2 ds}},$$

where $W(s)$ is a standard Brownian motion that does not depend on i. This is the limiting distribution of the DF statistic, see Breitung and Das (2008). Thus, $\tau_{LLC} = O_p(\sqrt{N})$, and moreover,

$$\begin{aligned}\lim_{N,T\to\infty} P(\tau_{LLC} < -1.645) &= \lim_{N,T\to\infty} P\left(\frac{1}{\sqrt{N}}\tau_{LLC} < -\frac{1.645}{\sqrt{N}}\right) \\ &= P\left(\int_0^1 W(s)dW(s) < 0\right) \simeq 0.7,\end{aligned}$$

showing that the LLC test is upwards biased in the case of strong cross-section dependence.

Let us now consider the case when there is a stationary factor that enters the AR process in the following fashion:

$$y_{it} = \theta_i f_t + y_{it}^s, \tag{13}$$

where y_{it}^s is as in (2), while the common factor f_t is assumed to satisfy $f_t = \delta f_{t-1} + u_t$ with $|\delta| < 1$ and u_t is white noise. This means that under H_0,

$$\Delta y_t = \Theta\Delta f_t + \Delta y_t^s = \Theta((1-\delta)f_t + u_t) + \varepsilon_t,$$

suggesting that

$$\text{cov}(\Delta y_t) = \frac{(1-\delta)^2}{1-\delta^2}\Theta\Theta' + \sigma^2 I_N.$$

To see how this is going to affect the limiting distribution of the LLC test write

$$\frac{1}{NT}M_{12} = \frac{1}{NT}\sum_{t=2}^{T}(f_{t-1}\Theta'\Theta\Delta f_t + (y_{t-1}^s)'\varepsilon_t + f_{t-1}'\Theta'\varepsilon_t + (y_{t-1}^s)'\Theta\Delta f_t),$$

where, since ε_{it} is assumed to be independent of ε_{jt} and f_t,

$$\frac{1}{\sqrt{NT}}\sum_{t=2}^{T} f_{t-1}'\Theta'\varepsilon_t = \frac{1}{N}\sum_{i=1}^{N}\theta_i\frac{1}{\sqrt{T}}\sum_{t=2}^{T} f_{t-1}\varepsilon_{it} = O_p(1).$$

Moreover, since

$$\frac{1}{T\sqrt{N}}\sum_{t=2}^{T}(y_{t-1}^s)'\varepsilon_t = O_p(1),$$

where $\frac{1}{T\sqrt{N}}\sum_{t=2}^{T}(y_{t-1}^s)'\Theta\Delta f_t$ is of the same order, we obtain

$$\frac{1}{NT}M_{12} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\theta_i^2 f_{t-1}\Delta f_t + O_p\left(\frac{1}{\sqrt{N}}\right) \xrightarrow{p} -\frac{1-\delta}{1-\delta^2}\overline{\theta^2}$$

as $N, T \to \infty$, where $\overline{\theta^2} = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\theta_i^2$ (Breitung and Das, 2008).

Also,

$$\frac{1}{NT^2}M_{22} = \frac{1}{NT^2}\sum_{t=2}^{T}(\Theta f_{t-1}+y^s_{t-1})'(\Theta f_{t-1}+y^s_{t-1})$$
$$= \frac{1}{NT^2}\sum_{t=2}^{T}(y^s_{t-1})'y^s_{t-1}+O_p\left(\frac{1}{T}\right)+O_p\left(\frac{1}{T\sqrt{N}}\right) \xrightarrow{p} \frac{\sigma^2}{2}.$$

By collecting these results we obtain

$$\frac{\hat{\sigma}}{\sqrt{N}}\tau_{LLC} \xrightarrow{p} -\frac{\frac{1-\delta}{1-\delta^2}\overline{\theta^2}}{\sqrt{\frac{\sigma^2}{2}}},$$

or $\tau_{LLC} = O_p(\sqrt{N})$, suggesting that the size of the LLC test will tend to one as $N \to \infty$.

