SUPPLEMENT TO “A SIMPLE TEST FOR NONSTATIONARITY IN MIXED PANELS: A FURTHER INVESTIGATION”: EXTENSIONS

Joakim Westerlund*
Deakin University
Australia
April 28, 2014

Abstract
This supplement extends the results reported in Westerlund (2014a) in several directions, including (i) nonzero initial values, (ii) cross-section dependence, (iii) heteroskedasticity, (iv) serial correlation, and (v) incidental trends.

1 Introduction

Many of the assumptions used in Westerlund (2014a) are stronger than the ones often encountered in the literature based on letting both $N$ and $T$ to infinity, but are standard when testing for a unit root in a fixed-$T$ setting (see Harris and Tzavalis, 1999; Kruiniger, 2009). In particular, the assumption that $\epsilon_{i,t}$ is independent across both $i$ and $t$ is standard, and is difficult to dispense with unless one is willing to make very specific assumptions regarding the structure for the dependence (when $T$ is fixed). For example, if the correlation structure can be assumed to be homogenous, it is also possible to allow for serial correlation in $\epsilon_{i,t}$ (see De Blander and Dhaene, 2013, for such an assumption). Of course, the situation is quite different when $N, T \to \infty$, which enables more general types of dependencies. In this supplement we therefore discuss some of the allowances that can be made. In particular, in what remains we discuss in turn nonzero initial values (Section 2), cross-section dependence

*Deakin University, Faculty of Business and Law, School of Accounting, Economics and Finance, Melbourne Burwood Campus, 221 Burwood Highway, VIC 3125, Australia. Telephone: +61 3 924 46973. Fax: +61 3 924 46283. E-mail address: j.westerlund@deakin.edu.au.
(Section 3), heteroskedasticity (Section 4), serial correlation (Section 5), and incidental trends (Section 6).

2 Nonzero initial value

Ng (2008) assumes that $u_{1,0}, ..., u_{N,0}$ are $O_p(1)$, which is obviously less restrictive than the $u_{1,0} = ... = u_{N,0} = 0$ assumption in Westerlund (2014a). However, as with many of the other assumptions, zero initialization is not really necessary. Indeed, while restrictive in a model without deterministic terms, the assumption of zero initialization is irrelevant as long as the model includes a constant. In fact, under $N, T \to \infty$, we can even allow $u_{i,0} = o_p(\sqrt{T})$ without affecting the results. The reason is the following: The proofs of Theorems 1 and 2 in Westerlund (2014a, Appendix) start with the limiting behavior of $r_{i,t} = u_{i,t}/\sigma_\epsilon \sqrt{T}$. If $u_{1,0}, ..., u_{N,0}$ are unrestricted, then this variable can be written as

$$\frac{1}{\sigma_\epsilon \sqrt{T}} r_{i,t} = \frac{1}{\sigma_\epsilon \sqrt{T}} \alpha_i' u_{i,0} + \frac{1}{\sigma_\epsilon \sqrt{T}} \sum_{s=1}^{t} \alpha_i^{t-s} \epsilon_{i,s},$$

where the first term on the right-hand side is negligible as long as $u_{i,0} = o_p(\sqrt{T})$ (see Westerlund, 2014b, for a detailed discussion of the effect of initialization when testing for a unit root in panels). Unfortunately, if $T$ is fixed, then this is no longer the case. A sufficient condition for the finite-$T$ results to go through also in the case of nonzero but $O_p(1)$ initialization is that $\sum_{i=1}^{N} E[(u_{i,0} - \bar{u}_0)(\epsilon_{i,t} - \bar{\epsilon}_t)]/N = o(1/\sqrt{N})$ (with obvious definitions of $\bar{u}_0$ and $\bar{\epsilon}_t$).

3 Cross-section dependence

It is easy to verify that $\hat{\theta}$ is invariant with respect to transformations of the type $\{y_{i,t}\}_{t=1}^{T} \mapsto \{a_t + y_{i,t}\}_{t=1}^{T}$, suggesting that cross-section dependence in the form of a common time effect can be allowed without any correction needed. To appreciate this, suppose that instead of (2) in Westerlund (2014a) we have

$$y_{i,t} = \lambda_i + \gamma_t + u_{i,t},$$

where $\gamma_t$ is a common time effect. Under this DGP,

$$y_{i,t} - \bar{y}_t = \alpha_i - \bar{\alpha} + \epsilon_{i,t} - \bar{\epsilon}_t,$$

showing that $V_t = \sum_{t=1}^{T} (y_{i,t} - \bar{y}_t)^2/N$ is indeed invariant with respect to $\gamma_t$. But while common time effects do not affect $\hat{\theta}$, the test statistics need to be modified slightly. Specifically,
the estimators of $\sigma^2_\epsilon$ and $\kappa_\epsilon$ should be based on $y_{i,t} - \bar{y}_t$ rather than on $y_{i,t}$. The appropriate versions of $\hat{\phi}^2_\epsilon$ and $\hat{\kappa}_\epsilon$ to consider in this case are therefore given by $\hat{\phi}^2_\epsilon = \frac{\sum_{i=1}^N \sum_{t=2}^T \Delta(y_{i,t} - \bar{y}_t)^2}{NT}$ and $\hat{\kappa}_\epsilon = \frac{\sum_{i=1}^N \sum_{t=2}^T \Delta(y_{i,t} - \bar{y}_t)^4}{4\hat{\phi}^4 N T}$, respectively. The asymptotic results reported in Westerlund (2014a) are unaffected by this.

Common time effects represent a limiting form of cross-section dependence. For a detailed treatment of the case with a general common factor structure we make reference to Ng (2008, Section 3.2).

