

WHAT DETERMINES LOCAL POWER OF PANEL UNIT ROOT TESTS?*

Joakim Westerlund[†]
University of Gothenburg
Sweden

Rolf Larsson
Uppsala University
Sweden

September 6, 2010

Abstract

In this note we offer an explanation for the common observation that, in contrast to theoretical predictions, the local power of panel unit root tests does not seem to be determined completely by the mean of the drift parameter.

JEL Classification: C22; C23.

Keywords: Panel unit root test; Local power function.

1 Motivation

Except for the case with unit-specific deterministic trends that needs to be estimated, most results available regarding the local power of panel unit root tests consider a neighborhood of unity that shrink towards the null hypothesis at the rate $\frac{1}{\sqrt{NT}}$, where N and T denote the cross-section and time series dimensions, respectively. In this setting it has been shown that the local power of most existing tests depends on the mean of the drift parameter, c_i say, where $i = 1, \dots, N$, but not on the heterogeneity. This means that one can just as well treat c_i as being equal across i ; there are no additional insights to be gained by allowing c_i to vary, at least not from a power point of view.

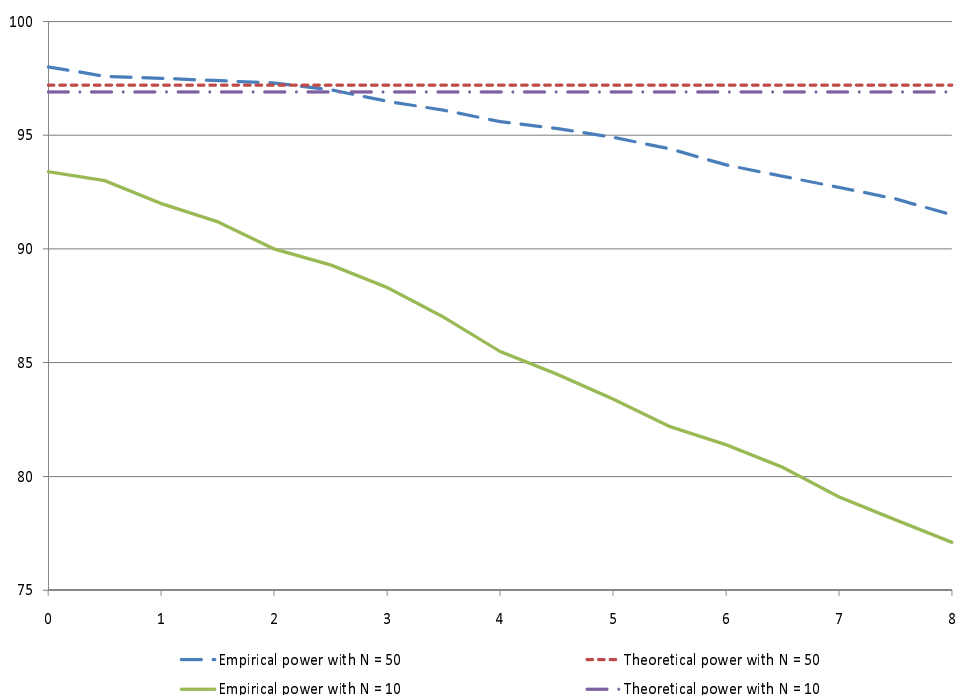
The fact that according to theory power should only depend on the mean drift is somewhat of an anomaly, because in small samples there is also a dependence on higher moments.

*The authors would like to thank the Jan Wallander and Tom Hedelius Foundation for financial support under research grant numbers P2005–0117:1 and 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.

Indeed, as Moon and Perron (2008, page 91) conclude from their simulations, “Despite our theoretical results, there is somewhat of a power loss against a heterogeneous alternative in finite samples.”¹ Let us illustrate this point using the Levin and Lin (1992) pooled t -test for an unit root. This test is one of the most widely researched, and we are going to use it in this note as a canonical example. In Figure 1 we plot the 5% power as a function of $\text{var}(c_i)$ when c_i is drawn from a uniform distribution with mean $E(c_i) = -5$. The data generating process of the observed data, y_{it} say, is the simplest possible; a first-order autoregression with standard normal errors.²

Figure 1: Power as a function of $\text{var}(c_i)$ when c_i is uniformly distributed with $E(c_i) = -5$.



There are four lines representing the empirical and asymptotic power when N is 10 and when it is 50. In both cases $T = 500$. Since the asymptotic power only depends on the mean of c_i , these lines are flat. We see that when $N = 10$ the empirical power function is way off the theoretical prediction. It starts at about 10% below and then the difference just gets larger as the variance increases. Of course, this is not completely unexpected, as N is rather

¹There is also evidence suggesting that it is not only the variance of c_i that matters for power, but also higher moments, see for example Moon *et al.* (2007).

