A Note on the Pooling of Individual PANIC Unit Root Tests

Westerlund, Joakim

2007

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

A Note on the Pooling of Individual PANIC Unit Root Tests*

Joakim Westerlund†

February 19, 2007

Abstract

One of the most cited studies in recent years within the field of nonstationary panel data analysis is that of Bai and Ng (2004, A PANIC Attack on Unit Roots and Cointegration. *Econometrica* 72, 1127-1177), in which the authors propose PANIC, a new framework for analyzing the nonstationarity of panels with idiosyncratic and common components. This paper shows that, although valid at the level of the individual unit, PANIC is not an asymptotically valid framework for pooling tests at the aggregate panel level.

JEL Classification: C21; C22; C23.

Keywords: Panel Unit Root Test; Pooling; Common Factor; Cross-Sectional Dependence.

1 Introduction

Consider the observed variable X_{it}, where $t = 1, ..., T$ and $i = 1, ..., N$ indexes the time series and cross-sectional units, respectively. The starting point of PANIC is to decompose X_{it} into two components, one that is common across i and one that is idiosyncratic. In this paper, we consider the simple setup with

*Previous versions of this paper were presented at Lund university and the University of Maastricht. The author would like to thank seminar participants, and in particular Jushan Bai, David Edgerton, Rolf Larsson, Serena Ng and Jean-Pierre Urbain for many valuable comments and suggestions. The author would also like to thank the Maastricht Research School of Economics of Technology and Organizations for its hospitality during a visit at the Department of Quantitative Economics at the University of Maastricht, where a part of this paper was written. Thank you also to the Jan Wallander and Tom Hedelius Foundation for financial support under research grant number W2006-0068:1. The usual disclaimer applies.

†Department of Economics, Lund University, P. O. Box 7082, S-220 07 Lund, Sweden. Telephone: +46 46 222 8670; Fax: +46 46 222 4118; E-mail address: joakim.westerlund@nek.lu.se.

an individual specific constant only, in which case X_{it} may be written as

$$X_{it} = c_i + \lambda_i' F_t + e_{it} = c_i + \sum_{j=1}^{r} \lambda_{ji} F_{jt} + e_{it}, \tag{1}$$

where the common factors F_{jt} and loadings λ_{ji} together represents the common component of X_{it}, while e_{it} represents the idiosyncratic component. These are assumed to be generated as

$$F_{jt} = \phi_j F_{jt-1} + u_{jt} \quad \text{and} \quad e_{it} = \rho_i e_{it-1} + \epsilon_{it}, \tag{2}$$

where we assume for simplicity that the errors u_{jt} and ϵ_{it} are serially uncorrelated. In this setup, the idiosyncratic component e_{it} has a unit root if $\rho_i = 1$ and it is stationary if $\rho_i < 1$. Similarly, if some of the ϕ_j parameters are equal to one, then X_{it} has as many common stochastic trends as the number of unit roots in F_t.

The objective of PANIC is to determine the number of common stochastic trends and test if $\rho_i = 1$ when F_t and e_{it} are estimated using the method of principal components. The problem is that if e_{it} is nonstationary, then this method cannot be applied to X_{it} because it will render the resulting estimate of λ_i inconsistent. Bai and Ng (2004) therefore suggest applying the principal components method to x_{it}, the first difference of X_{it}, rather than to X_{it} itself. To appreciate this point, note that x_{it} can be written as

$$x_{it} = \lambda_i' f_t + z_{it}, \tag{3}$$

where f_t and z_{it} are the first differences of F_t and e_{it}, respectively. In contrast to (1), all the components of this equation are stationary, which means that consistent estimates $\widehat{\lambda}_i$ and $\widehat{f}_t$ of λ_i and f_t can be obtained. The variables F_t and e_{it} can then be estimated by recumulating the first differenced series as

$$\widehat{F}_t = \sum_{s=2}^{t} \widehat{f}_s \quad \text{and} \quad \widehat{e}_{it} = \sum_{s=2}^{t} \widehat{z}_{is} \quad \text{where} \quad \widehat{z}_{is} = x_{is} - \widehat{\lambda}_i' \widehat{f}_s.$$

The idea behind PANIC is to test whether $\rho_i = 1$ by subjecting $\widehat{e}_{it}$ to any conventional unit root test, such as the classical Dickey and Fuller (1979) test, henceforth denoted $DF^c_{\widehat{e}}(i)$. The justification for testing in this particular way is that $DF^c_{\widehat{e}}(i)$ is asymptotically equivalent to $DF^c_{e}(i)$, the unit root test based on e_{it}. Similarly, knowing $\widehat{F}_t$ is as good as knowing F_t, in the sense that $DF^c_{\widehat{F}}(i)$ is asymptotically equivalent to $DF^c_{F}(i)$. This is very convenient as it implies that it is possible to disentangle the sources of potential nonstationarity in X_{it} by separately testing for unit roots in e_{it} and F_t.

Another interesting advantage of PANIC is that $DF^c_{\widehat{e}}(i)$ can be used to construct pooled tests for a unit root in e_{it}. The conventional way to construct such tests involves first demeaning the data, and then subjecting each of the demeaned series to a unit root test. If X_{it} is independent across i, the average of these tests converges to a normal variate under the null hypothesis of a unit root. Unfortunately, such tests are generally inappropriate as X_{it} will usually exhibit at least some form of dependence across i.[1] By contrast, pooled tests based on e_{it} are more widely applicable, since they are valid under the more plausible assumption that X_{it} admits to a common factor structure.

Yet another advantage, even in comparison to other studies that also permit for common factors, is that in PANIC the factors need not be stationary. This makes tests based on e_{it} very general indeed, and is probably one of the main reasons why PANIC has become so popular in both applied and theoretical work, see Breitung and Pesaran (2005).