4 Heteroskedasticity

Consider first the case when the heteroskedasticity is across time. Specifically, suppose that $E(\epsilon^2_{i,t}) = \sigma^2_\epsilon^{i,t}$ and $E(\epsilon^4_{i,t})/\sigma^4_\epsilon^{i,t} = \kappa_\epsilon^{i,t} < \infty$. Heteroskedasticity of this type can be permitted by simply replacing $\hat{\phi}^2_\epsilon$ and $\hat{\kappa}_\epsilon$ by $\hat{\phi}^2_{\epsilon,t} = \frac{\sum_{i=1}^N \Delta y_{i,t}^2}{N}$ and $\hat{\kappa}_{\epsilon,t} = \sum_{i=1}^N \Delta y_{i,t}^4 / (\hat{\phi}_{\epsilon,t}^4 N T)$, respectively. Since these estimates are consistent for $N \rightarrow \infty$, heteroskedasticity across time can be permitted even when $T$ is fixed. In Westerlund (2014a) $\hat{\phi}^* = \hat{\phi}^2_\epsilon$, which is equivalent to using $\hat{\phi}$ while replacing $y_{i,t}$ by $y_{i,t}' = y_{i,t} / \hat{\phi}_{\epsilon,t}$. The appropriate estimator to consider under heteroskedasticity is the same but with $y_{i,t}$ replaced by $y_{i,t}' = y_{i,t} / \hat{\phi}_{\epsilon,t}$. The $\tau_{1,T}^*$ statistic is the same as under homoskedasticity, but with $\hat{\phi}^2_{\tilde{\theta},T}$ replaced by

$$\hat{\phi}^2_{\tilde{\theta},T} = \frac{2(T^2 - 1)}{T^2} + \frac{1}{T^2} \sum_{i=2}^T (\hat{\kappa}_{\epsilon,i} - 3).$$

If the heteroskedasticity is across the cross-section, then $T$ cannot be finite anymore. Assume that $E(\epsilon^2_{i,t}) = \sigma^2_\epsilon^{i,t}$ and $E(\epsilon^4_{i,t})/\sigma^4_\epsilon^{i,t} = \kappa_\epsilon^{i,t} < \infty$. Valid inference in this case requires replacing $\hat{\phi}^2_\epsilon$ in $\tau_{\theta_0}^*$ (and $\tau_{1}^*$) by $\hat{\phi}^2_{\epsilon,t} = \frac{\sum_{i=1}^N \Delta y_{i,t}^2}{T}$. The obvious rationale for this practice is that $\hat{\phi}_{\epsilon,t} = \sigma_{\epsilon,t}^2 + o_p(1)$ as $T \rightarrow \infty$ (see also Ng, 2008, Section 3.1). The appropriate estimator of $\theta$ to use in this case is $\hat{\theta}^*$ with $\hat{\phi}^2_\epsilon = 1$ and $y_{i,t}$ in $\hat{\theta}$ replaced by $y_{i,t}' = y_{i,t} / \hat{\phi}_{\epsilon,t}$. Given this change, the $\tau_{1}^*$ statistic is the same as under homoskedasticity. In Section 5 we report some Monte Carlo evidence for the case when the errors are both cross-section heteroskedastic and serially correlated.
5 Error serial correlation

In Westerlund (2014a) it is assumed that the error driving $u_{i,t}$ is serially uncorrelated. This is not necessary. Suppose therefore that instead of (3) in Westerlund (2014a), we have

$$u_{i,t} = \alpha_i u_{i,t-1} + e_{i,t},$$

(2)

$$e_{i,t} = \phi_i(L)e_{i,t},$$

(3)

where $\phi_i(L) = \sum_{n=0}^{\infty} \phi_i n L^n$ has all its roots outside the unit circle, and $e_{i,t}$ is iid with $E(e_{i,t}) = 0$, $E(e_{i,t}^2) = \sigma_{e,i}^2$ and $E(e_{i,t}^4)/\sigma_{e,i}^4 = \kappa_{e,i} < \infty$. In this DGP the “long-run variance” of $e_{i,t}$ is given by $\omega^2_{e,i} = \phi_i(1)^2 \sigma_{e,i}^2$.

In order to illustrate the effect of the above DGP, we revisit Proof of Theorem 1. As pointed out in Section 4, with cross-section heteroskedasticity the appropriate version of $\hat{\theta}^*$ to use is $\hat{\theta}$ with $y_{i,t}^* = y_{i,t}/\omega_{e,i}$. The appropriate estimator in the current case is the same but with $\hat{\sigma}_{e,i}^2$ replaced by $\hat{\omega}_{e,i}^2$, which can be any consistent estimator of $\omega_{e,i}^2$. Later in this section we provide an example of how $\hat{\omega}_{e,i}^2$ may be constructed. However, for now we assume that $\omega_{e,i}^2$ is known, such that $y_{i,t}^* = y_{i,t}/\omega_{e,i}$. Let us therefore define $a_{i,t}^* = a_{i,t}/\omega_{e,i}$ for any variable $a_{i,t}$. Redefine $V_t = \sum_{i=1}^{N} (y_{i,t}^* - \bar{y}_t^*)^2/N, r_{i,t} = u_{i,t}/\omega_{e,i} \sqrt{T}$ and $U_t = V_t/T$. From

$$y_{i,t}^* - \bar{y}_t^* = (\lambda_t^* - \bar{\lambda}_t^*) + (u_{i,t}^* - \bar{u}_t^*),$$

$$U_t = A_t + B_t + C_t,$$

(4)

where

$$A_t = 1/T \bar{\lambda}_t^2, N,$$

$$B_t = 1/N \sum_{i=1}^{N} (r_{i,t} - \bar{r}_t)^2,$$

$$C_t = 2 \sqrt{T}N \sum_{i=1}^{N} (\lambda_t^* - \bar{\lambda}_t^*)(r_{i,t} - \bar{r}_t).$$

Therefore, since $\Delta A_t = 0$, we obtain

$$\Delta U_t = \Delta B_t + \Delta C_t,$$

(5)

suggesting that

$$\hat{\theta} = \frac{1}{T} \sum_{t=2}^{T} \Delta V_t = \sum_{t=2}^{T} \Delta U_t = \sum_{t=2}^{T} (\Delta B_t + \Delta C_t).$$

(6)
Consider $B_t$. Clearly,

$$
\Delta B_t = \frac{1}{N} \sum_{i=1}^{N} \Delta r_{it}^2 - \Delta r_{t}^2,
$$

According to the Beveridge–Nelson decomposition, $\phi_t(L) = \phi_t(1) + \phi_t^*(L)(1 - L)$, where $\phi_t^*(L) = \sum_{n=0}^{\infty} \phi_{t,n}^* L^n$ with $\phi_{t,n}^* = \sum_{k=n+1}^{\infty} \phi_{t,k}$ has all roots outside the unit circle. Hence, letting $\epsilon_{it}^* = \phi_t^*(L)e_{i,t}$, $\epsilon_{it} = \phi_t(1)e_{i,t} + \Delta \epsilon_{it}^*$. Let $\epsilon_{it}^* = \alpha_i^{-}\epsilon_{it}$, such that

$$r_{it} = \frac{1}{\omega_{e,i}\sqrt{T}} u_{it} = \frac{1}{\omega_{e,i}\sqrt{T}} \sum_{s=1}^{t} \alpha_i^{t-s} \epsilon_{is},
$$

$$= \frac{1}{\sigma_{e,i}\sqrt{T}} \sum_{s=1}^{t} \alpha_i^{t-s} \epsilon_{is} - \frac{1}{\sigma_{e,i}\sqrt{T}} \sum_{s=1}^{t} \Delta \epsilon_{is}^* = \frac{1}{\sigma_{e,i}\sqrt{T}} \sum_{s=1}^{t} \alpha_i^{t-s} \epsilon_{is} - \frac{1}{\sigma_{e,i}\sqrt{T}} \sum_{s=1}^{t} \alpha_i^{t-s} \epsilon_{is}^*.
$$