As for the IPS test, note that

$$\frac{1}{T}M_{12i} = \frac{1}{T}\sum_{t=2}^{T}y_{it-1}\Delta y_{it} = \frac{1}{T}\sum_{t=2}^{T}(\theta f_{t-1}+y^s_{it-1})(\theta_i\Delta f_t+\varepsilon_{it})$$
$$= \frac{1}{T}\sum_{t=2}^{T}y^s_{it-1}\varepsilon_{it}+\theta_i\frac{1}{T}\sum_{t=2}^{T}y^s_{it-1}\Delta f_t+\theta_i^2\frac{1}{T}\sum_{t=2}^{T}f_{t-1}\Delta f_t+O_p\left(\frac{1}{\sqrt{T}}\right),$$

while

$$\hat{\sigma}_i^2\frac{1}{T^2}M_{22i} = \hat{\sigma}_i^2\frac{1}{T^2}\sum_{t=2}^{T}y^2_{it-1} = \hat{\sigma}_i^2\frac{1}{T^2}\sum_{t=2}^{T}(\theta_i f_{t-1}+y^s_{it-1})^2 = Q_i+O_p\left(\frac{1}{T}\right),$$

where $Q_i = \sigma^2\frac{1}{T^2}\sum_{t=2}^{T}(y^s_{it-1})^2$. Hence,

$$\tau_i = \frac{M_{12i}}{\hat{\sigma}_i\sqrt{M_{22i}}} = \tau_i^s+\frac{\theta_i}{\sqrt{Q_i}}\frac{1}{T}\sum_{t=2}^{T}y^s_{it-1}\Delta f_t+\frac{\theta_i^2}{\sqrt{Q_i}}\frac{1}{T}\sum_{t=2}^{T}f_{t-1}\Delta f_t+O_p\left(\frac{1}{\sqrt{T}}\right),$$

where τ_i^s is the DF test based on y^s_{it}. Using the law of iterated expectations with respect to the sigma field generated by f_t, $\mathcal{F}$ say, and passing $T \to \infty$, we get

$$E(\tau_i|\mathcal{F}) = E(\tau_i^s)+\frac{1}{T}\sum_{t=2}^{T}E\left(\frac{\theta_i}{\sqrt{Q_i}}y^s_{it-1}|\mathcal{F}\right)\Delta f_t$$
$$+ E\left(\frac{\theta_i^2}{\sqrt{Q_i}}|\mathcal{F}\right)\frac{1}{T}\sum_{t=2}^{T}f_{t-1}\Delta f_t+O_p\left(\frac{1}{\sqrt{T}}\right)$$

and therefore, $E(\tau_i) \neq E(\tau_i^s)$. In other words, the usual correction terms for the mean of τ_i are incorrect as they ignore the correlation between f_{t-1} and Δf_t. It follows that under H_0 the IPS statistic is $O_p(\sqrt{N})$.

Summarizing the results reported in this section we find that the presence of cross-section dependence is likely to lead to misleading inference. The extreme case being when the dependence is of the strong form, in which the test statistics actually become divergent. This last result is particularly interesting since in our setup f_t is stationary, and the presence of unit roots usually eliminates the effects of such variables. This is illustrated in Table 6, which depicts the size of the LLC and IPS tests in the presence of a single common factor. For simplicity the data are generated from (2) and (13) with no deterministic components or serial correlation. The factor is generated with AR coefficient $\delta = 0$ and loading $\theta_i = \theta$ for all i. The results show that the distortions are increasing in N, which is clearly in agreement with the theoretical predictions.

The lesson we learn from there results is that since cross-section dependence is more of a rule rather than the exception, some form of correction is usually needed, see Breitung and Pesaran (2008) and the references therein.

Table 6: Size in the presence of a stationary common factor.

θ	N	T	LLC	IPS
0	10	50	7.4	4.3
	20	50	6.1	3.3
	10	100	6.7	4.0
	20	100	6.1	4.6
1	10	50	29.7	41.4
	20	50	53.2	64.1
	10	100	47.3	59.9
	20	100	73.3	84.5
2	10	50	92.0	95.0
	20	50	99.4	99.8
	10	100	99.1	99.7
	20	100	100.0	100.0

Notes: The table reports the 5% rejection frequencies under H_0. θ refers to the factor loading which is identical for all i.

Fact 9: The IPS and LLC tests fail under cross-unit cointegration

Suppose that (13) holds, and that $\rho_i = \rho$ for all i with $|\rho| < 1$ such that y_{it}^s is stationary. Suppose also that $\delta = 1$, so that f_t is unit root non-stationary. Under these conditions, y_{it} and y_{jt} are cointegrated, a situation commonly referred to as cross-unit cointegration.

If we assume that $\theta_i \neq 0$ for all i, the vector y_t has a unit root and therefore a panel unit root test should accept the null hypothesis that all N units have a unit root. However, it is not difficult to show that in this case the LLC and IPS tests are going to be oversized, leading to the wrong conclusion.

