Panel Unit Root Tests with Cross-Section Dependence: A Further Investigation

Jushan Bai* Serena Ng †

November 19, 2007

Abstract

An effective way to control for cross-section correlation when conducting panel unit root test is to remove the common factors from the data. However, there remains many ways to use the defactored residuals to construct a test. In this paper, we compare tests based on a pooled estimate of the autoregressive coefficient with those that do not. We first use the PANIC residuals to form two tests, one estimates the autoregressive coefficient and one simply uses a sample moment, and establish their large sample properties. We then provide a bias correction to improve the size of Moon-Perron tests. Upon comparing the properties of these tests, we find that whether we project the data on the factors or subtract the factors from the data, the pooled autoregressive estimate has little power in rejecting the unit root hypothesis in the presence of incidental trends. Tests with some power in such cases are the ones that do not depend on the estimated autoregressive coefficient.

*Department of Economics, NYU, 269 Mercer St, New York, NY 10003 Email: Jushan.Bai@nyu.edu.

†Department of Economics, University of Michigan and Columbia University, New York, NY 10027 Email: Serena.Ng@Columbia.edu

We would like to thank three referees, the co-editor, and Benoit Perron for many helpful comments and suggestions. We also acknowledge financial support from the NSF (grants SES-0137084, SES-0136923, SES-0551275, and SES-0549978)

1 Introduction

Cross-section dependence can pose serious problems for testing the null hypothesis that all units in a panel are non-stationary. As first documented in O'Connell (1998), much of what appeared to be power gains in panel unit root tests developed under the assumption of cross-section independence over individual unit root tests are in fact the consequence of non-trivial size distortions. Many tests have been developed to relax the cross-section independence assumption. See Chang (2002), Chang and Song (2002), and Pesaran (2007), among others. An increasingly popular approach is to model the cross-section dependence using common factors. The PANIC framework of Bai and Ng (2004) (hereafter BN) enables the common factors and the idiosyncratic errors to be tested separately, while Moon and Perron (2004) (hereafter MP) tests the orthogonal projection of the data on the common factors. Most tests are formulated as an average of the individual statistics or their p-values. The MP tests retain the spirit of the original panel unit root test of Levin et al. (2002), which estimates and tests the pooled first order autoregressive parameter. As pointed out by Maddala and Wu (1999), the Levin et al. (2002) type tests have good power when autoregressive roots are identical over the cross sections. On the other hand, pooling individual test statistics may be more appropriate when there is heterogeneity in the dynamic parameters.

This paper provides a better understanding of tests based on the pooled autoregressive coefficient and those that do not. To do so, we first develop MP type tests using the PANIC residuals, and a panel version of the Modified Sargan-Bhargava test (hereafter PMSB) that simply uses the sample moments of these residuals. We also suggest an iterative procedure to improve the size properties of the MP tests. We then use simulations to show that autoregressive coefficient based tests have minimal power when there are incidental trends, irrespective of how the factors are removed. Of all the tests considered, only the original PANIC test that pools the p values and, to some extent, the PMSB test have any power against the trend stationary alternative. These two tests do not estimate the pooled autoregressive coefficient. Many papers have compared finite sample properties of panel unit root tests. One contribution of our paper is to trace the difference in finite sample properties to whether or not the pooled autoregressive coefficient is estimated.

The rest of the paper is organized as follows. In Section 2, we consider two data generating processes that involve common factors, introduce necessary notations, and discuss model assumptions. In Section 3, we consider the PANIC residual-based MP type and PMSB tests. Section 4 presents an improved version of the MP test. Section 5 provides finite sample evidence via Monta Carlo simulations. Concluding remarks are given in Section 6, and the proofs are given in the appendix.

2 Preliminaries

We consider two data generating processes (DGP). Let $D_{it} = \sum_{j=0}^{p} \delta_{ij} t^j$ be the deterministic component. When $p = 0$, $D_{it} = \delta_i$ is the individual specific fixed effect, and when $p = 1$, an individual specific time trend is also present. When there is no deterministic term, D_{it} is null and we will refer to this as case $p = -1$. Throughout the paper, the projection matrix M_0 demeans the data and M_1 demeans and detrends the data.

DGP 1 Following Bai and Ng (2004), the data are generated as

$$
\begin{aligned}
X_{it} &= D_{it} + \lambda_i' F_t + e_{it} \\
(1-L)F_t &= C(L)\eta_t \\
e_{it} &= \rho_i \, e_{it-1} + \varepsilon_{it}
\end{aligned}
$$

where $C(L) = \sum_{j=0}^{\infty} c_j L^j$ is an one-sided lag polynomial, F_t is an $r \times 1$ vector of common factors that induce correlation across units, λ_i is a vector of factor loadings, and e_{it} is an idiosyncratic error, ε_{it} itself is a linear process such that $\varepsilon_{it} = \sum_{j=0}^{\infty} d_{ij} v_{it-j}$ with v_{it} being iid over t, zero mean and finite eighth moment. We assume summability condition such that $\sum_{j=1}^{\infty} j|c_j| < M$ and $\sum_{j=1}^{\infty} j|d_{ij}| < M$ uniformly in i. The number of common stochastic trends is determined by r_1, the rank of $C(1)$. When $r_1 = 0$, ΔF_t is overdifferenced and the common factors are stationary. For the purpose of this note, it is useful to consider an alternative representation of F_t:

$$F_t = \Phi_1 F_{t-1} + \eta_t.$$

The number of non-stationary factors is determined by the number of unit roots in the polynomial matrix equation, $\Phi(L) = I - \Phi_1 L = 0$. Under DGP 1, X_{it} can be non-stationary when $\Phi(L)$ has a unit root, or $\rho_i = 1$, or both. Clearly, if the common factors share a stochastic trend, X_{it} will all be non-stationary. The DGP used in Phillips and Sul (2003) is a special case of DGP 1, as they only allow for one factor, and the idiosyncratic errors are independently distributed across time. Choi (2006) also assumes one factor, but the idiosyncratic errors are allowed to be serially correlated. However, the units are restricted to have a homogeneous response to F_t (i.e., $\lambda_i = 1$).

An important feature of DGP 1 is that the common and the idiosyncratic components can have different orders of integration. It is only when we reject non-stationarity in both components that we can say that the data are inconsistent with unit root non-stationarity.

DGP 2 The DGP used in Moon and Perron (2004) and Moon et al. (2005), is

$$\begin{aligned} X_{it} &= D_{it} + X_{it}^0 \\ X_{it}^0 &= \rho_i X_{it-1}^0 + u_{it} \\ u_{it} &= \lambda_i' f_t + \varepsilon_{it}. \end{aligned}$$

where f_t and ε_{it} are I(0) linear processes . The series X_{it} has a unit root if $\rho_i = 1$.

DGP 1 differs from DGP 2 in that the former specifies the dynamics of the unobserved components, whereas the latter specifies the dynamics of the observed series. Assuming $X_{i0}^0 = 0$ and $\rho_i = \rho$ for all i, DGP 2 can be rewritten as DGP 1 as follows:

$$X_{it} = D_{it} + \lambda_i' F_t + e_{it}$$

where $(1-\rho L)F_t = f_t$ and $(1-\rho L)e_{it} = \varepsilon_{it}$. When $\rho_i = 1$ for all i, we have $F_t = F_{t-1} + f_t$ and $e_{it} = e_{it-1} + \varepsilon_{it}$. In this case, both F_t and e_{it} are I(1). When $\rho_i = \rho$ with $|\rho| < 1$ for all i, we have $F_t = \rho F_{t-1} + f_t$ and $e_{it} = \rho e_{it-1} + \varepsilon_{it}$ so both F_t and e_{it} are I(0). Thus in DGP 2, the common and idiosyncratic components are restricted to have the same order of integration. Note that when ρ_i are heterogeneous, DGP 2 cannot be expressed in DGP 1.