Hence, for $t \geq s$,

$$E(r_{it}r_{is}|c_i) = \frac{1}{\sigma_{e,i}^2 T} E \left[ \left( \sum_{n=1}^{t} \alpha_i^{t-n} \epsilon_{in} - \epsilon_{is}^* \right) \left( \sum_{m=1}^{s} \alpha_i^{s-m} \epsilon_{im} - \epsilon_{is}^* \right) \big| c_i \right]
$$

$$= \frac{1}{\sigma_{e,i}^2 T} \sum_{n=1}^{t} \sum_{m=1}^{s} \alpha_i^{t+s-n-m} E(\epsilon_{i,n} \epsilon_{i,m}) - \frac{1}{\sigma_{e,i}^2 T} \sum_{m=1}^{s} \alpha_i^{s-m} E(\epsilon_{i,m} \epsilon_{i,s}^*)
$$

$$- \frac{1}{\sigma_{e,i}^2 T} \sum_{n=1}^{t} \alpha_i^{t-n} E(\epsilon_{i,n} \epsilon_{i,s}^*) + \frac{1}{\sigma_{e,i}^2 T} E(\epsilon_{i,s}^* \epsilon_{i,s}^*),
$$

where

$$\frac{1}{\sigma_{e,i}^2 T} \sum_{m=1}^{s} \alpha_i^{s-m} E(\epsilon_{i,m} \epsilon_{i,t}) = \frac{1}{\sigma_{e,i}^2 T} \sum_{m=1}^{s} \phi_{t,n}^* \alpha_i^{s-n} E(\epsilon_{i,m} \epsilon_{i,t-n})
$$

$$= \frac{1}{\sigma_{e,i}^2 T} \sum_{n=1}^{s} \phi_{t,n}^* \alpha_i^{s-n} E(\epsilon_{i,n}^2) = \frac{1}{T} \sum_{n=1}^{s} \phi_{t,n}^* \alpha_i^{s-n} = O \left( \frac{1}{T} \right),
$$

which holds, because $\phi_{t,n}^*$ is summable. We can similarly show that $E(\epsilon_{i,s}^* \epsilon_{i,s}^*) = O(1)$, giving

$$E(r_{it}r_{is}|c_i) = \frac{1}{\sigma_{e,i}^2 T} \sum_{n=1}^{t} \sum_{m=1}^{s} \alpha_i^{t+s-n-m} E(\epsilon_{i,n} \epsilon_{i,m}) + O \left( \frac{1}{T} \right).
$$

(7)

All results used for evaluating $\sum_{t=2}^{T} u_{it}$, which are based on $E(r_{it}r_{is}|c_i)$, are therefore the same as in Proof of Theorem 1, up to a $O_p(1/T)$ remainder term. Hence,

$$E(S_T) = \sum_{t=2}^{T} \sqrt{N}E(u_{it}) = O \left( \sqrt{\frac{N}{T}} \right).
$$

It follows that $S_T$ is no longer exactly mean zero, but only asymptotically so, provided that $\sqrt{N}/T \to 0$. Hence, by a central limit theorem, letting $\sigma_{\theta,N,T}^2 = E(S_T^2)$,

$$S_T \sim \sqrt{N}\sigma_{\theta,N,T}N(0,1)
$$

(8)
as $N, T \to \infty$ with $\sqrt{N}/T \to 0$.

Of course, in practice $\omega^2_{\epsilon,t}$ is unknown and we therefore require a consistent estimator to be used in its place. Such an estimator can be constructed by using known results from the literature on heteroskedasticity and autocorrelation consistent (HAC) variance estimators (see, for example, Andrews, 1991). If $\eta > 0$, then we may use

$$\hat{\sigma}^2_{\epsilon,t} = \hat{\gamma}(0) + 2 \sum_{n=1}^{M-1} K(n/M)\hat{\gamma}(n),$$

where $\hat{\gamma}(n) = \sum_{t=0}^{T} \Delta y_{i,t}\Delta y_{i,t-n}/T$, $K(n/M)$ is a kernel function and $M$ is a bandwidth truncation parameter. As when allowing for heteroskedasticity across the cross-section, the presence of error serial correlation requires letting both $N$ and $T$ to infinity. As already mentioned, the appropriate version of $\hat{\theta}^*$ to use is $\hat{\theta}$ with $y_{i,t}$ replaced by $y^*_i = y_{i,t}/\hat{\omega}_{\epsilon,t}$. The appropriate variance correction factor to put in the denominator of $\tau^*_1$ depends on the formula for $\sigma^2_{\theta,NT}$. In the simple DGP considered in Westerlund (2014a), the appropriate formula under $T \to \infty$ simplifies to $\sigma^2_{\theta,NT} = 2 + o_p(1)$. Unfortunately, evaluating this term in the presence of serial correlation is extremely tedious. Preliminary calculations suggest that the effect of the serial correlation is negligible and therefore that $\sigma^2_{\theta,NT}$ has the same simple form as before, which is supported by unreported simulation results. However, in this supplement we follow Ng (2008, Section 3.2) and replace $\sigma^2_{\theta,NT}$ by a direct estimator, which is obtained as the HAC estimator based on $\sqrt{N/T}(\Delta V_t - 1)$. Hence, in terms of the above formula for $\hat{\sigma}^2_{\epsilon,t}$, we simply replace $\Delta y_{i,t}$ by $\sqrt{N/T}(\Delta V_t - 1)$.