In this paper, we point out a weakness in PANIC that seems to have been overlooked in the literature. In particular, we show that the theoretical results provided by Bai and Ng (2004) are not enough to ensure that PANIC can be used for the purpose of pooling individual unit root tests. This is because the asymptotic error incurred when replacing $DF_e^c(i)$ with $DF_{\widehat{e}}^c(i)$ is not small enough to vanish as N increases, but manifests itself as a nuisance parameter in the asymptotic distribution of the pooled test. Thus, although still valid at the level of the individual unit, PANIC is not a valid approach for constructing pooled tests.

2 Asymptotic results

In this section, we give the main theoretical results, using as an example the $DF_{\widehat{e}}^c(i)$ statistic, which was also considered by Bai and Ng (2004). However, the results apply to all panel tests that are based on pooling across individual test statistics or their p-values. The data generating process is taken directly from Bai and Ng (2004), and consists of (1) and (2) plus their Assumptions A through E. As in that paper, we also assume that u_{it} and ϵ_{it} are serially uncorrelated when the $DF_{\widehat{e}}^c(i)$ test is used.[2]

In the appendix we show that under the null hypothesis that $\rho_i = 1$, as $N, T \to \infty$

$$DF_{\widehat{e}}^c(i) \;=\; DF_e^c(i) + \mathcal{R}_i \;=\; DF_e^c(i) + O_p\left(\frac{1}{C_{NT}}\right) \;\Rightarrow\; \mathcal{B}_i, \tag{4}$$

where $\mathcal{R}_i$ is a reminder term, $C_{NT} = \min\{\sqrt{T}, \sqrt{N}\}$ and $\mathcal{B}_i$ is the usual Dickey and Fuller (1979) test distribution. The by far most common way of pooling

[1] See Breitung and Pesaran (2005) for a recent survey of the existing panel unit root literature.

[2] Violations can be easily accommodated by using any serial correlation corrected test, such as the augmented Dickey and Fuller (1979) test.

statistics of this sort is to take the cross-sectional average, which in the current context involves computing

$$DF_{\widehat{e}}^{c}(N) \;=\; \frac{1}{N}\sum_{i=1}^{N} DF_{\widehat{e}}^{c}(i).$$

Bai and Ng (2004) argues that since $DF_{\widehat{e}}^{c}(i)$ is asymptotically equivalent to $DF_{e}^{c}(i)$ and $\mathcal{B}_i$ is independent across i, then

$$\sqrt{N}(DF_{\widehat{e}}^{c}(N) - E(\mathcal{B}_i)) \;\Rightarrow\; N(0, \mathrm{var}(\mathcal{B}_i)) \quad \text{as} \quad N, T \to \infty.$$

However, this is not correct, as can be seen by writing

$$\begin{aligned}
\sqrt{N}(DF_{\widehat{e}}^{c}(N) - E(\mathcal{B}_i)) &= \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} DF_{\widehat{e}}^{c}(i) - E(\mathcal{B}_i)\right) \\
&= \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} DF_{e}^{c}(i) + O_p\left(\frac{1}{C_{NT}}\right) - E(\mathcal{B}_i)\right) \\
&= \sqrt{N}(DF_{e}^{c}(N) - E(\mathcal{B}_i)) + O_p\left(\frac{\sqrt{N}}{C_{NT}}\right). \qquad (5)
\end{aligned}$$

This result is summarized in the following theorem.

Theorem 1. Under the null hypothesis that $\rho_i = 1$ for all $i = 1, ..., N$, as $N, T \to \infty$

(a) $DF_{\widehat{e}}^{c}(N) \;\to_p\; E(\mathcal{B}_i)$,

(b) $\sqrt{N}(DF_{\widehat{e}}^{c}(N) - E(\mathcal{B}_i)) \;\Rightarrow\; N(0, \mathrm{var}(\mathcal{B}_i)) + \sqrt{N}\mathcal{R}_\infty$.

A detailed account of the reminder $\mathcal{R}_\infty$ is provided in the appendix. However, it is instructive to note that $\mathcal{R}_N = \frac{1}{N}\sum_{i=1}^{N}\mathcal{R}_i \to_p \mathcal{R}_\infty$, where

$$\begin{aligned}
\sqrt{N}\mathcal{R}_i &= O_p(1)\sqrt{N}\left(\frac{1}{T}e_{iT}A_{iT}\right) + O_p(1)\sqrt{N}\left(\frac{1}{T^2}\sum_{t=2}^{T} e_{it-1}A_{it-1}\right) \\
&+ O_p\left(\frac{\sqrt{N}}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right), \qquad (6)
\end{aligned}$$

where A_{it} denotes the cumulative sum of $a_{it} = \lambda_i' f_t - \widehat{\lambda}_i' \widehat{f}_t$. Thus, given that $N/T \to 0$, the last two terms on the right-hand side of (6) vanish as $N, T \to \infty$. The first and second terms, however, are $O_p(\sqrt{N}/C_{NT})$ and therefore do not vanish. Consider for example the first term, which under the null hypothesis of a unit root in e_{it} can be written as

$$\sqrt{N}\left(\frac{1}{T}e_{iT}A_{iT}\right) \;=\; \left(\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\epsilon_{it}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=2}^{T}(\sqrt{N}a_{it})\right) \;=\; O_p(1)O_p(1),$$

where the last equality follows from the fact that both terms in the product are normally distributed as $N, T \to \infty$ with $N/T \to 0$, see Bai (2003).

In other words, PANIC is not a valid approach for pooling tests unless λ_i and f_t are known so that a_{it} is zero. Specifically, the problem is that although the asymptotic distribution of $\sqrt{N}(DF^c_{\widehat{e}}(N) - E(\mathcal{B}_i))$ is correctly centered as indicated in Theorem 1 (a), it is no longer normal with variance $\text{var}(\mathcal{B}_i)$. Using the PANIC approach for an individual unit, however, will lead to an asymptotic distribution that is free of nuisance parameters, although nonnormal as seen in (4).