Pesaran (2007) considered the DGP

$$\begin{aligned} X_{it} &= (1-\rho_i L)D_{it} + \rho_i X_{it-1} + X_{it}^0 \\ X_{it}^0 &= \lambda_i f_t + \varepsilon_{it}. \end{aligned}$$

Written in terms of DGP 1, this becomes $X_{it} = D_{it} + \lambda_i F_t + e_{it}$ with $F_t = F_{t-1} + f_t$ if $\rho_i = 1$, and $F_t = \rho_i F_{t-1} + f_t$ when $\rho_i < 1$. Apart from the treatment of the deterministic term, this is essentially DGP 2 and we do not consider it separately.

The construction of the test statistics based on defactored processes requires the short run, long run, and one-sided variance of ε_{it} defined as

$$\sigma_{\varepsilon i}^2 = E(\varepsilon_{it}^2), \quad \omega_{\varepsilon i}^2 = \lim_{T\to\infty} T^{-1} E(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{it})^2, \quad \lambda_{\varepsilon i} = (\omega_{\varepsilon i}^2 - \sigma_{\varepsilon i}^2)/2$$

respectively. As in Moon and Perron (2004), we assume the following limits exist

$$\omega_\varepsilon^2 = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \omega_{\varepsilon i}^2, \quad \sigma_\varepsilon^2 = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \sigma_{\varepsilon i}^2, \quad \lambda_\varepsilon = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \lambda_{\varepsilon i}, \quad \phi_\varepsilon^4 = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \omega_{\varepsilon i}^4.$$

For future reference, let

$$\widehat{\omega}_\varepsilon^2 = \frac{1}{N}\sum_{i=1}^{N}\widehat{\omega}_{\varepsilon i}^2, \qquad \widehat{\sigma}_\varepsilon^2 = \frac{1}{N}\sum_{i=1}^{N}\widehat{\sigma}_{\varepsilon i}^2, \qquad \widehat{\lambda}_\varepsilon = \frac{1}{N}\sum_{i=1}^{N}\widehat{\lambda}_{\varepsilon i}, \qquad \widehat{\phi}_\varepsilon^4 = \frac{1}{N}\sum_{i=1}^{N}\widehat{\omega}_{\varepsilon i}^4. \tag{1}$$

be consistent estimates of ω_ε^2, σ_ε^2, λ_ε, and ϕ_ε^4, respectively.

Throughout the paper, we maintain that Assumptions A-E of Bai and Ng (2004, page 1130) hold for F_t and e_{it}. These assumptions largely pertain to weak cross-sectional and weak serial correlations in ε_{it}. In particular, even if $E(\varepsilon_{it}\varepsilon_{jt}) \neq 0$ $(i \neq j)$, the number of factors and the factor processes can be consistently estimated so that defactored processes can be obtained. The assumptions of Moon and Perron (2004) are similar to those of BN, but require ε_{it} to be cross-sectionally independent. However, when coming to pool individual test statistics, BN also assumes cross-sectional independence of ε_{it}. Under the null hypothesis that $\rho_i = 1$ for all i, DGP 1 covers DGP 2. It follows that the assumptions of BN are also applicable to DGP 2 for the PANIC procedure. In this paper, the nonparametric kernel function is assumed to satisfy the requirements in Moon and Perron (2004) to permit estimation of the long-run and one-sided long run variances. These assumptions will not be restated here to focus on the main issues we want to highlight here.

3 PANIC Pooled Tests

In Bai and Ng (2004) we showed using DGP 1 that testing can still proceed even when both components are unobserved, and without knowing a priori whether e_{it} is non-stationary. The strategy is to obtain consistent estimates of the space spanned by F_t (denoted by $\widehat{F}_t$) and the idiosyncratic error (denoted by $\widehat{e}_{it}$)y. In a nutshell, we apply the method of principal components to the first differenced data, and then form $\widehat{F}_t$ and $\widehat{e}_{it}$ by re-cumulating the estimated factor components. More precisely, when D_{it} in DGP 1 is zero ($p = -1$) or an intercept ($i.e., p = 0$), the first difference of the model is

$$\Delta X_{it} = \lambda_i'\Delta F_t + \Delta e_{it}.$$

Denote $x_{it} = \Delta X_{it}$, $f_t = \Delta F_t$, and $z_{it} = \Delta e_{it}$. Then

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

is a pure factor model, from which we can estimate $(\widehat{\lambda}_1, ..., \widehat{\lambda}_N)$ and $(\widehat{f}_2, ..., \widehat{f}_T)$ and $\widehat{z}_{it}$ for all i and t. Define

$$\widehat{F}_t = \sum_{s=2}^{t}\widehat{f}_s \quad \text{and} \quad \widehat{e}_{it} = \sum_{s=2}^{t}\widehat{z}_{is}.$$

When $p = 1$, we also need to remove the mean of the differenced data, which is the slope coefficient in the linear trend prior to differencing. This leads to $x_{it} = \Delta X_{it} - \overline{\Delta X_i}$, $f_t = \Delta F_t - \overline{\Delta F}$, and $z_{it} = \Delta e_{it} - \overline{\Delta e_i}$, where $\overline{\Delta X_i}$ is the same mean of X_{it} over t, and $\overline{\Delta F}$, $\overline{\Delta e_i}$ are similarly defined.

Bai and Ng (2004) provide asymptotically valid procedures for (i) determining the number of stochastic trends in $\widehat{F}_t$, (ii) testing if $\widehat{e}_{it}$ are individually I(1) using augmented Dickey-Fuller regressions, and (iii) testing if the panel is I(1) based on the p values of tests derived from (ii) under the additional assumption that e_{it} is independent across i. If π_i is the p-value of the ADF test for the i-th cross-section unit, the pooled test is

$$P_{\widehat{e}} = \frac{-2\sum_{i=1}^{N} \log \pi_i - 2N}{\sqrt{4N}}.$$

The estimation procedure is identical when $p = -1, 0$ and 1.