As an illustration we revisit the DGP used for generating Table 4 in Westerlund (2014a). The only difference is that now $u_{i,t} = u_{i,t-1} + e_{i,t}$, where $e_{i,t} = \phi_t(L)e_{i,t}, e_{i,t} \sim N(0,1)$, $\phi_t(L) = \phi_{t,0} + \phi_{t,1}L$ and $\phi_{t,0} = 1$ and $\phi_{t,1} = 0.5$. Two different versions of $\tau^*_1$ are considered. The first, denoted $\tau^*_{1,ac}$, is the one described in the above with $K(x) = (1 - |x|)1(|x| \leq 1)$ (the Bartlett kernel) and $M = \lfloor T^{1/3} \rfloor$, where $1(A)$ is the indicator function for the event $A$ and $\lfloor x \rfloor$ is the integer part of $x$. The second version, $\tau^*_1$, is the same as in Westerlund (2014a), which is constructed under the assumption of no serial correlation. The results based on 3,000 replications are reported in Table 1. As expected, we see that while $\tau^*_{1,ac}$ is correctly sized, $\tau^*_1$ is severely undersized in all experiments considered. The reason for this difference can be seen by looking at the mean and variance results, which in case of $\tau^*_1$ are way off the theoretical predictions under no serial correlation.

While $\tau^*_{1,ac}$ is not identical to the test statistic considered by Ng (2008, Section 3), they
should be asymptotically equivalent. In spite of this, unreported simulation evidence suggests that the test of Ng (2008) generally suffers from severe size distortions. Hence, just as in the case without serial correlation, the test statistic considered here leads to improved small-sample performance.

Table 1: 5% size, mean and variance of \( \tau_{1}^{*} \) and \( \tau_{1,ac}^{*} \).

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>( \tau_{1}^{*} )</th>
<th>( \tau_{1,ac}^{*} )</th>
<th>( \tau_{1}^{*} )</th>
<th>( \tau_{1,ac}^{*} )</th>
<th>( \tau_{1}^{*} )</th>
<th>( \tau_{1,ac}^{*} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>25</td>
<td>0.30</td>
<td>4.60</td>
<td>1.65</td>
<td>−0.10</td>
<td>3.15</td>
<td>0.85</td>
</tr>
<tr>
<td>10</td>
<td>50</td>
<td>0.20</td>
<td>5.90</td>
<td>1.70</td>
<td>−0.17</td>
<td>3.38</td>
<td>0.92</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>0.20</td>
<td>7.70</td>
<td>1.75</td>
<td>−0.20</td>
<td>3.69</td>
<td>1.01</td>
</tr>
<tr>
<td>10</td>
<td>200</td>
<td>0.40</td>
<td>8.50</td>
<td>1.73</td>
<td>−0.25</td>
<td>3.66</td>
<td>1.08</td>
</tr>
<tr>
<td>20</td>
<td>25</td>
<td>0.10</td>
<td>2.50</td>
<td>2.68</td>
<td>0.11</td>
<td>3.58</td>
<td>0.83</td>
</tr>
<tr>
<td>20</td>
<td>50</td>
<td>0.00</td>
<td>3.90</td>
<td>2.74</td>
<td>0.03</td>
<td>3.59</td>
<td>0.89</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>0.00</td>
<td>5.60</td>
<td>2.68</td>
<td>−0.07</td>
<td>3.54</td>
<td>0.93</td>
</tr>
<tr>
<td>20</td>
<td>200</td>
<td>0.00</td>
<td>6.30</td>
<td>2.70</td>
<td>−0.10</td>
<td>3.70</td>
<td>0.99</td>
</tr>
<tr>
<td>40</td>
<td>25</td>
<td>0.00</td>
<td>1.10</td>
<td>3.93</td>
<td>0.28</td>
<td>3.57</td>
<td>0.79</td>
</tr>
<tr>
<td>40</td>
<td>50</td>
<td>0.00</td>
<td>2.80</td>
<td>3.96</td>
<td>0.14</td>
<td>3.70</td>
<td>0.87</td>
</tr>
<tr>
<td>40</td>
<td>100</td>
<td>0.00</td>
<td>3.30</td>
<td>4.04</td>
<td>0.10</td>
<td>3.56</td>
<td>0.89</td>
</tr>
<tr>
<td>40</td>
<td>200</td>
<td>0.00</td>
<td>3.30</td>
<td>4.04</td>
<td>0.07</td>
<td>3.56</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Notes: \( \tau_{1,ac}^{*} \) refers to the serial correlation robust version of \( \tau_{1}^{*} \).

See Westerlund (2014a) for a description of the DGP.

6 Incidental trends

The DGP in Westerlund (2014a) excludes the possibility that \( y_{i,t} \) is trending deterministically, thereby excluding the so-called “incidental trends” case (see, for example, Moon et al., 2007). While in principle there is nothing that prevents a generalization of the test approach considered here to accommodate such trends, preliminary calculations suggest that the proofs in the finite-\( T \) case would be extremely tedious, if not impossible. In Westerlund (2014a) we therefore only consider the intercept-only case. Fortunately, the required calculations simplify under \( N, T \to \infty \). Suppose therefore that instead of (2) in Westerlund (2014a) we have

\[
y_{i,t} = \lambda_i + \beta_i t + u_{i,t}. \tag{9}
\]
In what follows we provide the asymptotic distribution (as \(N, T \to \infty\)) of \(\sqrt{N/T}\hat{\theta}\) under this specification. In so doing we will assume that \(\eta > 0\), which means that we will basically derive a constant and trend version of Theorem 1. Since many of the results can be taken directly from Proof of Theorem 1 in Westerlund (2014a, Appendix), only essential details are given.

We begin by defining \(r_{i,t} = u_{i,t}/\sigma_e\sqrt{T}, \sigma^2_{\lambda,N} = \sum_{i=1}^{N}(\lambda_i - \bar{\lambda})^2/N, \sigma^2_{\beta,N} = \sum_{i=1}^{N}(\beta_i - \bar{\beta})^2/N\) and \(U_t = V_t/\sigma^2_T\). Substitution of \(y_{i,t} - \bar{y}_t = (\lambda_i - \bar{\lambda}) + (\beta_i - \bar{\beta})t + (u_{i,t} - \bar{u}_t)\) into \(U_t\) yields

\[
U_t = A_t + B_t + C_t + D_t + E_t + F_t, \quad (10)
\]

where

\[
\begin{align*}
A_t &= \frac{1}{\sigma^2_e T} \sigma^2_{\lambda,N}, \\
B_t &= \frac{1}{N} \sum_{i=1}^{N} (r_{i,t} - \bar{r}_t)^2, \\
C_t &= \frac{2}{\sigma_e \sqrt{T} N} \sum_{i=1}^{N} (\lambda_i - \bar{\lambda}) (r_{i,t} - \bar{r}_t), \\
D_t &= \frac{1}{\sigma^2_e T} \sigma^2_{\beta,N} t^2, \\
E_t &= \frac{2}{\sigma_e \sqrt{T} N} \sum_{i=1}^{N} (\beta_i - \bar{\beta}) (r_{i,t} - \bar{r}_t) t, \\
F_t &= \frac{2}{\sigma_e N T} \sum_{i=1}^{N} (\lambda_i - \bar{\lambda}) (\beta_i - \bar{\beta}) t,
\end{align*}
\]

suggesting

\[
\Delta U_t = \Delta B_t + \Delta C_t + \Delta D_t + \Delta E_t + \Delta F_t. \quad (11)
\]