With N and T finite there is the additional problem that a_{it} will in general be different from zero, and hence we might also expect to observe some miscentering in small samples. In other words, replacing $DF^c_{\widehat{e}}(i)$ with $DF^c_e(i)$ induces both an asymptotic and a small-sample bias, which presents us with a theoretical difficulty that needs to be resolved. The proposal of this paper is very simple and involves applying the method of principal components to the whole panel, but using only a subset, M say, for constructing the pooled test.

Theorem 2. Suppose that $N/T \to 0$ and $M/N \to 0$ as $M, N, T \to \infty$, then under the null hypothesis that $\rho_i = 1$ for $i = 1, ..., M$

(a) $DF^c_{\widehat{e}}(M) \ \to_p \ E(\mathcal{B}_i)$,

(b) $\sqrt{M}(DF^c_{\widehat{e}}(M) - E(\mathcal{B}_i)) \ \Rightarrow \ N(0, \text{var}(\mathcal{B}_i))$.

This result follows naturally by noting that

$$\sqrt{M}(DF^c_{\widehat{e}}(M) - E(\mathcal{B}_i)) \quad = \quad \sqrt{M}(DF^c_e(M) - E(\mathcal{B}_i)) + \sqrt{M}\mathcal{R}_N, \qquad (7)$$

where the second term on the right-hand side is $O_p(\sqrt{M}/C_{NT})$, which vanishes under the condition that $N/T \to 0$ and $M/N \to 0$ suggesting that $\sqrt{M}(DF^c_{\widehat{e}}(M) - E(\mathcal{B}_i)) \Rightarrow N(0, \text{var}(\mathcal{B}_i))$ as $M, N, T \to \infty$. In other words, if we take as the null hypothesis a unit root in e_{it} for M out of the N units, then the problem disappears.[3] Note also that this result does not depend on the assumed serial independence, which can be easily relaxed by replacing $DF^c_{\widehat{e}}(i)$ with any serial correlation corrected test, see Bai and Ng (2004).

3 Monte Carlo simulations

In this section we look a little more closely at the small-sample effects of pooling individual PANIC unit root tests. The idea is to decompose $\mathcal{R} = \widehat{e}_N - e_N$, where

[3] Another way to get around the problem is to use *a priori* information about λ_i and f_t. For example, with full knowledge of λ_i and f_t, then a_{it} is zero and so the problem disappears. Of course, in applied work such information is rarely available, in which case bootstrapping might be a more feasible alternative.

$\widehat{e}_N$ and e_N are the standardized pooled tests based on $\widehat{e}_{it}$ and e_{it}, respectively. If PANIC is a valid pooling approach then the remainder $\mathcal{R}$ should collapse to zero as N and T grow, while if PANIC is not valid, then $\mathcal{R}$ should be nondegenerate.

Since our primary focus lies in examining the null distribution of the pooled unit root test, the data are generated according to (1) and (2) with $\rho_i = 1$ for all i. For simplicity, we further assume that ϕ_j is equal to ϕ for all j, that $r = 2$, that $\lambda_i \sim N(0, 1)$ and that $c_i = 1$ for all i. The errors u_{jt} and ϵ_{it} are both assumed to be mean zero and normally distributed with variance one and $\sigma^2_{\epsilon i} \sim U(0, b)$, respectively. The parameter b determines the relative variance of the idiosyncratic component, and therefore also the accuracy of the estimated common component, and is key in the simulations. If b is small, then λ_i and f_t will be estimated with high precision suggesting that the reminder $\mathcal{R}$ will be close to zero. Conversely, if b is large, the precision of the estimated common component is expected to be low, and so the effect of replacing e_{it} with $\widehat{e}_{it}$ is expected to be much larger. All computations have been performed in GAUSS using 5,000 replications. The results reported in Table 1 may be summarized as follows.[4]

Firstly, looking at the three rightmost columns, we see that $\mathcal{R}$ is a significant contributor to the variation of $\widehat{e}_N$ with a variance share close to 50% in most cases. We also see that this share does not tend to disappear as N and T grows, which supports the asymptotic results.

Another interesting observation is that $\mathcal{R}$ and e_N are negatively correlated, and that this correlation increases with b. In other words, the effect on $\widehat{e}_N$ of an increased variance in $\mathcal{R}$ is compensated by its negative correlation with e_N, which suggests that the asymptotic bias effect should be small. Indeed, a closer look at the three leftmost columns reveals that the size of $\widehat{e}_N$ is not affected very much by the variance of $\mathcal{R}$.

Secondly, it interesting to note how the centering of $\mathcal{R}$ is affected by N and T on the one hand, and by b on the other hand. With b fixed, we see that while decreasing in T, a larger N actually pushes the distribution of $\mathcal{R}$, and hence also that of $\widehat{e}_N$, to the right, thus making positive outcomes more likely. Hence, since the critical region is in the left tail of the normal distribution, this will make the test more conservative. By contrast, if we fix N and T, and instead let b increase, we see that the distribution of $\mathcal{R}$ tends to the left, thus causing $\widehat{e}_N$ to become oversized. These two effects are manifestations of the small-sample bias mentioned earlier, and are due to the low precision in the estimated common component.

Thirdly, concerning the stationarity of F_t, we see that setting $\phi = 1$ generally

[4] To better isolate the effect of pooling, we have assumed that the true number of factors is known. Also, as in the previous sections, we do not provide any results for the case with serial correlation. Interested readers are referred to the paper of Bai and Ng (2004) for some results when the data are serially correlated.

improves the performance of $\widehat{e}_N$. This is to be expected as this is the only case where f_t is equal to u_t. The fact that the variance of $\mathcal{R}$ varies with ϕ confirms that it is indeed dependent on the estimation of the common component.