The pooled test in Bai and Ng (2004) does not require a pooled OLS estimate of the AR(1) coefficient in the idiosyncratic errors. Pooling p values has the advantage that more heterogeneity in the units are permitted. However, a test based on a pooled estimate of ρ can be easily constructed by estimating a panel autoregression in the (cumulated) idiosyncratic errors estimated by PANIC, ie. $\widehat{e}_{it}$. Specifically, for $p = -1$ or $p = 0$, pooled OLS estimation of the model

$$\widehat{e}_{it} = \rho \widehat{e}_{it-1} + \varepsilon_{it}.$$

yields

$$\widehat{\rho} = \frac{\mathrm{tr}(\widehat{e}'_{-1}\widehat{e})}{\mathrm{tr}(\widehat{e}'_{-1}\widehat{e}_{-1})}$$

where $\widehat{e}_{-1}$ and $\widehat{e}$ are $(T-2) \times N$ matrices. For $p = 1$, we add an intercept and a linear trend in the above regression, so the pooled estimator is

$$\widehat{\rho} = \frac{\mathrm{tr}(\widehat{e}'_{-1} M_1 \widehat{e})}{\mathrm{tr}(\widehat{e}'_{-1} M_1 \widehat{e}_{-1})}$$

where M_1 is the usual projection matrix that demeans and detrends the data. The bias-corrected pooled PANIC autoregressive estimator for $\widehat{e}$ can be written as

$$\widehat{\rho}^{+} = \frac{\mathrm{tr}(\widehat{e}'_{-1} M_z \widehat{e} - NT\widehat{\psi})}{\mathrm{tr}(\widehat{e}'_{-1} M_z \widehat{e}_{-1})}$$

where $\widehat{\psi}$ is the bias correction estimated from $\widehat{\varepsilon} = M_z\widehat{e} - \widehat{\rho}M_z\widehat{e}_{-1}$. Let

$$P_a = \frac{\sqrt{N}T(\widehat{\rho}^+ - 1)}{\sqrt{K_a\widehat{\phi}_\varepsilon^4/\widehat{\omega}_\varepsilon^4}} \tag{2}$$

$$P_b = \sqrt{N}T(\widehat{\rho}^+ - 1)\sqrt{\frac{1}{NT^2}\mathrm{tr}(\widehat{e}_{-1}M_z\widehat{e}'_{-1})K_b\frac{\widehat{\omega}_\varepsilon^2}{\widehat{\phi}_\varepsilon^4}} \tag{3}$$

where $\widehat{\sigma}_\varepsilon^2$, $\widehat{\lambda}_\varepsilon$, $\widehat{\omega}_\varepsilon^2$, and $\widehat{\phi}_\varepsilon^4$ are defined in (1), and are estimated from the N series contained in matrix $\widehat{\varepsilon} = M_z\widehat{e} - \widehat{\rho}M_z\widehat{e}_{-1} = [\widehat{\varepsilon}_1, \widehat{\varepsilon}_2, ..., \widehat{\varepsilon}_N]$ with $\widehat{\varepsilon}_i$ being $(T-2)\times 1$ for all i. The bias adjustment and the tests depend on the treatment of the deterministic terms and are summarized below.

Parameters for pooled PANIC tests

p	M_z	$\widehat{\psi}$	K_a	K_b
-1, or 0	I_{T-2}	$\widehat{\lambda}_\varepsilon$	2	1
1	M_1	$-\widehat{\sigma}_\varepsilon^2/2$	15/4	4

where I_{T-2} is a $(T-2)\times(T-2)$ identity matrix.

Theorem 1 *Let $\widehat{\rho}^+$ be the bias-corrected pooled autoregressive coefficient for the idiosyncratic errors estimated by PANIC, where the true errors are assumed to be independent across i. Then as $N, T \to \infty$ with $N/T \to 0$, and under the null hypothesis that $\rho_i = 1$ for all i, $P_a \xrightarrow{d} N(0,1)$ and $P_b \xrightarrow{d} N(0,1)$.*

Theorem 1 shows that the t tests of the pooled autoregressive coefficient in the idiosyncratic errors are asymptotically normal. The proof is based on a sequential limit argument, although the result also holds jointly. The P_a and P_b are the analog of t_a and t_b of MP, except that (i) the tests are based on PANIC residuals and (ii) the method of 'defactoring' of the data is different from MP. By taking first differences of the data to estimate the factors, we also simultaneously remove the individual fixed effects. Thus when $p = 0$, the $\widehat{e}_{it}$ obtained from PANIC can be treated as though they come from a model with no fixed effect. It is also for this reason that in Bai and Ng (2004), the augmented Dickey-Fuller test for $\widehat{e}_{it}$ has a limiting distribution that depends on standard Brownian motions, and not its demeaned variant.

When $p = 1$, the adjustment parameters used in $P_{a,b}$ are, however, the same as those of $t_{a,b}$ of MP. In this case, the PANIC residuals $\widehat{e}_{it}$ have the property that $T^{-1/2}\widehat{e}_{it}$ converges to a Brownian bridge, and a Brownian bridge takes on the value of zero at the boundary.

In consequence, the Brownian motion component in the numerator of the autoregressive estimate vanishes. Our construction of $\widehat{\rho}^+$ uses the M_1 matrix to remove the degeneracy. Since our $\widehat{\rho}^+$ is obtained from a regression of $\widehat{e}_{it}$ on $\widehat{e}_{it-1}$, plus an individual specific constant and a trend, it is not surprising that the tests use the same adjustments as t_a and t_b.

3.1 The Pooled MSB

An important feature that distinguishes stationary from non-stationary processes is that their sample moments require different rates of normalization in order to be bounded asymptotically. In the univariate context, a simple test based on this idea is the test of Sargan and Bhargava (1983). If for a given i, $\Delta e_{it} = \varepsilon_{it}$ has mean zero, unit variance, and is serially uncorrelated, then $Z_i = T^{-2}\sum_{t=1}^T e_{it}^2 \Rightarrow \int_0^1 W_i(r)^2 dr$. However, if x_{it} is stationary, $Z_i = O_p(T^{-1})$. Stock (1990) developed the modified Sargan-Bhargava test (MSB) to allow $\varepsilon_{it} = \Delta e_{it}$ to be serially correlated with short and long run variance σ_i^2 and ω_i^2, respectively. In particular, if $\widehat{\omega}_i^2$ is an estimate of ω_i^2 that is consistent under the null and is bounded under the alternative, $MSB = Z_i/\widehat{\omega}_i^2 \Rightarrow \int_0^1 W_i^2(r)dr$ under the null and degenerates to zero under the alternative. Thus the null is rejected when the statistic is too small. This motivates the following simple panel non-stationarity test for the idiosyncratic errors, denoted the PMSB (pooled MSB). Let $\widehat{e}_{it}$ be obtained from PANIC:

$$PMSB = \frac{\sqrt{N}(\mathrm{tr}(\frac{1}{NT^2}\widehat{e}'\widehat{e}) - \widehat{\psi})}{\sqrt{\widehat{\phi}_\varepsilon^4/K_{msb}}}. \tag{4}$$

where $\widehat{\psi}$ estimates the asymptotic mean of $\frac{1}{NT^2}\mathrm{tr}(\widehat{e}'\widehat{e})$ and the denominator estimates its standard deviation. These parameters are given in the table below

Parameters for PMSB

p	$\widehat{\psi}$	K_{msb}
-1, or 0	$\widehat{\omega}_\varepsilon^2/2$	3
1	$\widehat{\omega}_\varepsilon^2/6$	45

The variables $\widehat{\omega}_\varepsilon^2$ and $\widehat{\phi}_\varepsilon^4$ are defined in (1) and their estimation is discussed below. The null hypothesis is rejected for small values of PMSB. We have the following result:

Theorem 2 *Let PMSB be defined as in (4). Under the null hypothesis that e_{it} are all non-stationary and are independent across i, then as $N, T \to \infty$ with $N/T \to 0$,*

$$PMSB \xrightarrow{d} N(0,1).$$

The theorem is again proved using a sequential limit argument. The result should hold jointly as $T, N \to \infty$ with $N/T \to 0$ by verifying the generalized Lindeberg conditions as stated in Phillips and Moon (1999). But the sequential asymptotic framework provides the intuition for the result. For a given i, $Z_i = T^{-2}\sum_{t=1}^{T}\widehat{e}_{it}^2$ converges in distribution to $\omega_i^2\int_0^1 V_i(r)^2dr$ when $p=1$, where V_i is a Brownian bridge. Demeaning these random variables and averaging over the i give the stated result.