²We make 3,000 replications with the first 100 time series observations discarded in order to reduce the effect of the initial value, which is set to zero.

small in this case. However, the same pattern is observed also when $N = 50$, although the vertical distance to the asymptotic line is not as large as before. In other words, there seem to be a variance effect at work here that cannot be explained by theory, and that seems to go away, although rather slowly, as N increases.

This example suggests that the asymptotic framework commonly used for analyzing the local power of panel unit root tests may not be sharp enough to capture actual behavior. It is therefore necessary to consider alternative frameworks that are sensible also for panels in which N is moderately large, and this note can be seen as a step in this direction. The approach we take is to derive a finite- N expansion of the test statistic that retains not only the mean of c_i but also higher moments, which are shown to exert a second-order effect. This enables us to analyze what happens for heterogeneous alternatives that are local-to-unity in T but not necessarily in N . The results go a long way towards explaining the variance effect seen in Figure 1.

2 Model and assumptions

Consider the following very simple data generating process:

$$y_{it} = \rho_i y_{it-1} + \epsilon_{it}, \quad (1)$$

where $\epsilon_{it} \sim \text{iid}(0, \sigma^2)$ with $E(\epsilon_{it}^4) < \infty$ and $y_{i0} = 0$. As for the autoregressive coefficient ρ_i , the following local-to-unity model is used:

$$\rho_i = \exp\left(\frac{c_i}{N^\eta T^\delta}\right) = 1 + \frac{c_i}{N^\eta T^\delta} + o_p(1), \quad (2)$$

where the drift parameter c_i is assumed to be iid and independent ϵ_{it} . All moments of c_i exist, and in what follows it will be convenient to denote these as $\mu_m = E(c_i^m)$ for $m \geq 0$ with $\mu_0 = 1$. For simplicity, to rule out explosive outcomes, we will further assume that the support of c_i is of the form $(-\infty, 0]$.

It should be pointed out here that none of these assumptions are really necessary, and that they can be relaxed at the expense of added technical complexity, which is unnecessary given our purpose to illustrate. In fact, as we shall see, the above model is able to deliver significant insights, despite its simplicity.

The null hypothesis of a unit root can be written as $H_0 : c_i = 0$ for all i , while the alternative hypothesis of stationarity can be written as $H_1 : c_i < 0$ for at least some i . The

exact nature of the alternative depends on η and δ . If $\eta = \delta = 0$, then $\rho_i = 1 + c_i < 1$, in which case the alternative is fixed in the sense that the degree of mean reversion does not depend on N or T . However, with such a specification we only learn if the test is consistent and, if so, at what rate. Therefore, to be able to evaluate the power analytically, in this paper we consider alternatives in which ρ_i is local-to-unity as $N, T \rightarrow \infty$, which means that $\eta, \delta > 0$. However, as we will demonstrate later, the rate at which ρ_i approaches one cannot be meaningfully interpreted unless one is precise about the relative expansion rate of N and T , and we are therefore going to assume that $T = N^\theta$ with $\theta > 0$.

3 Main result

The pooled least squares t -statistic for a unit root in (1) is given by

$$t_\rho = \frac{\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=2}^T \Delta y_{it} y_{it-1}}{\hat{\sigma} \sqrt{\frac{1}{NT} \sum_{i=1}^N \sum_{t=2}^T y_{it-1}^2}} = \frac{A_{NT}}{\hat{\sigma} \sqrt{B_{NT}}},$$

where $\hat{\sigma}^2$ is any consistent estimator of σ^2 . This is the Levin and Lin (1992) test statistic in the simple case with no deterministic terms, no serial correlation and homoskedastic errors.

Consider A_{NT} , and let us introduce $\alpha = \frac{T^{1-\delta}}{N^\eta}$ such that $\rho_i = \exp(c_i / N^\eta T^\delta) = \exp(\alpha c_i / T) = 1 + \frac{\alpha c_i}{T} + o_p(1)$. This, together with the fact that $\Delta y_{it} = (\rho_i - 1)y_{it-1} + \epsilon_{it}$, suggests

$$A_{NT} = \sqrt{N} \alpha A_{1NT} + A_{2NT} + o_p(1), \quad (3)$$

where

$$A_{1NT} = \frac{1}{NT^2} \sum_{i=1}^N \sum_{t=2}^T c_i y_{it-1}^2,$$

$$A_{2NT} = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=2}^T y_{it-1} \epsilon_{it}.$$