Finally, note that while the performance of the pooled tests seems to be greatly affected by the parametrization of the data generating process, the performance of the individual test, denoted $\widehat{e}_i$, is essentially unaffected. This supports our claim that it is only when pooling that PANIC is no longer valid. It also implies that pooling only a subset of N should result in more accurate tests. Unreported simulation results confirm that this is indeed the case.

4 Concluding remarks

In this paper, we point out a weakness in the PANIC methodology developed by Bai and Ng (2004). The problem lies in the asymptotic error incurred when replacing e_{it} by $\widehat{e}_{it}$, which is not small enough to vanish when aggregating over N. This induces a nondegenerate bias, which makes the asymptotic distribution of the pooled test nonnormal and dependent on nuisance parameters that reflect, among other things, the accuracy of the estimated common component.

The proposed solution is very simple and follows from the fact that although decreasing at rate $\sqrt{N}$ at the individual level, the pooling across N nevertheless makes the aggregate error nondegenerate. This naturally leads to a subsampling approach where the size of the subsample is required to go to infinity at a slower rate than N. The simulation results confirm that pooling individual PANIC tests results in bias, but that the effect of this bias is generally quite small unless the variance of the idiosyncratic component is relatively large.

Finally, it should be stressed that the results reported here apply when pooling across individual test statistics or their p-values, which does not mean that PANIC cannot be used in combination with other pooling approaches. For example, as shown by Bai and Ng (2007), PANIC can still be used to construct asymptotically valid panel tests based on the pooled estimator of the autoregressive parameter.

Appendix: Mathematical proofs

In this appendix, we prove Theorems 1 and 2. In order to do so, however, we need some preliminary results. Throughout, we will make use of the fact that the common factor can only be identified up to a scale matrix H, say. Thus, what we will consider here is the rotation HF_t of F_t. As usual, $||A||$ will denote the Euclidean norm $(\text{tr}(A'A))^{1/2}$ of the matrix A.

Lemma A.1. Let $R_{NT} = \min\{\sqrt{T}, N\}$, then as $N, T \to \infty$

$$\begin{aligned}
&\text{(a)} \quad \frac{1}{T}\sum_{t=2}^{T}(\Delta\widehat{e}_{it})^2 \;=\; \frac{1}{T}\sum_{t=2}^{T}(\Delta e_{it})^2 + O_p\left(\frac{1}{C_{NT}^2}\right),\\
&\text{(b)} \quad \frac{1}{T}\sum_{t=2}^{T}\widehat{e}_{it-1}\Delta\widehat{e}_{it} \;=\; \frac{1}{T}\sum_{t=2}^{T}e_{it-1}\Delta e_{it} - \frac{1}{T}e_{iT}A_{iT} + O_p\left(\frac{1}{R_{NT}}\right),\\
&\text{(c)} \quad \frac{1}{T^2}\sum_{t=2}^{T}\widehat{e}_{it-1}^2 \;=\; \frac{1}{T^2}\sum_{t=2}^{T}e_{it-1}^2 - \frac{2}{T^2}\sum_{t=2}^{T}e_{it-1}A_{it-1} + O_p\left(\frac{1}{R_{NT}}\right).
\end{aligned}$$

Proof of Lemma A.1.

We begin with (a). From the text, we have that the defactored and first differentiated residuals can be written as

$$\Delta\widehat{e}_{it} \;=\; x_{it} - \widehat{\lambda}_i'\widehat{f}_t \;=\; \Delta e_{it} + \lambda_i' f_t - \widehat{\lambda}_i'\widehat{f}_t, \tag{A1}$$

which, by some algebra, can be restated as

$$\begin{aligned}
\Delta\widehat{e}_{it} &= \Delta e_{it} + \lambda_i' H^{-1}Hf_t - \lambda_i' H^{-1}\widehat{f}_t + \lambda_i' H^{-1}\widehat{f}_t - \widehat{\lambda}_i'\widehat{f}_t\\
&= \Delta e_{it} - \lambda_i' H^{-1}(\widehat{f}_t - Hf_t) - (\widehat{\lambda}_i - H^{-1\prime}\lambda_i)'\widehat{f}_t\\
&= \Delta e_{it} - \lambda_i' H^{-1}v_t - d_i'\widehat{f}_t,
\end{aligned} \tag{A2}$$

where $d_i = \widehat{\lambda}_i - H^{-1\prime}\lambda_i$ and $v_t = \widehat{f}_t - Hf_t$. Let $a_{it} = \lambda_i' H^{-1}v_t - d_i'\widehat{f}_t$, so that (A2) becomes $\Delta\widehat{e}_{it} = \Delta e_{it} - a_{it}$. This implies

$$\begin{aligned}
\frac{1}{T}\sum_{t=2}^{T}(\Delta\widehat{e}_{it})^2 &= \frac{1}{T}\sum_{t=2}^{T}(\Delta e_{it} - a_{it})^2\\
&= \frac{1}{T}\sum_{t=2}^{T}(\Delta e_{it})^2 + \frac{1}{T}\sum_{t=2}^{T}a_{it}^2 - \frac{2}{T}\sum_{t=2}^{T}\Delta e_{it}a_{it}\\
&= \frac{1}{T}\sum_{t=2}^{T}(\Delta e_{it})^2 + I - II, \quad \text{say.}
\end{aligned} \tag{A3}$$