Consider first the LLC test. By using

$$\begin{aligned}
\frac{1}{NT} M_{12} &= \frac{1}{NT} \sum_{i=1}^{N}\sum_{t=2}^{T} \theta_i^2 f_{t-1}\Delta f_t + \frac{1}{NT} \sum_{i=1}^{N}\sum_{t=2}^{T} y_{it-1}^s \Delta y_{it} + O_p\left(\frac{1}{\sqrt{N}}\right) \\
&\Rightarrow \overline{\theta^2} \int_0^1 W(s)dW(s) - \frac{1-\rho}{1-\rho^2}\sigma^2, \\
\frac{1}{NT^2} M_{22} &= \frac{1}{NT^2} \sum_{t=2}^{T} (\Theta f_{t-1} + y_{t-1}^s)'(\Theta f_{t-1} + y_{t-1}^s) \\
&= \frac{1}{NT^2} \sum_{i=1}^{N}\sum_{t=2}^{T} \theta_i^2 f_{t-1}^2 + O_p\left(\frac{1}{\sqrt{N}}\right) \Rightarrow \overline{\theta^2} \int_0^1 W(s)^2 ds,
\end{aligned}$$

it follows that

$$\frac{1}{\sqrt{N}} \tau_{LLC} \Rightarrow \frac{\overline{\theta^2} \int_0^1 W(s)dW(s) - \frac{1-\rho}{1-\rho^2}\sigma^2}{\sqrt{\overline{\theta^2} \int_0^1 W(s)^2 ds}}$$

or $\tau_{LLC} = O_p(\sqrt{N})$, showing that the size of the LLC test is tending to one as $N \to \infty$.

The analysis of the IPS test is very similar to the case when f_t what white noise. We begin by noting that

$$\begin{aligned}
\frac{1}{T} M_{12i} &= \frac{1}{T} \sum_{t=2}^{T} (\theta_i f_{t-1} + y_{it-1}^s)(\theta_i u_t + \Delta y_{it}^s) \\
&= \theta_i^2 \frac{1}{T} \sum_{t=2}^{T} f_{t-1} u_t + \theta_i \frac{1}{T} \sum_{t=2}^{T} f_{t-1}\Delta y_{it}^s + \frac{1}{T} \sum_{t=2}^{T} y_{t-1}^s \Delta y_{it}^s + O_p\left(\frac{1}{\sqrt{T}}\right), \\
\hat{\sigma}_i^2 \frac{1}{T^2} M_{22i} &= \hat{\sigma}_i^2 \frac{1}{T^2} \sum_{t=2}^{T} (\theta_i f_{t-1} + y_{it-1}^s)^2 = V_i + O_p\left(\frac{1}{T}\right).
\end{aligned}$$

where $V_i = \sigma_i^2 \frac{1}{T^2} \sum_{t=2}^{T} \theta_i^2 f_{t-1}^2$. Thus, using τ^f to denote the DF test based on f_t,

$$\tau_i = \tau^f + \frac{\theta_i}{\sqrt{V_i}} \frac{1}{T} \sum_{t=2}^{T} f_{t-1}\Delta y_{it}^s + \frac{1}{\sqrt{V_i}} \frac{1}{T} \sum_{t=2}^{T} y_{t-1}^s \Delta y_{it}^s + O_p\left(\frac{1}{\sqrt{T}}\right),$$

whose asymptotic distribution as $T \to \infty$ is clearly different from the usual DF distribution, as the second and third terms on the right-hand side are $O_p(1)$. The IPS statistic is therefore $O_p(\sqrt{N})$.

Consequently, the presence of cross-unit cointegrating relationships causes the IPS and LLC statistics to become divergent, which is in agreement with the simulation results of Banerjee *at al.* (2005), showing that the presence of such relationships can lead to severe size distortions. One way to alleviate this problem in practice is to use the PANIC approach of Bai and Ng (2004).