Note that

$$\frac{1}{\sqrt{T}} y_{it} = \frac{1}{\sqrt{T}} \sum_{s=1}^t \rho_i^{t-s} \epsilon_{is} = \frac{1}{\sqrt{T}} \sum_{s=1}^t \exp\left(\frac{\alpha c_i (t-s)}{T}\right) \epsilon_{is} \rightarrow_w \sigma J_i(r)$$

as $T \rightarrow \infty$, where \rightarrow_w signifies weak convergence and

$$J_i(r) = \int_0^r \exp(\alpha c_i (r-u)) dW_i(u).$$

Moreover, by Taylor expansion around $c_i = 0$,

$$\exp(2\alpha c_i) = \sum_{j=0}^{\infty} \frac{(2\alpha c_i)^j}{j!},$$

suggesting that, since $J_i(r)|c_i \sim N(0, (\exp(2r\alpha c_i) - 1)/2\alpha c_i)$,

$$\begin{aligned} E\left(c_i \int_0^1 J_i(r)^2 dr\right) &= \int_0^1 E(c_i J_i(r)^2) dr = \int_0^1 E\left(\frac{\exp(2r\alpha c_i) - 1}{2\alpha}\right) dr \\ &= E\left(\frac{\exp(2\alpha c_i) - 2\alpha c_i - 1}{(2\alpha)^2 c_i}\right) = \sum_{j=0}^{\infty} \frac{(2\alpha)^j}{(j+2)!} \mu_{j+1} = \sum_{j=0}^{\infty} \phi_j \mu_{j+1}. \end{aligned}$$

Thus, since the conditions for the law of large numbers given in Phillips and Moon (1999, Corollary 1) are satisfied,

$$A_{1NT} = \frac{1}{NT^2} \sum_{i=1}^N \sum_{t=2}^T c_i y_{it-1}^2 \rightarrow_p \sigma^2 \sum_{j=0}^{\infty} \phi_j \mu_{j+1} \quad (4)$$

as $N, T \rightarrow \infty$, where \rightarrow_p signifies convergence in probability.

Next, consider A_{2NT} . Clearly, $E(A_{2NT}) = 0$. Therefore, since $dW_i(r)$ is independent across both i and r ,

$$\text{var}(A_{2NT}) \rightarrow \sigma^4 E\left[\left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \int_0^1 J_i(r) dW_i(r)\right)^2\right] = \sigma^4 \frac{1}{N} \sum_{i=1}^N \int_0^1 E(J_i(r)^2) dr = \sigma^4 \sum_{j=0}^{\infty} \phi_j \mu_j$$

as $T \rightarrow \infty$. The conditions for the central limit theorem of Phillips and Moon (1999, Theorem 2) are satisfied. Hence, letting \rightarrow_d denote convergence in distribution,

$$A_{2NT} \rightarrow_d N\left(0, \sigma^4 \sum_{j=0}^{\infty} \phi_j \mu_j\right) \quad (5)$$

as $N, T \rightarrow \infty$, and by another application of Corollary 1 of Phillips and Moon (1999),

$$B_{NT} = \frac{1}{NT^2} \sum_{i=1}^N \sum_{t=2}^T y_{it-1}^2 \rightarrow_p \sigma^2 \sum_{j=0}^{\infty} \phi_j \mu_j \quad (6)$$

as $N, T \rightarrow \infty$.

Putting everything together, since $\hat{\sigma}^2 = \sigma^2 + o_p(1)$,

$$t_\rho = \frac{A_{NT}}{\hat{\sigma} \sqrt{B_{NT}}} = \frac{\sqrt{N} \alpha A_{1NT} + A_{2NT}}{\hat{\sigma} \sqrt{B_{NT}}} + o_p(1), \quad (7)$$

where

$$\frac{A_{1NT}}{\hat{\sigma} \sqrt{B_{NT}}} = \frac{\sum_{j=0}^{\infty} \phi_j \mu_{j+1}}{\sqrt{\sum_{j=0}^{\infty} \phi_j \mu_j}} + o_p(1),$$

$$\frac{A_{2NT}}{\hat{\sigma} \sqrt{B_{NT}}} \rightarrow_d N(0, 1)$$

as $N, T \rightarrow \infty$.

The finite- N power function in (7) is our main result. It shows how the power depends not only on μ_1 , but also on higher moments, and also how the effect of these moments depend on α . There are two possibilities.