By using the same arguments as in Lemma B.1 of Bai and Ng (2004), part I is $O_p(1/C_{NT}^2)$. Part II can be written as

$$
\begin{aligned}
II &= 2\lambda_i' H^{-1} \frac{1}{T}\sum_{t=2}^{T} \Delta e_{it} v_t - 2d_i' \frac{1}{T}\sum_{t=2}^{T} \Delta e_{it} \widehat{f}_t \\
&= -2d_i' \frac{1}{T}\sum_{t=2}^{T} \Delta e_{it} \widehat{f}_t + O_p\left(\frac{1}{C_{NT}^2}\right) \\
&= -2d_i' \frac{1}{T}\sum_{t=2}^{T} \Delta e_{it} (\widehat{f}_t - Hf_t) - 2d_i' H \frac{1}{T}\sum_{t=2}^{T} \Delta e_{it} f_t + O_p\left(\frac{1}{C_{NT}^2}\right),
\end{aligned}
$$

where the second equality follows by Lemma B.1 of Bai (2003). By applying $||AB|| \leq ||A||||B||$ and the triangle inequality to the above expression, we get

$$
\begin{aligned}
|II| &\leq 2||d_i|| \left(\frac{1}{T}\sum_{t=2}^{T} ||\Delta e_{it}(\widehat{f}_t - Hf_t)||\right) \\
&+ 2||d_i||||H|| \left(\frac{1}{T}\sum_{t=2}^{T} ||\Delta e_{it} f_t||\right) + O_p\left(\frac{1}{C_{NT}^2}\right),
\end{aligned}
$$

which, by applying the Cauchy-Schwarz inequality to the first term on the right-hand side, reduces to

$$
\begin{aligned}
|II| &\leq 2||d_i|| \left(\frac{1}{T}\sum_{t=2}^{T} (\Delta e_{it})^2\right)^{1/2} \left(\frac{1}{T}\sum_{t=2}^{T} ||\widehat{f}_t - Hf_t||^2\right)^{1/2} \\
&+ 2||d_i||||H|| \left(\frac{1}{T}\sum_{t=2}^{T} ||\Delta e_{it} f_t||\right) + O_p\left(\frac{1}{C_{NT}^2}\right) \\
&= O_p\left(\frac{1}{R_{NT}}\right) O_p(1) O_p\left(\frac{1}{C_{NT}}\right) + \left(\frac{1}{\sqrt{T} R_{NT}}\right) O_p(1) + O_p\left(\frac{1}{C_{NT}^2}\right),
\end{aligned}
$$

where $||H|| = O_p(1)$ by construction and $||d_i|| = O_p(1/R_{NT})$ by Lemma 1 (c) of Bai and Ng (2004). Also, from Lemma A.1 of Bai (2003), we have

$$
\frac{1}{T}\sum_{t=1}^{T} ||\widehat{f}_k - Hf_k||^2 = O_p\left(\frac{1}{C_{NT}^2}\right). \tag{A4}
$$

This implies that II is $O_p(1/C_{NT}^2)$, which in turn implies

$$
\frac{1}{T}\sum_{t=2}^{T} (\Delta \widehat{e}_{it})^2 = \frac{1}{T}\sum_{t=2}^{T} (\Delta e_{it})^2 + O_p\left(\frac{1}{C_{NT}^2}\right).
$$

This proves (a).

In order to prove (b) we use that $\widehat{e}_{it}^2$ may be written as $\widehat{e}_{it}^2 = (\widehat{e}_{it-1}+\Delta\widehat{e}_{it})^2 = \widehat{e}_{it-1}^2 + (\Delta\widehat{e}_{it})^2 + 2\widehat{e}_{it-1}\Delta\widehat{e}_{it}$, from which it follows that

$$\begin{aligned}\frac{1}{T}\sum_{t=2}^{T}\widehat{e}_{it-1}\Delta\widehat{e}_{it} &= \frac{1}{2T}\sum_{t=2}^{T}(\widehat{e}_{it}^2 - \widehat{e}_{it-1}^2 - (\Delta\widehat{e}_{it})^2) \\ &= \frac{1}{2T}\widehat{e}_{iT}^2 - \frac{1}{2T}\widehat{e}_{i1}^2 - \frac{1}{2T}\sum_{t=2}^{T}(\Delta\widehat{e}_{it})^2. \qquad \text{(A5)}\end{aligned}$$

Similarly, by applying the same trick to e_{it}^2, we have

$$\frac{1}{T}\sum_{t=2}^{T}e_{it-1}\Delta e_{it} = \frac{1}{2T}e_{iT}^2 - \frac{1}{2T}e_{i1}^2 - \frac{1}{2T}\sum_{t=2}^{T}(\Delta e_{it})^2. \qquad \text{(A6)}$$

Now, the terms in the middle of the right-hand side of (A5) and (A6) are clearly $O_p(1/T)$ as $\widehat{e}_{i1} = 0$ and $e_{i1} = O_p(1)$ by assumption. Also, by using (a) the difference between the third terms is $O_p(1/C_{NT}^2)$. For the first term, let $\widehat{e}_{it} = e_{it} - e_{i1} - A_{it}$, where A_{it} is the cumulative sum of a_{it}.

$$\begin{aligned}\frac{1}{T}\widehat{e}_{iT}^2 &= \frac{1}{T}(e_{iT} - e_{i1} - A_{iT})^2 \\ &= \frac{1}{T}e_{iT}^2 + \frac{1}{T}e_{i1}^2 + \frac{1}{T}A_{iT}^2 - \frac{2}{T}e_{iT}A_{iT} - \frac{2}{T}e_{i1}(e_{iT} + A_{iT}) \\ &= \frac{1}{T}e_{iT}^2 + I + II - III - IV, \quad \text{say.} \qquad \text{(A7)}\end{aligned}$$

By using $(A+B)^2 \leq 2(A^2+B^2)$, the triangle inequality, $||AB|| \leq ||A||||B||$ and then (A4), part II can be written as

$$\begin{aligned}II &= \frac{1}{T}(\lambda_i' H^{-1}V_T - d_i'\widehat{F}_T)^2 \\ &\leq 2||\lambda_i'H^{-1}||^2\left(\frac{1}{T}\sum_{t=2}^{T}||v_t||^2\right) - 2||d_i||^2\left(\frac{1}{T}\sum_{t=2}^{T}||\widehat{f}_t||^2\right) \\ &= O_p\left(\frac{1}{C_{NT}^2}\right) + O_p\left(\frac{1}{R_{NT}^2}\right)O_p(1).\end{aligned}$$

Hence, II is $O_p(1/C_{NT}^2)$.