Fact 10: Factor based tests can be restrictive

An important feature of the factor model in (13) is that f_t and y^s_{it} can have different orders of integration, see for example Bai and Ng (2008). In most other work on panel unit root tests with common factors this is not the case. In particular, consider the data generating process adopted by Moon and Perron (2004), Moon *et al.* (2007), Pesaran (2007) and Phillips and Sul (2003), which in the case with a single factor and no deterministic components can be written as in (12). This specification differs from (13) in that it essentially specifies the dynamics of the observed series, whereas (12) specifies the dynamics of unobserved components. Assuming $y_{i0} = 0$ and $\rho_i = \rho$ for all i, (12) can be written in terms of (13) as follows:

$$y_{it} \;=\; \theta_i(\rho f_{t-1} + g_t) + (\rho y^s_{it-1} + \epsilon_{it}) \;=\; \theta_i f_t + y^s_{it}.$$

It follows that if $\rho = 1$, then both components are non-stationary. Conversely, if $|\rho| < 1$, then both components are stationary. The common and idiosyncratic components of the above model are therefore restricted to have the same order of integration. Note also that when ρ_i is heterogeneous the above model cannot be expressed in terms of (13). However under the null that $\rho_i = 1$ for all i, then it is nested in (13). Studies that explicitly takes (13) as the data generating process are Bai and Ng (2004) and Breitung and Das (2008). As it is difficult to motivate that the factor and idiosyncratic component can be represented by the same process, these studies consider a more realistic scenario.

Fact 11: Sequential limits need not imply joint limits

As pointed out by Phillips and Moon (1999), the sequential limit theory wherein $T \to \infty$ is followed by $N \to \infty$, is very straightforward to apply and generally leads to quick asymptotic results in a variety of settings. The main reason for this is that passing $T \to \infty$ while holding N fixed in the first step allows one to focus only on the first-order terms, as the higher order terms are eliminated prior to averaging over N. However, this feature can also be deceptive in its simplicity because it hides the need to control the relative expansion rate of the two dimensions. Indeed, as Phillips and Moon (1999) show, sequential convergence does not necessarily imply convergence in the joint limit as $N, T \to \infty$ simultaneously. In some situations the sequential limit theory may therefore break down. The problem is that the connection between the two methods is not well understood, and many researchers view breakdowns as extreme events.

Consider as an example the generalized least squares test of Breitung and Das (2005) and Harvey and Bates (2005),

$$\tau_{GLS} = \frac{\sum_{t=2}^{T} y'_{t-1}\hat{\Omega}^{-1}\Delta y_t}{\sqrt{\sum_{t=2}^{T} y'_{t-1}\hat{\Omega}^{-1}y_{t-1}}},$$

which is suitable for testing H_0 in the presence of weak cross-section dependence. Here y_t is again the vector stacking y_{it}, while $\hat{\Omega}$ is such that $\sqrt{T}\,(\hat{\Omega}-\Omega) = O_p(1)$.

We begin by deriving the sequential limit distribution of τ_{GLS}, and then we show that this distribution need not be the same as the one obtained when using joint limits.

Applying a Taylor expansion yields

$$\hat{\Omega}^{-1} = \Omega^{-1} + O_p\left(\frac{1}{\sqrt{T}}\right).$$

Moreover, by a functional central limit theorem, $\frac{1}{\sqrt{T}} y_t \Rightarrow B(s)$ as $T \to \infty$, where $B(s) = \Omega^{1/2}W(s)$ and $W(s)$ is now a N-dimensional vector standard Brownian motion. It follows that as $T \to \infty$ with N kept fixed,

$$\tau_{GLS} = \frac{\frac{1}{T}\sum_{t=2}^{T} y'_{t-1}\hat{\Omega}^{-1}\Delta y_t}{\sqrt{\frac{1}{T^2}\sum_{t=2}^{T} y'_{t-1}\hat{\Omega}^{-1}y_{t-1}}} \Rightarrow \frac{\int_0^1 B(s)'\Omega^{-1}dB(s)}{\sqrt{\int_0^1 B(s)'\Omega^{-1}B(s)ds}} = \frac{\int_0^1 W(s)'dW(s)}{\sqrt{\int_0^1 W(s)'W(s)ds}}.$$

But the elements of $W(s)$ are independent, suggesting that as $N \to \infty$

$$\frac{\int_0^1 W(s)'dW(s)}{\sqrt{\int_0^1 W(s)'W(s)ds}} = \frac{\frac{1}{\sqrt{N}}\int_0^1 W(s)'dW(s)}{\sqrt{\frac{1}{N}\int_0^1 W(s)'W(s)ds}} \xrightarrow{d} \mathcal{N}(0,1).$$