Comparing the PMSB with $P_{a,b}$ tests when $p=1$ is of special interest. From Bai and Ng (2004), $\widehat{e}_{it} = e_{it} - e_{i1} - (e_{iT} - e_{i1})\frac{t-1}{T-1} + o_p(1)$, which has a time trend component. With the $P_{a,b}$ tests, an intercept and a time trend are included in the pooled OLS regression that estimates $\widehat{\rho}$ to remove the linear trend component $(e_{iT} - e_{i1})\frac{t-1}{T-1}$ in $\widehat{e}_{it}$. Such a detrending procedure is not necessary when pooling $\widehat{e}_{it}^2$ directly as is the case with the PSMB since the limit of $T^{-1/2}\widehat{e}_{it}$ is simply a Browinian bridge. The PMSB essentially exploits this fact. We will see that it has better finite sample properties than any of the t tests when $p=1$, at least with large T.

To implement PMSB, we need an estimate of ω_ε^2 and ϕ_ε^4, the averaged zero frequency spectrum and squared spectrum of Δe_{it}, respectively. A kernel estimate based on $\Delta\widehat{e}_{it}$, although consistent under the null hypothesis that all units are non-stationary, is degenerate under the specific alternative that all units are stationary. Accordingly, we estimate ω_i^2 using $\widehat{\varepsilon}_{it} = \widehat{e}_{it} - \widehat{\rho}\widehat{e}_{it-1}$, where $\widehat{\rho}$ is the pooled least squares estimate by regressing $\widehat{e}_{it}$ on $\widehat{e}_{it-1}$, for $p=-1,0$. If $p=1$, we add an intercept and linear trend in the regression of $\widehat{e}_{it}$ on its lag to obtain $\widehat{\varepsilon}_{it}$, from which we estimate $\widehat{\omega}_{\varepsilon i}^2$ for each i, and obtain $\widehat{\omega}_\varepsilon^2$ by averaging over i. Averaging $\widehat{\omega}_{\varepsilon i}^4$ will give $\widehat{\phi}_\varepsilon^4$, see (1).

4 Improved MP Tests

Moon and Perron (2004) considered estimates of ρ obtained from

$$X_{it} = D_{it} + \rho X_{it-1} + u_{it}.$$

Their Model A assumes D_{it} is null. Model B assumes $D_{it} = a_i$, and Model C has $D_{it} = a_i + b_i t$. Let $\Lambda = (\lambda_1, \ldots \lambda_N)'$, X and X_{-1} be $T-1$ by N matrices. Based on the first step estimator $\widehat{\rho} = \frac{\mathrm{tr}(X'_{-1}M_zX)}{\mathrm{tr}(X'_{-1}M_zX_{-1})}$, one computes the residuals $\widehat{u} = M_zX - \widehat{\rho}M_zX_{-1}$, from which a factor model is estimated to obtain $\widehat{\Lambda} = (\widehat{\lambda}_1, ..., \widehat{\lambda}_N)'$. The bias-corrected, defactored pooled OLS estimator defined in MP is

$$\widehat{\rho}^+ = \frac{\mathrm{tr}(X'_{-1}M_zXM_{\widehat{\Lambda}} - NT\widehat{\psi}_\varepsilon)}{\mathrm{tr}(X'_{-1}M_zX_{-1}M_{\widehat{\Lambda}})}$$

where $\widehat{\psi}_\varepsilon$ is a bias correction term defined on the residuals of the defactored data $\widehat{\varepsilon} = [M_z X - \widehat{\rho} M_z X_{-1}] M_{\widehat{\Lambda}}$. The MP tests, denoted t_a and t_b, are as defined in (2) for P_a and (3) for P_b, with X and X_{-1} replacing $\widehat{e}$ and $\widehat{e}_{-1}$ both in $\widehat{\rho}^+$ and in the tests.

Apart from the way that the factors are controlled before testing, $P_{a,b}$ and $t_{a,b}$ also differ in that the deterministic terms in the MP framework are removed by least squares estimation of the incidental parameters, while PANIC takes a first differencing approach. The parameters for Models A and C are the same as $p = -1$ and 1 for $P_{a,b}$, respectively. However, the adjustment terms used in $P_{a,b}$ when $p = 0$ coincide with those used in MP for Model A (no incidental parameters). Moon and Perron (2004) used Model A (i.e., $M_z = I_{T-1}$) even when $p = 0$, as the tests are apparently consistent when the fixed effects are ignored. For completeness, the parameters for the MP tests under Model B (demeaning is performed) are $M_z = M_0$ (the projection matrix of demeaning), $\widehat{\psi}_\varepsilon = -\widehat{\sigma}^2_\varepsilon/2, K_a = 3$, and $K_b = 2$.

The bias corrected estimator and the t tests require consistent estimates of the averages of the long and short run variances of the ε_{it}, the residuals of the defactored model. MP estimated these by averaging over $\widehat{\omega}^2_{\varepsilon i}$ and $\widehat{\sigma}^2_{\varepsilon i}$, where these latter are defined using $\widehat{\varepsilon} = \widehat{u} M_{\widehat{\Lambda}}$. Importantly, $\widehat{\omega}_{\varepsilon i}$ is a function of $\widehat{\rho}$ which, is biased. However, the unbiased $\widehat{\rho}^+$ itself depends on $\widehat{\omega}_{\varepsilon i}$. This problem can be remedied by iterating $\widehat{\rho}^+$ till convergence is achieved.

<u>An Iterated MP:</u> Initialize $\widehat{\rho}_0 = \widehat{\rho}$

1 Estimate $\widehat{\omega}^2_\varepsilon, \widehat{\phi}^4_\varepsilon$ and $\widehat{\lambda}_\varepsilon$ from $\widehat{\varepsilon}_t(\widehat{\rho}_0)$.

2 Let $\widehat{\rho}_1 = \widehat{\rho}_0 - NT\psi(\widehat{\rho}_0)/\mathrm{tr}(X'_{-1} M_z X_{-1} M_{\widehat{\Lambda}})$

3 if $|\widehat{\rho}_1 - \widehat{\rho}_0|$ exceeds a tolerance level, $\widehat{\rho}_0 = \widehat{\rho}_1$ and return to step 1. Otherwise $\widehat{\rho}^+ = \widehat{\rho}_1$ and construct t_a, t_b.

To keep the tests distinct, we will refer to these tests that iterate on $\widehat{\rho}$ as $It_{a,b}$. Notably, iterated tests using the PANIC residuals can likewise be defined.