Part III is simply

$$\begin{aligned}III &= \frac{2}{T}e_{iT}(\lambda_i'H^{-1}V_T - d_i\widehat{F}_T) \\ &= 2\left(\frac{1}{\sqrt{T}}e_{iT}\right)\lambda_i'H^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=2}^{T}v_t\right) - \left(\frac{1}{\sqrt{T}}e_{iT}\right)d_i'\left(\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\widehat{f}_t\right) \\ &= O_p(1)O_p\left(\frac{1}{C_{NT}}\right) + O_p(1)O_p\left(\frac{1}{R_{NT}}\right)O_p(1),\end{aligned}$$

where we have used equation (A.3) of Bai and Ng (2004), which says that

$$\frac{1}{\sqrt{T}}\sum_{j=2}^{t} v_j \;=\; O_p\left(\frac{1}{C_{NT}}\right).$$

Thus, III is $O_p(1/C_{NT})$. Part IV is dominated by e_{iT}/T, which is $O_p(1/\sqrt{T})$. Therefore, by adding the terms, (A7) simplifies to

$$\frac{1}{T}\widehat{e}_{iT}^2 \;=\; \frac{1}{T}e_{iT}^2 - \frac{2}{T}e_{iT}A_{iT} + O_p\left(\frac{1}{R_{NT}}\right),$$

which it turn implies that

$$\frac{1}{T}\sum_{t=2}^{T}\widehat{e}_{it-1}\Delta\widehat{e}_{it} \;=\; \frac{1}{T}\sum_{t=2}^{T}e_{it-1}\Delta e_{it} - \frac{1}{T}e_{iT}A_{iT} + O_p\left(\frac{1}{R_{NT}}\right).$$

This establishes part (b).

Finally, consider (c). By using that $\widehat{e}_{it} = e_{it} - e_{i1} - A_{it}$, this part may be written as

$$\begin{aligned}
\frac{1}{T^2}\sum_{t=2}^{T}\widehat{e}_{it}^2 \;&=\; \frac{1}{T^2}\sum_{t=2}^{T}(e_{it} - e_{i1} - A_{it})^2 \\
&=\; \frac{1}{T^2}\sum_{t=2}^{T}e_{it}^2 + \frac{1}{T^2}e_{i1}^2 + \frac{1}{T^2}\sum_{t=2}^{T}A_{it}^2 - \frac{2}{T^2}\sum_{t=2}^{T}e_{it}A_{it} \\
&-\; 2e_{i1}\left(\frac{1}{T^2}\sum_{t=2}^{T}e_{it} + \frac{1}{T^2}\sum_{t=2}^{T}A_{it}\right) \\
&=\; \frac{1}{T^2}\sum_{t=2}^{T}e_{it}^2 + I + II - III - IV, \quad \text{say.}
\end{aligned} \tag{A8}$$

Now, I is obviously $O_p(1/T^2)$. The next step is to show that II is $O_p(1/C_{NT}^2)$. In so doing, by first applying the triangle inequality and then $(A + B)^2 \leq 2(A^2 + B^2)$, we obtain

$$\begin{aligned}
|II| \;&=\; \frac{1}{T^2}\left|\left|\sum_{t=2}^{T}(\lambda_i' H^{-1}V_t - d_i'\widehat{F}_t)^2\right|\right| \\
&\leq\; 2||\lambda_i' H^{-1}||^2\left(\frac{1}{T^2}\sum_{t=2}^{T}||V_t||^2\right) + 2||d_i||^2\left(\frac{1}{T^2}\sum_{t=2}^{T}||\widehat{F}_t||^2\right) \\
&=\; O_p(1/N) + O_p\left(\frac{1}{R_{NT}^2}\right)O_p(1),
\end{aligned}$$

where we have used (A4), from which it follows that

$$\frac{1}{T^2}\sum_{t=2}^{T}||V_t||^2 \;=\; O_p\left(\frac{1}{C_{NT}^2}\right). \tag{A9}$$

Part III can be rewritten as

$$\begin{aligned} III &= 2\lambda_i' H^{-1} \frac{1}{T^2} \sum_{t=2}^{T} e_{it} V_t - 2d_i' \frac{1}{T^2} \sum_{t=2}^{T} e_{it} \widehat{F}_t \\ &= 2\lambda_i' H^{-1} \frac{1}{T^2} \sum_{t=2}^{T} e_{it} V_t - 2d_i' \frac{1}{T^2} \sum_{t=2}^{T} e_{it} (\widehat{F}_t - HF_t) \\ &+ 2d_i' H \left(\frac{1}{T^2} \sum_{t=2}^{T} e_{it} F_t \right) \\ &= O_p\left(\frac{1}{C_{NT}}\right) + O_p\left(\frac{1}{R_{NT}}\right) O_p\left(\frac{1}{C_{NT}}\right) + O_p\left(\frac{1}{R_{NT}}\right) O_p(1), \end{aligned}$$

where the order of the first two terms on the right-hand follows from first using Cauchy-Schwarz inequality and then (A9), as seen by writing

$$\frac{1}{T^2} \sum_{t=2}^{T} e_{it} V_t \le \left(\frac{1}{T^2} \sum_{t=2}^{T} e_{it}^2 \right)^{1/2} \left(\frac{1}{T^2} \sum_{t=2}^{T} ||V_t||^2 \right)^{1/2} = O_p(1) O_p\left(\frac{1}{C_{NT}}\right).$$

If follows that III is $O_p(1/C_{NT})$.