Thus, in the sequential limit as $T \to \infty$ and then $N \to \infty$

$$\tau_{GLS} \xrightarrow{d} \mathcal{N}(0,1).$$

Consider next the joint limit of the same test statistic. By using the results of Breitung and Das (2005),

$$\begin{aligned}
\frac{1}{T\sqrt{N}}\sum_{t=2}^{T} y'_{t-1}\hat{\Omega}^{-1}\Delta y_t &= \frac{1}{T\sqrt{N}}\sum_{t=2}^{T} y'_{t-1}\Omega^{-1}\Delta y_t + O_p\left(\frac{N}{\sqrt{T}}\right), \\
\frac{1}{NT^2}\sum_{t=2}^{T} y'_{t-1}\hat{\Omega}^{-1}y_{t-1} &= \frac{1}{NT^2}\sum_{t=2}^{T} y'_{t-1}\Omega^{-1}y_{t-1} + O_p\left(\frac{\sqrt{N}}{\sqrt{T}}\right)
\end{aligned}$$

as N, $T \to \infty$, from which it follows that

$$\tau_{GLS} = \frac{\frac{1}{T\sqrt{N}}\sum_{t=2}^{T} y'_{t-1}\hat{\Omega}^{-1}\Delta y_t}{\sqrt{\frac{1}{NT^2}\sum_{t=2}^{T} y'_{t-1}\hat{\Omega}^{-1}y_{t-1}}} = \frac{\frac{1}{T\sqrt{N}}\sum_{t=2}^{T} y'_{t-1}\Omega^{-1}\Delta y_t}{\sqrt{\frac{1}{NT^2}\sum_{t=2}^{T} y'_{t-1}\Omega^{-1}y_{t-1}}} + O_p\left(\frac{N}{\sqrt{T}}\right)$$

Using the previous results we obtain

$$\frac{\frac{1}{T\sqrt{N}}\sum_{t=2}^{T}(y_{t-1}^*)'\Delta y_t^*}{\sqrt{\frac{1}{NT^2}\sum_{t=2}^{T}(y_{t-1}^*)'y_{t-1}^*}} \xrightarrow{d} \mathcal{N}(0,1).$$

Only if we assume that $\frac{N}{\sqrt{T}} \to 0$ as $N, T \to \infty$ we end up with the same asymptotic distribution resulting from a sequential limit theory. In other words, T must be large relative to N^2 to achieve a reliable approximation by the standard normal distribution. This makes sense even from an empirical point of view, as $\hat{\Omega}$ is singular unless $T \geq N$, a fact not accounted for when using the sequential limit theory. It also explains the poor small sample properties of the test if T is small relative to N^2, as documented by Breitung and Das (2005).

Of course, this example is still quite specific, which makes it difficult to appreciate the generality of the critique. Let us therefore consider another example. In particular, let us reconsider the limiting null distribution of τ_{IPS} under the assumption that ε_{it} is normal, independent and identically distributed, in which case we know from Phillips (1987) that

$$\frac{1}{\sqrt{T}} y_{it} = \sigma W_i(s) + O_p\left(\frac{1}{T}\right).$$

Using a Taylor series expansion and $\sqrt{T}(\hat{\sigma}_i^2 - \sigma^2) = O_p(1)$ we obtain

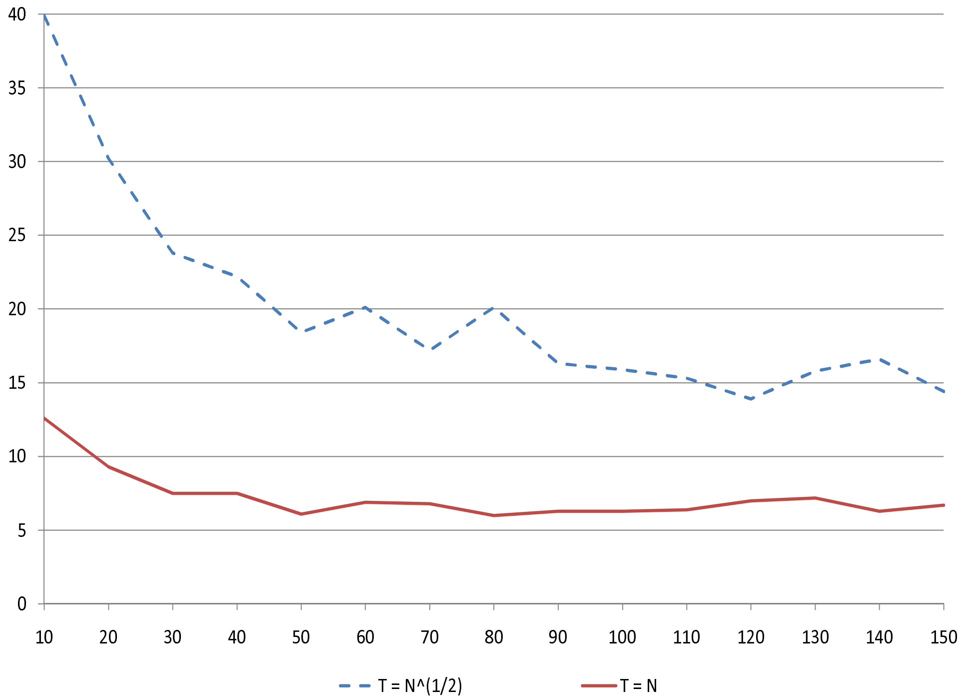
$$\tau_i = \frac{M_{12i}}{\hat{\sigma}_i\sqrt{M_{22i}}} = \frac{M_{12i}^\circ}{\sqrt{M_{22i}^\circ}} + O_p\left(\frac{1}{T}\right),$$