Because the MP tests are applied directly to the observable series, many researchers infer that t_a and t_b are testing the observed panel of data. It is worth re-iterating that after the common factors are controlled for, one must necessarily be testing the properties of the idiosyncratic errors. This is clearly true for DGP 2 since both the common and idiosyncratic terms have the same order of integration. Although less obvious, the statement is also true for DGP 1, in which e_{it} and F_t are not constrained to have the same order of integration. To see this, assume no deterministic component for simplicity. DGP 1 can be rewritten as

$$X_{it} = \rho_i X_{it-1} + \lambda'_i F_t - \rho_i \lambda'_i F_{t-1} + \varepsilon_{it}. \tag{5}$$

Since the defactored approach will remove the common factors, we can ignore them in the equations. It is then obvious that the MP tests will determine if the average of ρ_i is unity. The same holds for the test of Pesaran (2007), who estimates augmented autoregressions of the form (suppressing deterministic terms for simplicity and adapted to our notation)

$$\Delta X_{it} = (\rho_i - 1)X_{it-1} + d_0\bar{X}_t + d_1\Delta\bar{X}_t + e_{it},$$

where $\bar{X}_t = \frac{1}{N}\sum_{i=1}^{N} X_{it}$. Although $\Delta\bar{X}_t$ is observed, it plays the same role as $(1 - \rho_i F_t)$. As such, the CADF, which takes an average of the t ratios on ρ_i, is also a test of whether the idiosyncratic errors are non-stationary. Others control for cross-sectional correlation by adjusting the standard errors of the pooled estimate. But the method still depends on whether the factors and/or the errors are non-stationary, see Breitung and Das (2008). To make statements concerning non-stationarity for the observed series X_{it}, researchers still have to separately test if the common factors are non-stationary. As far as we are aware of, PANIC remains the only framework that can establish if the components are stationary in a coherent manner.

5 Finite Sample Properties

In this section, we compare the properties of three autoregressive coefficient based tests, ($t_{a,b}$, $It_{a,b}$, $P_{a,b}$) with $P_{\hat{e}}$ and PMSB. As $p = -1$ is not usually a case of practical interest, we only report results for $p = 0$ and $p = 1$. For the MP tests, we follow Moon and Perron (2004) and use Model A for testing (i.e. no demeaning) even when $p = 0$. To make this clear, we denote the results by $t_{a,b}^A$. Jang and Shin (2005) explored the sensitivity of the MP tests to the method of demeaning, but did not consider the case of incidental trends. Furthermore, they averaged the t tests of the PANIC residuals, rather than pool the p values as in Bai and Ng (2004). Gengenbach et al. (2006) also compared the MP tests with PANIC, but also for $p = 0$ only. As well, all these studies consider alternatives with little variation in the dynamic parameters. Here, we present new results by focusing on mixed I(1)/I(0) units, and with more heterogeneity in the dynamic parameters. We report results for 4 models. Additional results are available on request.

Models 1 and 3 are configurations of DGP 1 while Model 4 is a configuration of DGP 2. The common parameters are $r = 1$, λ_i is drawn from the uniform distribution such that $\lambda_i \sim U[-1, 3]$; $\eta_t \sim N(0, 1)$, $\varepsilon_{it} \sim N(0, 1)$. The model specific parameters are:

1 $\Phi_1 = 1$, $\rho_i = 1$ for all i;

2 $\Phi_1 = .5$, $\rho_i \sim U[.9, .99]$;

3 $\Phi_1 = .5$, $\rho_i = 1$ for $i = 1, \ldots N/5$, $\rho_i \sim U[.9, .99]$ otherwise;

4 $\rho_i \sim U[.9, .99]$;

where Φ_1 is the AR coefficient in $F_t = \Phi_1 F_{t-1} + \eta_t$. We consider combinations of N, T taking on values of 20, 50, and 100. In practice, the number of factors is not known; it can be consistently estimated using the criteria developed in Bai and Ng (2002). We hold the number of factors to the true value. This seems appropriate when evaluating the adequacy of the asymptotic approximations because the theory does not incorporate sampling variability due to the number of factors. The number of replications is 5000.

Results are reported in Table 1 and for $p = 0$, and 1 respectively. The rejection rates of Model 1 correspond to finite sample size when the nominal size is 5%. Models 2, 3 and 4 report power. The power is not size-adjusted to focus on rejection rates that one would obtain in practice. Table 1 shows that for $p = 0$, there is little gain from iterating $\widehat{\rho}$. The $P_{\widehat{e}}, t_{a,b}, It_{a,b}$ have size distortion when T is small. The PANIC based PMSB and $P_{a,b}$ seem to have better size properties. Apart from size discrepancies when T is small, all tests have similar properties.

As mentioned earlier, Moon and Perron (2004) used Model A for testing even when $p = 0$. In results not reported, the MP tests are found to have huge size distortions once the fixed effects are explicitly removed. The iterative tests yield significant size improvements but at the cost of power. Our conjecture is that the MP tests are generally vulnerable to the presence of incidental parameters. As will be clear from the results for $p = 1$, this power problem applies to autoregressive coefficient based tests, and is not a consequence of the way MP defactored the data.

Results for $p = 1$ in Table 2 show that all tests based on a pooled estimate of ρ (i.e. $P_{a,b}, t_{a,b}$ and $It_{a,b}$) are grossly oversized. The most accurate size is achieved by $It_{a,b}$ only when T=100, at which point it has no power. This corroborates the analytical results of Moon et al. (2005), who find the MP tests have no local power against the alternative of incidental trends. Our simulations suggest that this loss of power is more general and applies to the PANIC based $P_{a,b}$ as well. Of all the tests considered, only the new PMSB and $P_{\widehat{e}}$, ie. the original PANIC test that pools p-values, have any power against local alternatives. Assuming cross-section independence, Phillips and Ploberger (2002) proposed a panel unit root test in the presence of incidental trends that maximizes average power. It has some resemblance to the Sargan-Bhargava test. Although optimality of the PMSB is not shown here, the PMSB does appear to have better finite sample properties in the presence of incidental trends, even when it is applied to the idiosyncratic errors. However, both the PMSB and pooling p values require that N not to be too small, or else the test is oversized. Overall, tests of non-stationarity in panel data with incidental trends are quite unreliable without N and T being reasonably large.

Incidental parameters clearly create challenging problems for unit root testing using panel data. While MP estimates the deterministic terms by OLS, PANIC takes a first differencing approach. The question arises as to whether alternative methods of detrending might help. In unreported simulations, the finite sample size and power of the tests under GLS detrending are similar to those in Tables 1 when $p = 0$. When $p = 1$, the autoregressive coefficient based tests continue to have limited power. Even after considering GLS detrending, PMSB and $P_{\hat{e}}$ remain useful when $p = 1$.

In a recent paper, Westerlund and Larsson (2007) provide a detailed analysis of pooled PANIC test $P_{\hat{e}}$. Their justification of the procedure is more rigorous than what was given in Bai and Ng (2004). It should be stressed that $P_{\hat{e}}$ is *not* the only way to construct a pooled test in the PANIC framework. As we showed in this paper, $P_{a,b}$ and PMSB are PANIC based pooled tests. Our simulations show that PMSB and $P_{\hat{e}}$ have good finite sample properties.

6 Conclusion

In this paper, we (i) develop a PANIC based estimate of the pooled autoregressive coefficient, (ii) develop a PMSB test that does not rely on the polled autoregressive coefficient, and (iii) provide an iterative approach that can improve the size of the MP tests. Upon comparing their finite sample properties, we find that tests based on the autoregressive coefficient have no power against incidental trends. The PMSB and the original PANIC pooled test of the p values have better power properties.

It should again be emphasized that tests which control cross-section correlation only permit hypothesis concerning the idiosyncratic errors to be tested. To decide if the observed data are stationary or not, we still need the PANIC procedure to see if the factors are stationary. In fact, PANIC goes beyond unit root testing by showing that the common stochastic trends are well defined and can be consistently estimated even if e_{it} are I(1) for all i. This is in contrast with a fixed N spurious system in which common trends are meaningless.