Finally, consider part V. The first term within the parenthesis is $O_p(1/\sqrt{T})$. For the second term, we have

$$\begin{aligned} \frac{1}{T^2} \sum_{t=2}^{T} A_{it} &= \lambda_i' H^{-1} \left(\frac{1}{T^2} \sum_{t=2}^{T} V_t \right) + d_i' H \left(\frac{1}{T^2} \sum_{t=2}^{T} F_t \right) \\ &- d_i' \left(\frac{1}{T^2} \sum_{t=2}^{T} (\widehat{F}_t - HF_t) \right) \\ &= \left(\frac{1}{\sqrt{T} C_{NT}} \right) + \left(\frac{1}{\sqrt{T} R_{NT}} \right) O_p(1) + \left(\frac{1}{\sqrt{T} R_{NT}} \right) O_p\left(\frac{1}{C_{NT}}\right). \end{aligned}$$

Thus, by collecting all the terms, we can show that (A8) reduces to

$$\frac{1}{T^2} \sum_{t=2}^{T} \widehat{e}_{it}^2 = \frac{1}{T^2} \sum_{t=2}^{T} e_{it}^2 - \frac{2}{T^2} \sum_{t=2}^{T} e_{it} A_{it} + O_p\left(\frac{1}{R_{NT}}\right).$$

This establishes (c) and thus the proof of Lemma A.1 is complete. ■

Proof of Theorem 1.

By definition

$$DF_{\widehat{e}}^{c}(i) \;=\; \left(\widehat{\sigma}_{\epsilon i}^{2}\frac{1}{T^{2}}\sum_{t=2}^{T}\widehat{e}_{it}^{2}\right)^{-1/2}\frac{1}{T}\sum_{t=2}^{T}\widehat{e}_{it-1}\Delta\widehat{e}_{it},$$

where $\widehat{\sigma}_{\epsilon i}^{2} = \frac{1}{T}\sum_{t=1}^{T}\widehat{\epsilon}_{it}^{2}$ and $\widehat{\epsilon}_{it} = \widehat{e}_{it} - \widehat{\rho}_{i}\widehat{e}_{it-1} = \Delta\widehat{e}_{it} - (\widehat{\rho}_{i}-1)\widehat{e}_{it-1}$. Thus, given Lemma A.1 (a) and the null that $\rho_i = 1$ for all i, then

$$\begin{aligned}
\widehat{\sigma}_{\epsilon i}^{2} &= \frac{1}{T}\sum_{t=2}^{T}(\Delta\widehat{e}_{it} - (\widehat{\rho}_{i}-1)\widehat{e}_{it-1})^{2} \\
&= \frac{1}{T}\sum_{t=2}^{T}(\Delta\widehat{e}_{it})^{2} - 2(\widehat{\rho}_{i}-1)\left(\frac{1}{T}\sum_{t=2}^{T}\widehat{e}_{it-1}\Delta\widehat{e}_{it}\right) \\
&+ T(\widehat{\rho}_{i}-1)^{2}\left(\frac{1}{T^{2}}\sum_{t=2}^{T}\widehat{e}_{it}^{2}\right) \\
&= \frac{1}{T}\sum_{t=2}^{T}(\Delta\widehat{e}_{it})^{2} + O_p(1/T)O_p(1) + TO_p(1/T^{2})O_p(1) \\
&= \frac{1}{T}\sum_{t=2}^{T}(\Delta e_{it})^{2} + O_p\left(\frac{1}{C_{NT}^{2}}\right) \;\to_p\; \sigma_{\epsilon i}^{2}. \qquad \text{(A10)}
\end{aligned}$$

This result, together with Lemma A.1 (c), implies that

$$\widehat{\sigma}_{\epsilon i}^{2}\frac{1}{T^{2}}\sum_{t=2}^{T}\widehat{e}_{it-1}^{2} \;=\; \sigma_{\epsilon i}^{2}\frac{1}{T^{2}}\sum_{t=2}^{T}e_{it-1}^{2} - 2\sigma_{\epsilon i}^{2}\frac{1}{T^{2}}\sum_{t=2}^{T}e_{it-1}A_{it-1} + O_p\left(\frac{1}{R_{NT}}\right),$$

which, by a Taylor expansion of the inverse square root, yields

$$\begin{aligned}
\left(\widehat{\sigma}_{\epsilon i}^{2}\frac{1}{T^{2}}\sum_{t=2}^{T}\widehat{e}_{it-1}^{2}\right)^{-1/2} &= \left(\sigma_{\epsilon i}^{2}\frac{1}{T^{2}}\sum_{t=2}^{T}e_{it-1}^{2}\right)^{-1/2} \\
&+ \frac{1}{\sigma_{\epsilon i}}\left(\frac{1}{T^{2}}\sum_{t=2}^{T}e_{it-1}^{2}\right)^{-3/2}\frac{1}{T^{2}}\sum_{t=2}^{T}e_{it-1}A_{it-1} + O_p\left(\frac{1}{R_{NT}}\right).
\end{aligned}$$