implying that

$$\begin{aligned}\tau_{IPS} &= \frac{\sqrt{N}(\overline{\tau} - E(\tau))}{\sqrt{\text{var}(\tau)}} = \frac{1}{\sqrt{N\,\text{var}(\tau)}}\sum_{i=1}^{N}(\tau_i - E(\tau)) \\ &= \frac{1}{\sqrt{N\,\text{var}(\tau)}}\sum_{i=1}^{N}\left(\frac{M_{12i}^\circ}{\sqrt{M_{22i}^\circ}} - E(\tau)\right) + O_p\left(\frac{\sqrt{N}}{T}\right),\end{aligned}$$

where the first term on the right-hand side converges to a standard normal distribution as $N \to \infty$. Thus, for τ_{IPS} to be asymptotically normal, one needs to assume that $\frac{\sqrt{N}}{T} \to 0$ as $N, T \to \infty$, as otherwise the $O_p(\sqrt{N}/T)$ remainder will not disappear (see also IPS 2003). A similar result applies to τ_{LLC}. The point here is that if we use the sequential limit method where N is treated as fixed in the first step then this remainder is $O_p(1/T)$, which vanishes as $T \to \infty$. The sequential limit method therefore breaks down unless $\frac{\sqrt{N}}{T} \to 0$. But if this result holds in the current very restrictive case with normal innovations, it is expected to hold also under more relaxed conditions. In practice this means that in general one needs $T >> N$ for these tests to work properly.

Figure 4: Size of τ_{LLC} when T is set as a function of N.