Here \(\Delta B_t\) and \(\Delta C_t\) are as in Proof of Theorem 1 and

\[
\begin{align*}
\Delta D_t &= \frac{1}{\sigma^2_e T} \sigma^2_{\beta,N} \Delta t^2, \\
\Delta E_t &= \frac{2}{\sigma_e \sqrt{T} N} \sum_{i=1}^{N} (\beta_i - \bar{\beta}) \Delta [(r_{i,t} - \bar{r}_t)t], \\
\Delta F_t &= \frac{2}{\sigma_e N T} \sum_{i=1}^{N} (\lambda_i - \bar{\lambda}) (\beta_i - \bar{\beta}).
\end{align*}
\]

Consider \(\Delta D_t\). Since \(\Delta t^2 = t^2 - (t - 1)^2 = 2t - 1\), we may write

\[
\Delta D_t = \frac{1}{\sigma^2_e T} \sigma^2_{\beta,N} (2t - 1) = \mu_{\Delta D,NT}(t) + o(1)
\]
where

$$\mu_{\Delta B, NT}(t) = \frac{2}{\sigma_t^2 T} \sigma_{\beta, N}^2 t.$$

Making use of the results provided for $\Delta C_i$ in Proof of Theorem 1, it is clear that $\Delta E_i$ has zero mean. For the variance, we use $\Delta[(r_{i,t} - \bar{r}_t)t] = (r_{i,t} - \bar{r}_t)t - (r_{i,t-1} - \bar{r}_{t-1})(t - 1) = [\Delta(r_{i,t} - \bar{r}_t)]t + (r_{i,t-1} - \bar{r}_{t-1})$, giving $\Delta[(r_{i,t} - \bar{r}_t)t]^2 = [\Delta(r_{i,t} - \bar{r}_t)]t^2 + 2(r_{i,t-1} - \bar{r}_{t-1})[\Delta(r_{i,t} - \bar{r}_t)]t + (r_{i,t-1} - \bar{r}_{t-1})^2$, and so

$$NE[(\Delta E_i)^2] = \frac{4}{\sigma_t^2 NT} \sum_{i=1}^{N} E[(\beta_i - \bar{\beta})^2(\Delta[(r_{i,t} - \bar{r}_t)t]^2)]
= \frac{4}{\sigma_t^2 NT} \sum_{i=1}^{N} E[(\beta_i - \bar{\beta})^2]E[(\Delta r_{i,t} - \Delta \bar{r}_t)^2]t^2
+ \frac{8}{\sigma_t^2 NT} \sum_{i=1}^{N} E[(\beta_i - \bar{\beta})^2]E[(r_{i,t-1} - \bar{r}_{t-1})\Delta(r_{i,t} - \bar{r}_t)]t
+ \frac{4}{\sigma_t^2 NT} \sum_{i=1}^{N} E[(\beta_i - \bar{\beta})^2]E[(r_{i,t-1} - \bar{r}_{t-1})^2].$$

The first term on the right is $(t/T)^2$ times $NE[(\Delta C_i)^2]$ with $\sigma_{\Delta, N}^2$ replaced by $\sigma_{\beta, N}^2$. As for the third term, by cross-section independence,

$$E[(r_{i,t-1} - \bar{r}_{t-1})^2] = E(r_{i,t-1}^2 - 2r_{i,t-1}\bar{r}_{t-1} + \bar{r}_{t-1}^2) = E(r_{i,t-1}^2) + o(1)
= \rho_{r, NT}(t - 1, t - 1) + o(1).$$

where $\rho_{r, NT}(s, t) \sim 1/T$, as in Proof of Theorem 1. Hence, $E[(r_{i,t-1} - \bar{r}_{t-1})^2] = O(1)$, suggesting that the third term the expression for $NE[(\Delta E_i)^2]$ is $O(1/T)$. The same is true for the second term. Hence,

$$NE[(\Delta E_i)^2] = \frac{4}{\sigma_t^2 NT} \sum_{i=1}^{N} E[(\beta_i - \bar{\beta})^2]E[(\Delta r_{i,t} - \Delta \bar{r}_t)^2]t^2 + o(1) \sim \frac{4}{\sigma_t^2 T^2} \sigma_{\beta, N}^2 t^2,$$

which we can use to show that

$$\sqrt{N}\Delta E_i \sim \sqrt{NE[(\Delta E_i)^2]} N(0, 1). \quad (12)$$

For $\Delta F_i$,

$$\Delta F_i = \frac{2}{\sigma_t^2 TN} \sum_{i=1}^{N} (\lambda_i - \bar{\lambda})(\beta_i - \bar{\beta}) = \frac{2}{\sigma_t^2 T} \sigma_{\lambda, N} = o(1),$$

where $\sigma_{\lambda, N} = \sum_{i=1}^{N}(\lambda_i - \bar{\lambda})(\beta_i - \bar{\beta})/N$. Let $u_i = \Delta U_i - \mu_{\Delta B, NT}(t) - \mu_{\Delta D, NT}(t)$. The above results, together with the known orders of $\Delta B_i$ and $\Delta C_i$ (see Proof of Theorem 1), imply

$$\sqrt{N}u_i = \sqrt{N}[(\Delta B_i - \mu_{\Delta B, NT}(t)) + \Delta C_i + (\Delta D_i - \mu_{\Delta D, NT}(t)) + \Delta E_i + \Delta F_i]
\sim \sqrt{N}[(\Delta B_i - \mu_{\Delta B, NT}(t)) + \Delta C_i + \Delta E_i]. \quad (13)$$
Consider \( S_T = \sum_{t=2}^{T} \sqrt{N}u_t \), which in view of the above result for \( \sqrt{N}u_t \) may be written as follows:

\[
S_T \sim \sum_{t=2}^{T} \sqrt{N}[(\Delta B_t - \mu_{AB,NT}(t)) + \Delta C_t + \Delta E_t]
\]