Table 1: $p = 0$

N	T	$P_{\widehat{e}}$	$PMSB$	P_a	P_b	It_a^A	It_b^A	t_a^A	t_b^A
		Model 1: $F \sim I(1), e_{it} \sim I(1)$							
20	20	0.206	0.016	0.094	0.053	0.143	0.104	0.145	0.105
20	50	0.073	0.028	0.097	0.057	0.100	0.066	0.101	0.066
20	100	0.057	0.025	0.101	0.060	0.101	0.064	0.101	0.063
50	20	0.278	0.057	0.076	0.053	0.168	0.144	0.169	0.146
50	50	0.074	0.043	0.076	0.055	0.084	0.057	0.084	0.058
50	100	0.059	0.042	0.083	0.061	0.085	0.064	0.085	0.064
100	20	0.400	0.107	0.081	0.061	0.189	0.173	0.192	0.178
100	50	0.067	0.060	0.071	0.054	0.075	0.059	0.075	0.058
100	100	0.058	0.055	0.075	0.058	0.072	0.059	0.072	0.058
		Model 2: $F \sim I(0), e_{it} \sim I(0)$							
20	20	0.570	0.359	0.799	0.677	0.920	0.890	0.965	0.937
20	50	0.879	0.979	1.000	0.999	0.978	0.978	1.000	1.000
20	100	1.000	1.000	1.000	1.000	0.985	0.985	1.000	1.000
50	20	0.854	0.875	0.966	0.944	0.947	0.946	0.999	0.998
50	50	0.997	1.000	1.000	1.000	0.979	0.979	1.000	1.000
50	100	1.000	1.000	1.000	1.000	0.987	0.987	1.000	1.000
100	20	0.974	0.992	0.998	0.996	0.944	0.944	1.000	1.000
100	50	1.000	1.000	1.000	1.000	0.976	0.976	1.000	1.000
100	100	1.000	1.000	1.000	1.000	0.988	0.988	1.000	1.000
		Model 3: $F \sim I(0), e_{it}$ mixed I(0), I(1)							
20	20	0.472	0.197	0.547	0.405	0.838	0.787	0.879	0.828
20	50	0.738	0.743	0.922	0.875	0.952	0.932	0.969	0.949
20	100	0.995	0.928	0.979	0.963	0.974	0.961	0.981	0.968
50	20	0.751	0.610	0.780	0.687	0.938	0.928	0.979	0.970
50	50	0.970	0.984	0.996	0.992	0.977	0.977	0.999	0.998
50	100	1.000	0.999	1.000	1.000	0.990	0.989	1.000	1.000
100	20	0.929	0.903	0.940	0.905	0.956	0.956	0.998	0.998
100	50	1.000	1.000	1.000	1.000	0.982	0.982	1.000	1.000
100	100	1.000	1.000	1.000	1.000	0.988	0.988	1.000	1.000
		Model 4: $F \sim I(0), e_{it} \sim I(0)$							
20	20	0.498	0.286	0.715	0.590	0.801	0.719	0.818	0.734
20	50	0.785	0.922	0.984	0.971	0.972	0.966	0.987	0.981
20	100	0.996	0.994	0.999	0.998	0.989	0.989	0.999	0.998
50	20	0.777	0.780	0.901	0.858	0.913	0.893	0.934	0.913
50	50	0.959	0.973	0.983	0.978	0.978	0.973	0.988	0.982
50	100	1.000	0.997	0.998	0.997	0.992	0.991	0.998	0.997
100	20	0.925	0.961	0.961	0.949	0.952	0.944	0.968	0.961
100	50	0.994	0.996	0.994	0.993	0.989	0.987	0.997	0.995
100	100	1.000	0.999	1.000	0.999	0.995	0.995	1.000	1.000

Table 2: $p = 1$

N	T	$P_{\widehat{e}}$	$PMSB$	P_a	P_b	It_a^B	It_b^B	t_a^B	t_b^B
		Model 1: $F \sim I(1), e_{it} \sim I(1)$							
20	20	0.385	0.000	1.000	1.000	0.140	0.140	0.981	0.983
20	50	0.079	0.011	1.000	1.000	0.062	0.060	0.366	0.346
20	100	0.063	0.020	1.000	1.000	0.047	0.047	0.152	0.139
50	20	0.594	0.000	1.000	1.000	0.226	0.217	1.000	1.000
50	50	0.069	0.010	1.000	1.000	0.061	0.057	0.588	0.566
50	100	0.054	0.022	1.000	1.000	0.043	0.043	0.238	0.227
100	20	0.793	0.000	1.000	1.000	0.322	0.312	1.000	1.000
100	50	0.068	0.007	1.000	1.000	0.150	0.143	0.833	0.818
100	100	0.062	0.026	1.000	1.000	0.050	0.049	0.368	0.354
		Model 2: $F \sim I(0), e_{it} \sim I(0)$							
20	20	0.399	0.000	1.000	1.000	0.176	0.170	0.993	0.994
20	50	0.171	0.131	1.000	1.000	0.079	0.052	0.708	0.652
20	100	0.644	0.835	1.000	1.000	0.016	0.006	0.332	0.214
50	20	0.627	0.000	1.000	1.000	0.255	0.241	1.000	1.000
50	50	0.256	0.316	1.000	1.000	0.159	0.114	0.915	0.889
50	100	0.924	0.992	1.000	1.000	0.036	0.013	0.692	0.598
100	20	0.842	0.000	1.000	1.000	0.312	0.299	1.000	1.000
100	50	0.444	0.653	1.000	1.000	0.298	0.228	0.994	0.992
100	100	0.998	1.000	1.000	1.000	0.055	0.018	0.915	0.873
		Model 3: $F \sim I(0), e_{it}$ mixed I(0), I(1)							
20	20	0.394	0.000	1.000	1.000	0.182	0.174	0.994	0.995
20	50	0.149	0.089	1.000	1.000	0.098	0.067	0.710	0.667
20	100	0.514	0.581	1.000	1.000	0.033	0.015	0.385	0.262
50	20	0.621	0.000	1.000	1.000	0.260	0.250	1.000	1.000
50	50	0.204	0.188	1.000	1.000	0.200	0.148	0.917	0.899
50	100	0.794	0.891	1.000	1.000	0.090	0.046	0.754	0.675
100	20	0.839	0.000	1.000	1.000	0.326	0.314	1.000	1.000
100	50	0.320	0.359	1.000	1.000	0.388	0.311	0.995	0.993
100	100	0.975	0.995	1.000	1.000	0.172	0.094	0.947	0.927
		Model 4: $F \sim I(0), e_{it} \sim I(0)$							
20	20	0.393	0.000	1.000	1.000	0.121	0.121	0.985	0.988
20	50	0.143	0.098	1.000	1.000	0.028	0.023	0.380	0.324
20	100	0.506	0.647	1.000	1.000	0.010	0.007	0.148	0.087
50	20	0.621	0.000	1.000	1.000	0.197	0.190	1.000	1.000
50	50	0.198	0.214	1.000	1.000	0.023	0.018	0.608	0.563
50	100	0.747	0.841	1.000	1.000	0.016	0.014	0.245	0.182
100	20	0.829	0.000	1.000	1.000	0.280	0.270	1.000	1.000
100	50	0.332	0.441	1.000	1.000	0.066	0.055	0.824	0.796
100	100	0.936	0.942	1.000	1.000	0.022	0.017	0.407	0.322