1. If $\alpha > 0$ for all N and T , then $t_\rho = O_p(\sqrt{N})$. We also see that in this case all the moments of c_i are going to affect power, at least as long as $N < \infty$. Note in particular how the derivative of $\sum_{j=0}^{\infty} \phi_j \mu_{j+1} / \sqrt{\sum_{j=0}^{\infty} \phi_j \mu_j}$ with respect to μ_2 is given by

$$\frac{\sum_{j=0}^{\infty} \phi_j (2\phi_1 \mu_j - \phi_2 \mu_{j+1})}{2(\sum_{j=0}^{\infty} \phi_j \mu_j)^{3/2}},$$

whose sign is indeterminate, suggesting that for a given N power is not necessarily increasing in μ_2 , which is in agreement with the simulation evidence reported in Figure 1.

2. On the other hand, if α is permitted to go to zero with N and T , then the conclusion is rather different. In fact, since now

$$\frac{\sum_{j=0}^{\infty} \phi_j \mu_{j+1}}{\sqrt{\sum_{j=0}^{\infty} \phi_j \mu_j}} = \frac{\mu_1}{\sqrt{2}} + O(\alpha),$$

this means that μ_1 is going to have a dominating effect with the effect of higher-order moments going to zero as $N \rightarrow \infty$. In particular, let us consider the combinations of η and δ such that $\sqrt{N}\alpha = O(1)$, which means that the power function will be of the same stochastic order. By direct substitution, with $T = N^\theta$,

$$\sqrt{N}\alpha = T^{1-\delta} N^{1/2-\eta} = N^{\theta(1-\delta)+1/2-\eta}.$$

One way to make this $O(1)$ is to set $\delta = 1$ and $\eta = \frac{1}{2}$, and to let θ vary freely. This implies that $\alpha = \frac{1}{\sqrt{N}}$, and therefore

$$t_\rho = \frac{\mu_1}{\sqrt{2}} + \frac{A_{2NT}}{\hat{\sigma}\sqrt{B_{NT}}} + O_p\left(\frac{1}{\sqrt{N}}\right) \rightarrow_d \frac{\mu_1}{\sqrt{2}} + N(0,1) \quad (8)$$

as $N, T \rightarrow \infty$, which is exactly the same result derived by Breitung (2000, Theorem 2) in the special case when $\rho_i = 1 + \frac{c}{\sqrt{NT}}$. Our results show that (8) actually does not require $\delta = 1$ and $\eta = \frac{1}{2}$, but holds for all combinations of δ, η and θ such that $\theta(1 - \delta) + 1/2 - \eta = 0$. However, it is important to realize that if δ and η are set

differently, then θ is no longer free to move. Suppose, for example, that $\delta = \frac{1}{2}$ and $\eta = 1$, such that now we are considering alternatives of the type $\rho_i = 1 + \frac{c_i}{N\sqrt{T}}$. This means that for (8) to hold, we need $\theta = 1$, which obviously puts a limit on the applicability of the test.³ This relationship between δ , η and θ is something that seems to have been largely overlooked in the previous literature.

4 Simulations

A small-scale simulation study was conducted to assess the accuracy of our theoretical results in small samples. The data generating process is given by (1) and (2), where we set $\sigma^2 = 1$ such that $\epsilon_{it} \sim N(0, 1)$. As in Section 1 we further assume that the drift parameter c_i is generated as $c_i \sim U(a, b)$, suggesting that $\mu_m = \frac{1}{m+1} \sum_{s=0}^m a^s b^{m-s}$. The data are generated for 3,000 panels with $T = N^\theta$, where the first 100 time series observations are disregarded to reduce the effect of the initial value, which is set to zero. All tests are carried out at the 5% significance level, and the finite- N power function in (7) is truncated such that it only includes the first 10 moments of c_i .⁴

The results are reported in Table 1. We begin by considering the case when $\eta = 1/2$ and $\delta = 1$, in which $\alpha = \frac{1}{\sqrt{N}}$, suggesting that both the finite- N and asymptotic power functions in (7) and (8), respectively, are $O_p(1)$. In agreement with Figure 1, we see that the empirical power is decreasing in $\text{var}(c_i)$, and that this effect remains even in the largest panels considered. While the asymptotic power function completely misses this, the finite- N power function is much more successful in predicting the effect of $\text{var}(c_i)$. The accuracy of the prediction is far from perfect, though, in the sense that the theoretical rejection rate is typically much lower than its empirical counterpart. Of course, this is not totally unexpected, as the power function is based on a Taylor approximation around $c_i = 0$, whose accuracy depend on how distant the alternative is.⁵ However, we also see that the discrepancy diminishes as N and T increases, and that for the largest panels with $N = 50$ and $\theta = 2$ accuracy is actually quite good. We also see that, although there is a small increase among the smaller values of

³In the more realistic setting with unit-specific intercept and trend parameters that need to be estimated, it is usual to require $\frac{N}{T} \rightarrow 0$ in order for the estimation error to become negligible. To allow for this, we may consider $\delta = \frac{7}{8}$ and $\eta = \frac{3}{4}$, which implies $\theta = 2$, giving $\frac{N}{T} = \frac{1}{N} \rightarrow 0$ as $N \rightarrow \infty$.