From Proof of Theorem 1 we know that \( \sum_{t=2}^{T} \sqrt{N}(\Delta B_t - \mu_{AB,NT}(t)) \) and \( \sum_{t=2}^{T} \sqrt{N}\Delta C_t \) are both \( O_p(1) \). However, by using the above result for \( NE[(\Delta E_t)^2] \) it is not difficult to show that

\[
NE \left[ \left( \sum_{t=2}^{T} \Delta E_t \right)^2 \right] \sim T \frac{4}{\sigma^2_{\epsilon}} \sigma_{\beta,N}^2 T \left[ \sum_{t=2}^{T} t^2 \sim T \frac{4}{3\sigma^2_{\epsilon}} \sigma_{\beta,N}^2 \right]
\]

where the last result holds because \( \sum_{t=2}^{T} t^2 / T^3 \rightarrow \int_0^1 r^2 dr = 1/3 \). It follows that \( \sum_{t=2}^{T} \sqrt{N}\Delta E_t \sim O_p(\sqrt{T}) \), and therefore

\[
\frac{1}{\sqrt{T}}S_T \sim \sum_{t=2}^{T} \frac{\sqrt{N}}{\sqrt{T}}[(\Delta B_t - \mu_{AB,NT}(t)) + \Delta C_t + \Delta E_t] \sim \sum_{t=2}^{T} \frac{\sqrt{N}}{\sqrt{T}} \Delta E_t.
\]

We can therefore show that

\[
\frac{1}{\sqrt{T}}S_T \sim \frac{2}{\sqrt{3\sigma_{\epsilon}}} \sigma_{\beta,N} N(0,1).
\]

Since \( \sum_{t=2}^{T} t / T^2 \rightarrow \int_0^1 r dr = 1/2 \), we have

\[
\sum_{t=2}^{T} \frac{\sqrt{N}}{\sqrt{T}} \mu_{\Delta D,NT}(t) = \sqrt{NT} \frac{2}{\sigma_{\epsilon}^2} \sigma_{\beta,N}^2 T \left[ \sum_{t=2}^{T} t \sim \sqrt{NT} \frac{1}{\sigma_{\epsilon}^2} \sigma_{\beta,N}^2 \right],
\]

making use of this result and the fact that \( \mu_{AB,NT}(t) \sim \Delta \rho_{\epsilon,NT}(t) \) (see Proof of Theorem 1), the above result for \( S_T / \sqrt{T} \) can be rearranged as follows:

\[
\frac{1}{\sqrt{T}} \sum_{t=2}^{T} \Delta U_t \sim \sum_{t=2}^{T} \frac{\sqrt{N}}{\sqrt{T}} \left[ \mu_{AB,NT}(t) + \mu_{\Delta D,NT}(t) \right] + \frac{2}{\sqrt{3\sigma_{\epsilon}}} \sigma_{\beta,N} N(0,1)
\]

\[
\sim \sum_{t=2}^{T} \frac{\sqrt{N}}{\sqrt{T}} \Delta \rho_{\epsilon,NT}(t) + \sqrt{NT} \frac{1}{\sigma_{\epsilon}^2} \sigma_{\beta,N}^2 + \frac{2}{\sqrt{3\sigma_{\epsilon}}} \sigma_{\beta,N} N(0,1),
\]

or

\[
\frac{\sqrt{N}}{\sqrt{T}} \hat{\theta} \sim \sigma_{\epsilon}^2 \left[ \sum_{t=2}^{T} \frac{\sqrt{N}}{\sqrt{T}} \Delta \rho_{\epsilon,NT}(t) + \sqrt{NT} \sigma_{\beta,N}^2 + \frac{2\sigma_{\epsilon} \sigma_{\beta,N}}{\sqrt{3}} N(0,1) \right]. \tag{14}
\]

Inference based on this result requires an estimator of \( \sigma_{\beta,N}^2 \). Towards this end, note that the above result for \( \sqrt{N}u_t \) can be solved for \( \sqrt{N}[\Delta V_t / T - \sigma_{\epsilon}^2 \mu_{AB,NT}(t)] \), giving

\[
\sqrt{N} \left( \frac{1}{T} \Delta V_t - \sigma_{\epsilon}^2 \mu_{AB,NT}(t) \right) \sim \frac{2\sqrt{N}}{T} \sigma_{\beta,N}^2 t + \sigma_{\epsilon}^2 \sqrt{N}[\Delta B_t - \mu_{AB,NT}(t)] + \Delta C_t + \Delta E_t
\]

\[
= b2 \frac{t}{T} + \epsilon_t, \tag{15}
\]
where \( e_t = \sigma^2 \sqrt{N} ([D_{B_t} - \mu_{\Delta B, NT}(t)] + \Delta C_t + \Delta E_t) \) is normal with mean zero and variance \( \text{NE}([\Delta E_t]^2) \sim 4 \sigma^2_{\hat{\beta}, N} T/(2 \sigma^2_{\theta}) \) (this can be shown using the results provided in Proof of Theorem 1), \( \mu_{\Delta B, NT}(t) \sim \Delta \rho_{r, NT}(t) \sim 1/T \) and \( b = \sqrt{N} \sigma_{\hat{\theta}, N} \). This suggests that \( \sigma^2_{\hat{\theta}, N} \) can be estimated using \( \sigma^2_{\hat{\beta}, N} = \hat{b} / \sqrt{N} \), where \( \hat{b} \) is the least squares slope in a regression of \( \sqrt{N} [\Delta V_t - \sigma^2_e] / T \) onto \( 2t/T \). Clearly,

\[
\sqrt{T} (\hat{b} - b) = \sqrt{T} \frac{\sum_{t=2}^{T} (t/T) e_t}{2 \sum_{t=2}^{T} (t/T)^2} \sim \left( \frac{1}{\sigma^2_e} \frac{1}{\sum_{t=2}^{T} (t/T)^2} \right)^{1/2} \text{N}(0,1) \sim \frac{3 \sigma_{\hat{\theta}, N}}{\sqrt{5} \sigma_e} \text{N}(0,1). (16)
\]