Appendix

Lemma 1 *Let $C_{NT} = \min[\sqrt{N}, \sqrt{T}]$. The PANIC residuals $\widehat{e}_{it}$ satisfy, for $p = -1, 0$,*

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{e}_{it}^2 = \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}^2 + O_p(C_{NT}^{-2})$$

Proof. From Bai and Ng (2004, page 1154), $\widehat{e}_{it} = e_{it} - e_{i1} + \lambda_i' H^{-1} V_t - d_i'\widehat{F}_t$ where $V_t = \sum_{s=2}^{t} v_s$, $v_t = \widehat{f}_t - Hf_t$, and $d_i = \widehat{\lambda}_i - H^{-1\prime}\lambda_i$. Rewrite the above as $\widehat{e}_{it} = e_{it} + A_{it}$ with $A_{it} = -e_{i1} + \lambda_i' H^{-1} V_t - d_i'\widehat{F}_t$. Thus $\widehat{e}_{it}^2 = e_{it}^2 + 2e_{it}A_{it} + A_{it}^2$. It follows that

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{e}_{it}^2 = \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}^2 + 2\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}A_{it} + \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}A_{it}^2 = I + II + III \tag{6}$$

Bai and Ng (2004, p. 1163) show that $\frac{1}{T^2}\sum_{t=1}^{T} A_{it}^2 = O_p(C_{NT}^{-2})$ for each i. Averaging over i, it is still this order of magnitude. In fact, by the argument of BN,

$$III = \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}A_{it}^2 \le 3\frac{1}{T}(\frac{1}{N}\sum_{i=1}^{N}e_{i1}^2) + 3(\frac{1}{T^2}\sum_{t=1}^{T}\|V_t\|^2)(\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i H^{-1}\|^2)$$

$$+(\frac{1}{N}\sum_{i=1}^{N}\|d_i\|^2)\frac{1}{T^2}\sum_{t=1}^{T}\|\widehat{F}_t\|^2 = O_p(T^{-1}) + O_p(N^{-1}) + O_p([\min[N^2, T]^{-1}) = O_p(C_{NT}^{-2}).$$

Next, consider II, ignore the factor 2 and from the definition of A_{it},

$$II = -\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}e_{i1} + \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}\lambda_i' H^{-1}V_t - \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}d_i'\widehat{F}_t = a + b + c$$

The proof of $a = O_p(C_{NT}^{-2})$ is easy and is omitted (one can even assume $e_{i1} = 0$). Consider b.

$$\|b\| \le \|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_i e_{it}\|\|H^{-1}V_t\|$$

$$\le \Big(\frac{1}{T^2}\sum_{t=1}^{T}\|N^{-1/2}\sum_{i=1}^{N}\lambda_i e_{it}\|^2\Big)^{1/2} N^{-1/2}\Big(\frac{1}{T^2}\sum_{t=1}^{T}\|H^{-1}V_t\|^2\Big)^{1/2}$$

By (A.4) of BN (2004, page 1157), $N^{-1/2}\Big(\frac{1}{T^2}\sum_{t=1}^{T}\|H^{-1}V_t\|^2\Big)^{1/2} = O_p(N^{-1}) = O_p(C_{NT}^{-2})$. The first expression is $O_p(1)$ because $(NT)^{-1/2}\sum_{i=1}^{N}\lambda_i e_{it} = O_p(1)$. Thus $b = O_p(C_{NT}^{-2})$. Consider c.

$$\|c\| \le \frac{1}{\sqrt{T}}\Big(\frac{1}{T}\sum_{t=1}^{T}\Big[\frac{1}{N}\sum_{i=1}^{N}e_{it}d_i\Big]^2\Big)^{1/2}\Big(\frac{1}{T^2}\sum_{t=1}^{T}\|\widehat{F}_t\|^2\Big)^{1/2} = O_p(T^{-1/2})\Big(\frac{1}{T}\sum_{t=1}^{T}\Big[\frac{1}{N}\sum_{i=1}^{N}e_{it}d_i\Big]^2\Big)^{1/2}$$

Using equation (B.2) of Bai (2003), i.e.,

$$d_i = H\frac{1}{T}\sum_{s=1}^{T} f_s\varepsilon_{is} + O_p(C_{NT}^{-2}) \tag{7}$$

and ignoring H for simplicity, we have

$$\frac{1}{N}\sum_{i=1}^{N} e_{it}d_i = \frac{1}{N}\sum_{i=1}^{N} e_{it}\frac{1}{T}\sum_{s=1}^{T} f_s\varepsilon_{is} + T^{1/2}O_p(C_{NT}^{-2})$$

noting that $e_{it} = T^{1/2}O_p(1)$. If we can show for each t,

$$E\Big(\frac{1}{N}\sum_{i=1}^{N} e_{it}\frac{1}{T}\sum_{s=1}^{T} f_s\varepsilon_{is}\Big)^2 = O(T^{-1}) + O(N^{-1}) \tag{8}$$

then $c = O_p(T^{-1/2})[O_p(T^{-1/2}) + O_p(N^{-1/2}) + T^{1/2}O_p(C_{NT}^{-2})] = O_p(C_{NT}^{-2})$. But the above is proved for the case of $t = T$ below; the argument is exactly the same for every t. In summary $II = O_p(C_{NT}^{-2})$. This proves the lemma.

Lemma 2 *If $N/T^2 \to 0$, then the PANIC residuals $\widehat{e}_{it}$ satisfy, for $p = -1, 0$*

$$\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=1}^{T} \widehat{e}_{it-1}\Delta\widehat{e}_{it} = \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it-1}\Delta e_{it} + o_p(1)$$

Proof. Using the identity $\frac{1}{T}\sum_{t=2}^{T}\widehat{e}_{it-1}\Delta\widehat{e}_{it} = \frac{1}{2T}\widehat{e}_{iT}^2 - \frac{1}{2T}\widehat{e}_{i1}^2 - \frac{1}{2T}\sum_{t=2}^{T}(\Delta\widehat{e}_{it})^2$ and the corresponding identity for $\frac{1}{T}\sum_{t=2}^{T} e_{it-1}\Delta e_{it}$, then (ii) is a consequence of the lemma below.

Lemma 3 *(i)* $\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}(\widehat{e}_{i1}^2 - e_{i1}^2) = o_p(1)$
(ii) $\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}(\widehat{e}_{iT}^2 - e_{iT}^2) = o_p(1)$
(iii) $\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T}[(\Delta\widehat{e}_{it})^2 - (\Delta e_{it})^2] = o_p(1)$

Proof of (i). Since $\widehat{e}_{i1}$ is defined to be zero, it follows that the left hand side of (i) is $(\sqrt{N}/T)(\frac{1}{N}\sum_{i=1}^{N} e_{i1}^2) = o_p(1)$ if $\sqrt{N}/T \to 0$.