Another application of Lemma A.1 gives

$$\begin{aligned}
DF_{\widehat{e}}^{c}(i) &= DF_{e}^{c}(i) - \frac{1}{\sigma_{\epsilon i}}\left(\frac{1}{T^{2}}\sum_{t=2}^{T}e_{it-1}^{2}\right)^{-1/2}\frac{1}{T}e_{iT}A_{iT} \\
&- DF_{e}^{c}(i)\frac{1}{\sigma_{\epsilon i}^{2}}\left(\frac{1}{T^{2}}\sum_{t=2}^{T}e_{it-1}^{2}\right)^{-1}\frac{1}{T^{2}}\sum_{t=2}^{T}e_{it-1}A_{it-1} + O_p\left(\frac{1}{R_{NT}}\right) \\
&= DF_{e}^{c}(i) + \mathcal{R}_{i}, \qquad \text{(A11)}
\end{aligned}$$

where $\mathcal{R}_i = O_p(1/C_{NT})$. Thus, because $DF_e^c(i) \Rightarrow \mathcal{B}_i$ as $N, T \to \infty$ and $\mathcal{B}_i$ is independent across i, we can appeal to Theorem 1 of Phillips and Moon (1999) from which it follows that

$$\begin{aligned}
DF_{\widehat{e}}^c(N) &= \frac{1}{N}\sum_{i=1}^{N} DF_{\widehat{e}}^c(i) = \frac{1}{N}\sum_{i=1}^{N}\Big(DF_e^c(i) + \mathcal{R}_i\Big) = DF_e^c(N) + \mathcal{R}_N \\
&= DF_e^c(N) + O_p\left(\frac{1}{C_{NT}}\right) \to_p E(\mathcal{B}_i).
\end{aligned}$$

This establishes (a).

To prove (b) just write

$$\begin{aligned}
\sqrt{N}(DF_{\widehat{e}}^c(N) - E(\mathcal{B}_i)) &= \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} DF_{\widehat{e}}^c(i) - E(\mathcal{B}_i)\right) \\
&= \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} DF_e^c(i) + \mathcal{R}_N - E(\mathcal{B}_i)\right) \\
&= \sqrt{N}(DF_e^c(N) - E(\mathcal{B}_i)) + \sqrt{N}\mathcal{R}_N,
\end{aligned}$$

where $\sqrt{N}(DF_e^c(N) - E(\mathcal{B}_i)) \Rightarrow N(0, \mathrm{var}(\mathcal{B}_i))$ by Theorem 3 of Phillips and Moon (1999) and $\mathcal{R}_N \to_p \mathcal{R}_\infty$. ■

Proof of Theorem 2.

Part (a) follows from Theorem 1 (a) while part (b) is an immediate consequence of (7). ■

Table 1: Simulation results.

				Test size			Mean		Variance		
b	ϕ	T	N	$\widehat{e}_N$	e_N	$\widehat{e}_i$	$\mathcal{R}$	$\lvert\mathcal{R}\rvert$	e_N	$\mathcal{R}$	cov
1	0	50	50	3.6	4.2	4.7	0.26	0.63	74.0	42.0	−15.9
			100	3.0	4.3	4.8	0.35	0.64	75.8	40.5	−16.3
		100	50	5.4	4.6	4.7	0.13	0.56	73.1	39.3	−12.4
			100	5.3	4.4	4.9	0.19	0.57	72.2	36.3	−8.5
	0.5	50	50	4.3	4.5	4.8	0.26	0.63	73.6	42.2	−15.8
			100	2.9	3.7	4.8	0.33	0.63	75.0	39.3	−14.3
		100	50	5.4	4.5	4.8	0.11	0.57	74.0	39.5	−13.5
			100	5.5	5.0	4.9	0.17	0.58	71.5	37.3	−8.8
	1	50	50	6.1	3.6	5.1	0.01	0.68	72.9	51.2	−24.1
			100	5.2	3.0	5.1	0.03	0.67	69.7	50.9	−20.6
		100	50	6.7	4.6	5.0	−0.01	0.61	76.0	44.9	−20.9
			100	6.3	3.9	5.0	0.02	0.61	73.7	44.7	−18.3
10	0	50	50	22.7	4.8	7.5	−0.84	1.01	68.5	53.8	−22.3
			100	12.5	4.3	5.9	−0.35	0.75	70.2	49.3	−19.5
		100	50	20.8	4.7	7.1	−0.77	0.93	68.3	48.3	−16.6
			100	10.4	4.6	5.4	−0.21	0.64	69.8	41.4	−11.2
	0.5	50	50	24.9	4.6	8.0	−0.96	1.08	80.1	59.1	−39.2
			100	19.9	3.8	6.7	−0.77	0.96	69.9	51.8	−21.7
		100	50	29.3	4.4	8.3	−1.08	1.15	74.0	49.5	−23.6
			100	18.9	3.9	6.3	−0.64	0.84	68.2	42.6	−10.9
	1	50	50	7.3	4.5	5.2	−0.01	0.89	71.1	86.4	−57.5
			100	8.4	3.9	5.2	0.06	1.01	53.2	83.2	−36.4
		100	50	7.7	4.5	4.9	0.03	0.83	64.0	71.5	−35.5
			100	8.7	4.0	5.1	−0.02	0.86	56.7	69.1	−25.8

Notes: Let $\widehat{e}_N$ and e_N denote the standardized pooled tests based on $\widehat{e}_{it}$ and e_{it}, respectively, and let $\mathcal{R}$ denote the remainder, then $\widehat{e}_N = e_N + \mathcal{R}$. Also, $\widehat{e}_i$ refers to the individual test based on $\widehat{e}_{it}$, b refers to upper support for the variance of e_{it} and ϕ refers to the autoregressive parameter of F_t. The leftmost three columns report the size at the nominal 5% level, while the middle two columns report the average of $\mathcal{R}$ and its absolute value. The rightmost three columns report the percentage of the total variance in $\widehat{e}_N$ that is due to variance in e_N, $\mathcal{R}$ and their covariance, respectively.

References

Bai, J. (2003). Inferential Theory for Factor Models of Large Dimensions. *Econometrica*, Vol. 71, pp. 135-171.

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

Bai, J., and S. Ng (2007). Panel Unit Root Tests with Cross-Section Dependence: A Further Investigation. Unpublished manuscript.

Breitung, J., and M. H. Pesaran (2005). Unit Roots and Cointegration in Panels. In *The Econometrics of Panel Data*. Mátyás, L., and P. Sevestre (Eds.), Kluwer Academic Publishers.

Dickey, D. A., and W. A. Fuller (1979). Distribution of the Estimator for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, Vol. 74, pp. 427-431.

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