Since \( \sqrt{T} (\hat{b} - b) = \sqrt{NT} (\hat{\sigma}_e^2 - \sigma^2_{\hat{\theta}, N}) \), this implies

\[
\sqrt{NT} (\hat{\sigma}_e^2 - \sigma^2_{\hat{\theta}, N}) \sim \frac{3 \sigma_{\hat{\theta}, N}}{\sqrt{5} \sigma_e} \text{N}(0,1). (17)
\]

This distribution is independent of that of the normal variate in \( \hat{\theta} / \sqrt{T} \). Hence,

\[
\frac{\sqrt{N}}{\sqrt{T}} \hat{\theta} - \sqrt{NT} \hat{\sigma}^2_{\hat{\beta}, N} \sim \sigma^2_e \frac{\sum_{t=2}^{T} \sqrt{N} \Delta \rho_{r, NT}(t)}{\sqrt{T}} + \sqrt{NT} (\sigma^2_{\hat{\theta}, N} - \hat{\sigma}^2_e) + \frac{2 \sigma_e \sigma_{\hat{\beta}, N}}{\sqrt{3}} \text{N}(0,1)
\]

\[
\sim \sigma^2_e \frac{\sum_{t=2}^{T} \sqrt{N} \Delta \rho_{r, NT}(t)}{\sqrt{T}} + \sigma^2_e \sigma_0 \text{N}(0,1), \tag{18}
\]

where

\[
\sigma^2_e = \frac{(27 + 20 \sigma^4_e) \sigma^2_{\hat{\beta}, N}}{15 \sigma^2_e}.
\]

The appropriate test statistic to consider is therefore given by

\[
\tau^*_{i,t} = \frac{\sqrt{NT} (\hat{\theta}^* - (T - 1)/T - T \hat{\sigma}^2_{\hat{\beta}, N} / \hat{\sigma}^2_e)}{\hat{\sigma}_0} = \frac{\sqrt{NT} (\hat{\theta}^*_{BA,1,t} - 1)}{\hat{\sigma}_0},
\]

where \( \hat{\theta}^*_{BA,1,t} = \hat{\theta}^* + 1/T - T \hat{\sigma}^2_{\hat{\beta}, N} / \hat{\sigma}^2_e = \hat{\theta}^*_{BA,1} - T \hat{\sigma}^2_{\hat{\beta}, N} / \hat{\sigma}^2_e \), and \( \hat{\sigma}_0^2 \) is \( \sigma^2_e \) with \( \hat{\sigma}^2_e \) and \( \hat{\sigma}^2_{\hat{\beta}, N} \) replaced by \( \hat{\sigma}^2_e \) and \( \hat{\sigma}^2_{\hat{\beta}, N} \), respectively. The asymptotic distribution of this test statistic is easily seen to be

\[
\tau^*_{i,t} \sim \frac{1}{\hat{\sigma}_0} \sum_{t=2}^{T} \frac{\sqrt{N}}{\sqrt{T}} \left( \Delta \rho_{r, NT}(t) - \frac{1}{T} \right) + \text{N}(0,1), \tag{19}
\]

which reduces to \( \text{N}(0,1) \) under \( H_0 : \theta = 1 \).

It is interesting to consider the local power of \( \tau^*_{i,t} \), which is driven by the first (drift) term in the asymptotic distribution. By using the same arguments as in Westerlund (2014a), with \( T = N^\gamma \) and \( \alpha_i = \exp(c_i / N^\gamma) \),

\[
\sum_{t=2}^{T} \frac{\sqrt{N}}{\sqrt{T}} \left( \Delta \rho_{r, NT}(t) - \frac{1}{T} \right) = O((1 - \theta) N^{1/2 + \gamma/2 - \eta} \mu_{0,1}), \tag{20}
\]
which is less than $O((1 - \theta)N^{1/2+\gamma-\eta}\mu_{0,1})$, the corresponding expression in the constant-only case. This means that while the appropriate condition for non-negligible and non-increasing power is different than in the constant-only case. Specifically, while in the constant-only case the condition is $\gamma - \eta + 1/2 = 0$, in the current trend case the relevant condition is $\gamma/2 - \eta + 1/2 = 0$. Thus, if $\gamma = 1/4$, such that $T = N^{1/4}$, while $\tau^*_1$ has non-negligible power for $\eta = 3/4$, this is not the case for $\tau^*_{1,\mu}$, which has negligible power for all $\eta > 5/8$ (including $\eta = 3/4$). Hence, in agreement with the results of Moon et al. (2007), we find that the test statistic considered here has reduced power in the presence of incidental trends. This is illustrated in Table 2, which report the local power of $\tau^*_{1,\mu}$ for some constellations of $(\gamma, \eta)$ when $c_i \sim U(a, b)$. The DGP is the same as in Section 5, except that the equation for $y_{it}$ now includes a linear trend with coefficient $\beta_i \sim N(1, 0.5)$. Preliminary results suggest that the performance of the test is improved by the inclusion of a constant in (15). The results reported in Table 2 are therefore based on this regression. As expected, we see that power is very low and that it only rarely raises above the nominal 5% level. In order to get a feeling for the loss of power when compared to the constant-only case we look at Table 6 in Westerlund (2014a), which contain some power results for $\tau^*_1$. Clearly, for the combinations of $(a, b)$ considered here the loss of power is quite dramatic, which is in agreement with our theoretical prediction.

As mentioned in Westerlund (2014a), Ng (2008) only examines the asymptotic behavior of her test statistic under the null hypothesis. The finding that $\tau^*_{1,\mu}$ is subject to the so-called “incidental trends problem” is therefore a contribution relative to Ng (2008).\footnote{Moon and Phillips (1999) show that the maximum likelihood estimator of the local-to-unity parameter in near unit root panels with unit-specific trends 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 where $T$ is fixed.}

Another Monte Carlo simulation exercise was undertaken to investigate the small-sample accuracy of our asymptotic results under the null hypothesis. The DGP used for this purpose is the same as the one used in generating the local power results reported in Table 2. The only difference is that in this case $a = b = 0$ ($\alpha_1 = ... = \alpha_N = 1$). In addition to $\tau^*_{1,\mu}$ the test statistic considered by Ng (2008, Section 5), henceforth denoted $\tau_{1,\mu}$, which can be seen as a conventional $t$-test of the constant term in a regression of $\Delta V_t$ onto a constant and $\Delta t^2$, is also simulated. According to Table 3 $\tau_{1,\mu}$ is severely size distorted. In fact, unreported simulation results suggest that $N$ has to be as large as 1,000 for this test to be reasonably sized, which
Table 2: 5% power of $\tau^*_t$ when $T = N^\gamma$.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$a = b = -1$</th>
<th>$a = -2, b = 0$</th>
<th>$a = b = -2$</th>
<th>$a = b = -4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>11.3</td>
<td>11.3</td>
<td>12.4</td>
<td>15.5</td>
</tr>
<tr>
<td>80</td>
<td>7.3</td>
<td>7.4</td>
<td>7.8</td>
<td>9.5</td>
</tr>
<tr>
<td>160</td>
<td>1.5</td>
<td>1.5</td>
<td>1.7</td>
<td>2.2</td>
</tr>
<tr>
<td>320</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.3</td>
</tr>
<tr>
<td>640</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