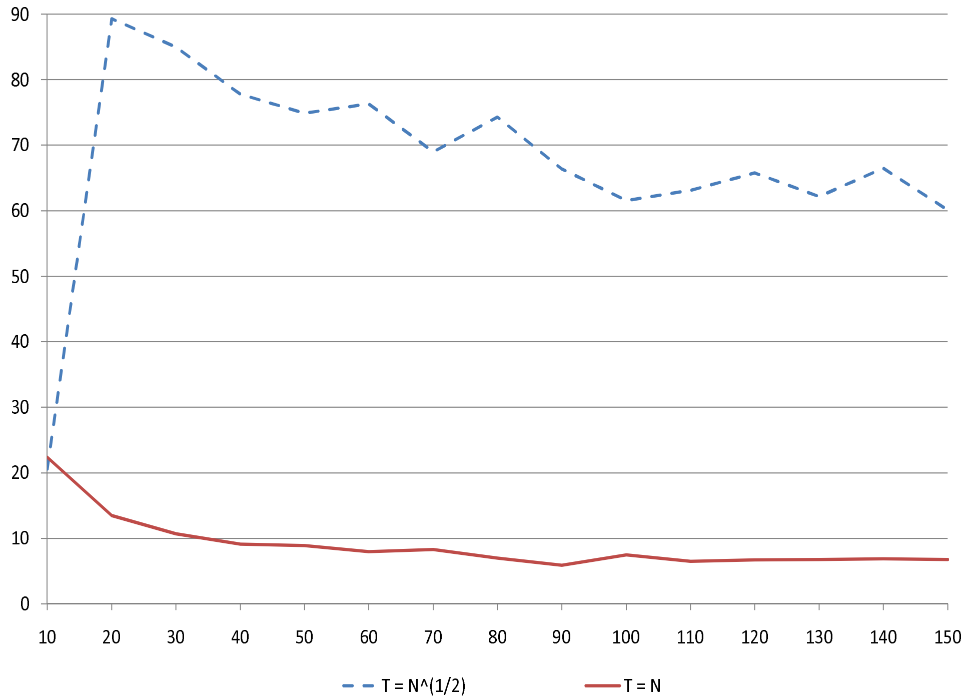


Figures 4 and 5 illustrate this point by plotting the size of the LLC and IPS tests at the 5% level when T is set as a function of N. If $T = N$ the condition that $\frac{\sqrt{N}}{T} \to 0$ is satisfied, while if $T = \sqrt{N}$, then the condition is violated. The model includes a constant but otherwise there are no nuisance parameters to correct for. In contrast to the case when $T = N$, in which both tests tend to perform well, we see that setting $T = \sqrt{N}$ generally leads to substantial size distortions, especially for the IPS test.

6 Concluding remarks

This paper points to a number of facts that are oftentimes overlooked when considering the problem of testing for a unit root in panel data. Suppose for example that one would like to test the null hypothesis that the variable y_{it} has a unit root versus the alternative that it is stationary with a nonzero mean, a very common research scenario. The by far most common way of carrying out such a test is to use the traditional DF approach of applying least squares to an intercept-augmented autoregression. Being so common it is easy to get the impression that demeaning in this way is the best way to accommodate the nonzero mean in y_{it}. But

Figure 5: Size of τ_{IPS} when T is set as a function of N.



this is not true. Indeed, as we show in the paper the inclusion of the intercept introduces a bias in the estimated AR coefficient, which then has to be corrected somehow. However, in so doing we find that the resulting corrected test is likely to suffer from low power and may even become inconsistent in some circumstances. As a response to this a few alternative demeaning procedures are suggested.

In this example, although applying traditional time series techniques leads to a larger computational burden and poorer small-sample performance, usually there are no fundamental shortcomings or flawed inference. Unfortunately, this is not always the case. Quite on the contrary, we find that ignoring these facts will in many cases lead to serious side-effects, including a complete breakdown of the whole test procedure. One example of such a situation is when y_{it} is contaminated by cross-section dependence in the form of common factors, in which case a failure to account for these factors can cause the test statistic to become divergent.

The implication is that one should be careful not to approach the testing problem from a too narrow and stylized perspective. In particular, we believe that the usual practice of looking at the problem from mainly a time series perspective can be deceptive in its simplicity, typically ignoring many important issues such as cross-sectional dependence, incidental trends and joint limit restrictions.

References

Andrews, D. W. K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica* **59**, 817–858.

Bai, J., and S. Ng (2004). A Panic Attack on Unit Roots and Cointegration. *Econometrica* **72**, 1127–1177.

Bai, J., and S. Ng (2008). Panel Unit Root Tests with Cross-Section Dependence: A Further Investigation. Forthcoming in *Econometric Theory*.

Bai, J., and Kao, C. (2006), On the Estimation and Inference of a Panel Cointegration Model with Cross-Sectional Dependence, in B. Baltagi (ed.) Contributions to Economic Analysis, Elsevier, 3–30.

Baltagi, B., and C. Kao (2000). Nonstationary Panels, Cointegration in Panels, and Dynamic Panels: A Survey. *Advances in Econometrics* **15**, 7-52.

Banerjee, A. (1999). Panel Data Unit Roots and Cointegration: An Overview. *Oxford Bulletin of Economics and Statistics* **61**, 607–630.

Banerjee, A., M. Marcellino and C. Osbat (2004). Testing for PPP: Should we Use Panel Methods? *Empirical Economics* **30**, 77–91.

Bowman, D. (1999). Efficient Tests for Autoregressive Unit Roots in Panel Data. International Finance Discussion Paper 646, Board of Governors of the Federal Reserve System.

Breitung, J. (2000). The Local Power of Some Unit Root Tests for Panel Data. In B. Baltagi (Ed.), *Advances in Econometrics: Nonstationary Panels, Panel Cointegration, and Dynamic Panels* **15**, 161–178. JAI, Amsterdam.

Breitung, J. and W. Meyer (1994). Testing for Unit Roots in Panel Data: Are Wages on Different Bargaining Levels Cointegrated? *Applied Economics* **26**, 353–361.

Breitung, J. and S. Das (2005). Panel Unit Root Tests Under Cross Sectional Dependence. *Statistica Neerlandica* **59**, 414–433.

Breitung, J. and S. Das (2008). Testing for Unit Roots in Panels With a Factor Structure. *Econometric Theory* **24**, 88–108.