⁴In all experiments considered the effect of higher-order moments was negligible.

⁵See for example Moon *et al.* (2007) for a similar discussion in the context of panel unit root testing with deterministic intercept and trend terms.

N and T , power is quite flat in the sample size, which is in accordance with our expectations, since asymptotically there is no dependence on N and T .

Consider next the results for the case when $\eta = 1/4$, $\delta = 3/4$ and $\theta = 1$, suggesting that $\alpha = 1$, and therefore $t_\rho = O_p(\sqrt{N})$. In agreement with this result, we see that the empirical power is strictly increasing in N , and that this effect is well-captured by the finite- N power function. By contrast, being based on the wrong specification of the local alternative, the asymptotic power function is now way off target. To take an extreme example, consider the case when $N = 50$ and $a = b = -1$, in which the empirical power is more than five times as large as the theoretical prediction. Again, the new power function is able to correctly predict the decrease in power as $\text{var}(c_i)$ increases.

5 Conclusion

Recently, much effort has been directed towards the analysis of the local power of panel unit root tests. The main thrust of this note is that the conventional asymptotic analysis, in which both N and T are passed to infinity, can be a rather unreliable guide to what happens in practice. Of course, this observation is in itself nothing new, but has been made in several recent studies, including Moon and Perron (2008), and Moon *et al.* (2007). However, as far as we are aware, so far there has been no attempts to try to explain the observation theoretically, and this note therefore offers some new results in this direction.

Many interesting extensions of the present work are possible. Most importantly, it would be interesting to derive the finite- N power functions of other tests, and also in cases of serial correlation and nonzero deterministic constant and trend terms.

References

- Levin, A., and C. Lin (1992). Unit Root Tests in Panel Data: Asymptotic and Finite-sample Properties. UC San Diego, Working Paper 92-23.
- Moon, H. R., and B. Perron (2008). 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.
- Phillips, P. C. B., and H. R. Moon (1999). Linear Regression Limit Theory of Nonstationary Panel Data. *Econometrica* **67**, 1057-1111.

Table 1: Simulation results.

a	b	Empirical power			Asymptotic theoretical power			Finite- N theoretical power		
		$N = 10$	$N = 20$	$N = 50$	$N = 10$	$N = 20$	$N = 50$	$N = 10$	$N = 20$	$N = 50$
		$\eta = 1/2, \delta = 1, \theta = 1$								
-2	-2	56.4	71.1	76.6	40.5	42.1	42.1	31.7	33.8	36.4
-3	-1	53.3	67.6	73.8	40.5	42.1	42.1	30.5	33.1	35.9
-4	0	40.4	51.2	63.9	40.5	42.1	42.1	27.1	30.8	34.0
		$\eta = 1/2, \delta = 1, \theta = 2$								
-2	-2	52.4	43.1	39.1	41.0	41.3	41.0	32.9	33.4	35.5
-3	-1	50.3	42.2	38.7	41.0	41.3	41.0	31.7	32.2	34.9
-4	0	42.8	39.6	36.8	41.0	41.3	41.0	28.9	29.6	33.4
		$\eta = 1/4, \delta = 3/4, \theta = 1$								
-1	-1	78.1	97.5	100.0	17.3	17.9	17.9	51.5	76.7	98.1
-2	0	54.3	77.3	98.5	17.3	17.9	17.9	42.2	65.8	94.0
-2	-2	98.4	100.0	100.0	40.5	42.1	42.1	87.2	99.1	100.0
-3	-1	96.8	100.0	100.0	40.5	42.1	42.1	86.0	98.8	100.0

Notes: The data is generated as $y_{it} = \rho_i y_{it-1} + \epsilon_{it}$, where $\epsilon_{it} \sim N(0, 1)$ and $\rho_i = 1 + \frac{c_i}{N^{1/T}}$ with $c_i \sim U(a, b)$. T is determined through $T = N^\theta$. The asymptotic power function is based on letting $N, T \rightarrow \infty$, while the finite- N power function retains terms that go to zero as $N \rightarrow \infty$.