$\gamma = 1/4, \eta = 3/4$

<table>
<thead>
<tr>
<th>$N$</th>
<th>$a = b = -1$</th>
<th>$a = -2, b = 0$</th>
<th>$a = b = -2$</th>
<th>$a = b = -4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>5.2</td>
<td>5.1</td>
<td>9.4</td>
<td>15.5</td>
</tr>
<tr>
<td>80</td>
<td>3.2</td>
<td>3.3</td>
<td>6.0</td>
<td>13.3</td>
</tr>
<tr>
<td>160</td>
<td>1.8</td>
<td>1.8</td>
<td>4.0</td>
<td>10.0</td>
</tr>
<tr>
<td>320</td>
<td>0.8</td>
<td>0.9</td>
<td>2.7</td>
<td>8.9</td>
</tr>
<tr>
<td>640</td>
<td>0.4</td>
<td>0.4</td>
<td>1.5</td>
<td>7.3</td>
</tr>
</tbody>
</table>

$\gamma = \eta = 1/2$

<table>
<thead>
<tr>
<th>$N$</th>
<th>$a = b = -1$</th>
<th>$a = -2, b = 0$</th>
<th>$a = b = -2$</th>
<th>$a = b = -4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>20.2</td>
<td>19.6</td>
<td>30.1</td>
<td>38.4</td>
</tr>
<tr>
<td>80</td>
<td>14.4</td>
<td>14.3</td>
<td>24.4</td>
<td>35.6</td>
</tr>
<tr>
<td>160</td>
<td>5.3</td>
<td>5.4</td>
<td>15.6</td>
<td>33.7</td>
</tr>
<tr>
<td>320</td>
<td>1.6</td>
<td>1.9</td>
<td>9.4</td>
<td>35.0</td>
</tr>
<tr>
<td>640</td>
<td>0.3</td>
<td>0.3</td>
<td>6.4</td>
<td>41.6</td>
</tr>
</tbody>
</table>

$\gamma = \eta = 1/4$

| Notes: $\tau^*_t$ refers to the test with incidental trends. The autoregressive parameter is generated as $a_i = \exp(c_i/N^\eta)$, where $c_i \sim U(a, b)$. See Westerlund (2014a) for a description of the rest of the DGP.

is largely in agreement with the results reported by Ng (2008, Table 4). By contrast, the test statistic proposed here seems to perform well even in relatively small samples. We also see that the mean and variance of the empirical distribution of $\tau^*_t$ are close to the theoretically predicted values (zero and one, respectively), and that the accuracy increases as $N$ and $T$ increase. The same cannot be said about $\tau_t$.

As in the constant-only case, $\tau^*_t$ is only suitable for testing $H_0 : \theta = 1$. The appropriate test statistic to use when testing $H_0 : \theta = \theta_0 \in (0, 1]$ under $\eta = 0$ is given by

$$\tau_{\theta_0,t}^* = \sqrt{N/T}(\hat{\theta}_{BA,\theta_0,t}^* - 1) / \sqrt{\hat{\theta}^*\hat{\sigma}_\theta}.$$ 

By using the same arguments as in Westerlund (2014a) we can show that if the null hypothesis is true, then

$$\tau_{\theta_0,t}^* \rightarrow_d N(0, 1)$$
Table 3: 5% size, mean and variance of $\tau_{1,t}^*$ and $\tau_{1,t}$.  

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>$\tau_{1,t}^*$</th>
<th>$\tau_{1,t}$</th>
<th>$\tau_{1,t}^*$</th>
<th>$\tau_{1,t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>25</td>
<td>3.5</td>
<td>36.9</td>
<td>0.54</td>
<td>-20.94</td>
</tr>
<tr>
<td>10</td>
<td>50</td>
<td>3.5</td>
<td>38.6</td>
<td>0.22</td>
<td>-17.06</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>3.9</td>
<td>39.5</td>
<td>0.18</td>
<td>-12.9</td>
</tr>
<tr>
<td>10</td>
<td>200</td>
<td>4.1</td>
<td>43.8</td>
<td>0.09</td>
<td>-10.3</td>
</tr>
<tr>
<td>20</td>
<td>25</td>
<td>3.4</td>
<td>34</td>
<td>0.4</td>
<td>-24.74</td>
</tr>
<tr>
<td>20</td>
<td>50</td>
<td>3</td>
<td>34.9</td>
<td>0.32</td>
<td>-19.98</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>3.9</td>
<td>37.8</td>
<td>0.18</td>
<td>-16.15</td>
</tr>
<tr>
<td>20</td>
<td>200</td>
<td>4</td>
<td>41.9</td>
<td>0.12</td>
<td>-12.98</td>
</tr>
<tr>
<td>40</td>
<td>25</td>
<td>3</td>
<td>27.5</td>
<td>0.58</td>
<td>-24.15</td>
</tr>
<tr>
<td>40</td>
<td>50</td>
<td>2.4</td>
<td>30.8</td>
<td>0.4</td>
<td>-22.11</td>
</tr>
<tr>
<td>40</td>
<td>100</td>
<td>2.8</td>
<td>33.9</td>
<td>0.28</td>
<td>-19.24</td>
</tr>
<tr>
<td>40</td>
<td>200</td>
<td>3.5</td>
<td>38.1</td>
<td>0.17</td>
<td>-16.05</td>
</tr>
</tbody>
</table>

Notes: $\tau_{1,t}^*$ refers to the test with incidental trends, with $\tau_{1,t}$ being the corresponding test in Ng (2008). See Westerlund (2014a) for a description of the DGP.

as $N, T \to \infty$ with $1/2 < \gamma < 1$. If, on the other hand, the null hypothesis is false, then $\tau_{1,t}^* = O_p(\sqrt{N})$, which is the same result as in the constant-only case. The incidental trends problem is therefore only an issue when considering a local alternative hypothesis ($\eta > 0$).
References