Breitung, J., and M. H. Pesaran (2008). Unit Roots and Cointegration in Panels. In Mátyás, L., and P. Sevestre (Eds.), *The Econometrics of Panel Data*, 279–322. Klüwer Academic Publishers.

Choi, I. (2006). Nonstationary Panels. *Palgrave Handbooks of Econometrics* **1**, 511–539. Palgrave Macmillan, New York.

Dickey, D. A., and W. A. Fuller (1979). Distribution of the Estimates for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association* **74**, 427–431.

Elliott, G., T. J. Rothenberg and J. H. Stock (1996). Efficient Tests for an Autoregressive Unit Root. *Econometrica* **64**, 813–836.

Hahn, J., and G. Kuersteiner (2002). Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects when both N and T are Large. *Econometrica* **70**, 1639–57.

Harvey, A., and D. Bates (2005). Multivariate Unit Root Tests, Stability and Convergence. DAE Working Paper No. 301, University of Cambridge.

Harris, R. D. F., and E. Tzavalis (1999). Inference for Unit Roots in Dynamic Panels where the Time Dimension is Fixed. *Journal of Econometrics* **91**, 201–226.

Harris, D., D. I. Harvey, S. J. Leybourne and N. D. Sakkas (2009). Local Asymptotic Power of the Im-Pesaran-Shin Panel Unit Root Test and the Impact of Initial Observations. Forthcoming in *Econometric Theory*.

Im, K. S., M. H. Peseran and Y. Shin (2003). Testing for Unit Roots in Heterogeneous Panels. *Journal of Econometrics* **115**, 53–74.

Kao, C. (1999). Spurious Regression and Residual-Based Tests for Cointegration in Panel Data. *Journal of Econometrics* **90**, 1–44.

Levin, A., C. Lin, and C.-J. Chu (2002). Unit Root Tests in Panel Data: Asymptotic and Finite-sample Properties. *Journal of Econometrics* **108**, 1–24.

Levin, A., and C. Lin (1992). Unit Root Tests in Panel Data: Asymptotic and Finite-sample Properties. University of California working paper 92–23.

Levin, A., and C. Lin (1993). Unit Root Tests in Panel Data: New Results. University of California working paper 93–56.

Madsen, E. (2010). Unit Root Inference in Panel Data Models where the Time-Series Dimension is Fixed: A Comparison of Different Tests. *Econometrics Journal* **13**, 63–94

Moon, R., and B. Perron (2004). Testing for Unit Root in Panels with Dynamic Factors. *Journal of Econometrics* **122**, 81–126.

Moon, H. R., and P. C. B. Phillips (1999). Maximum Likelihood Estimation in Panels with Incidental Trends. *Oxford Bulletin of Economics and Statistics* **61**, 771–748.

Moon, H. R., and P. C. B. Phillips (1999). Asymptotic Local Power of Pooled t-Ratio Tests for Unit Roots in Panels with Fixed Effects. *Econometrics Journal* **11**, 80–104.

Moon, H. R., B. Perron and P. C. B. Phillips (2007). Incidental Trends and the Power of Panel Unit Root Tests. *Journal of Econometrics* **141**, 416–459.

Newey, W., and K. West (1994). Autocovariance Lag Selection in Covariance Matrix Estimation. *Review of Economic Studies* **61**, 613–653.

Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica* **49**, 1417–1426.

Pedroni, P. (2004). Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis. *Econometric Theory* **20**, 597–625.

Pesaran, M. H. (2007). A Simple Panel Unit Root Test in the Presence of Cross Section Dependence. *Journal of Applied Econometrics* **22**, 265–312.

Phillips, P. C. B. (1987). Time Series Regression with a Unit Root. *Econometrica* **55**, 277–301.

Phillips, P. C. B., and H. R. Moon (1999). Linear Regression Limit Theory of Nonstationary Panel Data. *Econometrica* **67**, 1057–1111.

Phillips, P. C. B. and D. Sul (2003). Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence. *Econometrics Journal* **6**, 217–259.

Ploberger, W., and P. C. B. Phillips (2002). Optimal Testing for Unit Roots in Panel Data. Unpublished manuscript.

Quah, D. (1994). Exploiting Cross-Section Variations for Unit Root Inference in Dynamic Panels. *Economics Letters* **44**, 9–19.

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

Westerlund, J. (2010). A Note on the Use of the LLC Panel Unit Root Test. *Empirical Economics* **37**, 517–531.