Proof of (ii). From $\widehat{e}_{iT} = e_{iT} + A_{iT}$, it is sufficient to show the expression (a) $\frac{1}{\sqrt{N}T}\sum_{i=1}^{N} A_{iT}^2 = o_p(1)$ and that (b) $\frac{1}{\sqrt{N}T}\sum_{i=1}^{N} e_{iT}A_{iT} = o_p(1)$. Using $\|V_T\|^2/T = O_p(N^{-1})$ and $d_i = O_p(1/\min[\sqrt{T}, N])$, it is easy to show that the expression in (a) is $O_p(\sqrt{N}/T)$. Consider (b).

$$\frac{1}{\sqrt{N}T}\sum_{i=1}^{N} e_{iT}A_{iT} = \frac{-1}{\sqrt{N}T}\sum_{i=1}^{N} e_{iT}e_{i1} + \frac{1}{\sqrt{N}T}\sum_{i=1}^{N} e_{iT}\lambda_i' H^{-1}V_T - \frac{1}{\sqrt{N}T}\sum_{i=1}^{N} e_{iT}d_i'\widehat{F}_T \tag{9}$$

The first term on the right hand side can be shown to be $O_p(T^{-1/2})$. Consider the second term. Under $\rho_i = 1$, $e_{iT} = \sum_{t=1}^{T} \varepsilon_{it}$,

$$\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} e_{iT}\lambda_i' H^{-1} V_T\| \le \frac{\|V_T\|\|H^{-1}\|}{\sqrt{T}}\|\frac{1}{\sqrt{NT}}\Big(\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_i\varepsilon_{it}\Big)\| = O_p(1)\frac{\|V_T\|}{\sqrt{T}} = O_p(C_{NT}^{-1})$$

see Bai and Ng (2004, page 1157). For the last term of (9), from $T^{-1/2}\widehat{F}_T = O_p(1)$, we need

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} e_{iT} d_i = \Big(\frac{N}{T}\Big)^{1/2}\big(\frac{1}{N}\sum_{i=1}^{N} e_{iT} d_i\big) = o_p(1). \tag{10}$$

From $d_i = O_p(\min[N, \sqrt{T}]^{-1})$, we have $N^{-1}\sum_{i=1}^{N} e_{iT} d_i = O_p(1)$. Thus the above is $o_p(1)$ if $N/T \to 0$. But $N/T \to 0$ is not necessary. To see this, using d_i in (7),

$$\frac{1}{N}\sum_{i=1}^{N} e_{iT} d_i = \frac{1}{N}\sum_{i=1}^{N} e_{iT}\frac{1}{T}\sum_{s=1}^{T} f_s\varepsilon_{is} + O_p(T^{1/2})O_p(C_{NT}^{-2}). \tag{11}$$

We shall show

$$E\Big(\frac{1}{N}\sum_{i=1}^{N} e_{iT}\frac{1}{T}\sum_{s=1}^{T} f_s\varepsilon_{is}\Big)^2 = O(T^{-1}) + O(N^{-1}) \tag{12}$$

From $e_{iT} = \sum_{t=1}^{T}\varepsilon_{it}$, the left hand side above is

$$\frac{1}{N^2}\sum_{i,j}^{N}\frac{1}{T^2}\sum_{s,k,t,h}^{T} E(f_s f_k \varepsilon_{is}\varepsilon_{it}\varepsilon_{jk}\varepsilon_{jh}) \tag{13}$$

Consider $i \neq j$, from cross-sectional independence and the independence of factors with the idiosyncratic errors, $E(f_s f_k \varepsilon_{is}\varepsilon_{it}\varepsilon_{jk}\varepsilon_{jh}) = E(f_s f_k)E(\varepsilon_{is}\varepsilon_{it})E(\varepsilon_{jk}\varepsilon_{jh})$. To see the key idea, assume ε_{it} are serially uncorrelated, then $E(\varepsilon_{is}\varepsilon_{it}) = E(\varepsilon_{it}^2)$ for $s = t$ and 0 otherwise. Similarly, $E(\varepsilon_{jk}\varepsilon_{jh}) = E(\varepsilon_{jk}^2)$ for $h = k$ and 0 otherwise. We assume $E(\varepsilon_{it}^2) \le \sigma_i^2$ for all t. Thus terms involving $i \neq j$ have an upper bound

$$\frac{1}{N^2}\sum_{i\neq j}^{N}\sigma_i^2\sigma_j^2\frac{1}{T^2}\sum_{s,k}^{T}|E(f_s f_k)| = O(T^{-1})$$

since $T^{-1}\sum_{s,k}^{T}|E(f_s f_k)| \le M$ under weak correlation for f_s. If ε_{it} is serially correlated, then the sum in (13) for $i \neq j$ is bounded by

$$\frac{1}{N^2}\sum_{i\neq j}^{N}\Big(\frac{1}{T^2}\sum_{s,k}^{T}|E(f_s f_k)|\Big)\Big(\max_s\sum_{t=1}^{T}|E(\varepsilon_{is}\varepsilon_{it})|\Big)\Big(\max_k\sum_{h=1}^{T}|E(\varepsilon_{jk}\varepsilon_{jh})|\Big)$$

$$\leq \frac{1}{N^2}\sum_{i\neq j}^{N}\Big(\frac{1}{T}\sum_{s,k}^{T}|E(f_s f_k)|\Big)\Big(\sum_{\ell=0}^{\infty}|\gamma_i(\ell)|\Big)\Big(\sum_{\ell=0}^{\infty}|\gamma_j(\ell)|\Big)$$

where $\gamma_i(\ell)$ is the autocovariance of ε_{it} at lag ℓ, and $\gamma_j(\ell)$ is similarly defined. Replace σ_i^2 by $\sum_{\ell=0}^{\infty}|\gamma_i(\ell)|<\infty$ (and similarly for σ_j^2), the same conclusion holds.

Next consider the case of $i=j$. Then since $\frac{1}{T^2}\sum_{s,t,k,h}^{T}E(f_s f_k \varepsilon_{is}\varepsilon_{it}\varepsilon_{ik}\varepsilon_{ih})=O(1)$, we have $\frac{1}{N^2}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{s,t,k,h}^{T}E(f_s f_k \varepsilon_{is}\varepsilon_{it}\varepsilon_{ik}\varepsilon_{ih})=O(N^{-1})$, proving (12). Combining (10) to (12) and using $N/T^2\to 0$,

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}e_{iT}d_i=\Big(\frac{N}{T}\Big)^{1/2}[O_p(T^{-1/2})+O_p(N^{-1/2})+T^{1/2}O_p(C_{NT}^{-2})]=o_p(1).$$

Proof of (iii). From $\Delta\widehat{e}_{it}=\Delta e_{it}-a_{it}$, where $a_{it}=\lambda_i' H^{-1\prime}v_t+d_i'\widehat{f}_t$, we have

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}[(\Delta\widehat{e}_{it})^2-(\Delta e_{it})^2]=-\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}(\Delta e_{it})a_{it}+\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}a_{it}^2$$

From BN (page 1158), $T^{-1}\sum_{t=2}^{T}a_{it}^2=O_p(C_{NT}^{-2})$, thus the second term on the right hand side above is bounded by $\sqrt{N}O_p(C_{NT}^{-2})=o_p(1)$. Consider the first term

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}(\Delta e_{it})a_{it}=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}(\Delta e_{it})\lambda_i' H^{-1\prime}v_t+\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}(\Delta e_{it})d_i'\widehat{f}_t=I+II$$

By the Cauchy-Schwartz inequality,

$$I\leq\|H^{-1}\|\Big(\frac{1}{T}\sum_{t=2}^{T}\|N^{-1/2}\sum_{i=1}^{N}\Delta e_{it}\lambda_i\|\Big)^{1/2}\Big(\frac{1}{T}\sum_{t=2}^{T}\|v_t\|^2\Big)^{1/2}=O_p(1)O_p(C_{NT}^{-1})$$

For II, it can be shown that it suffices to show $II=o_p(1)$ when $\widehat{f}_t$ is replaced by f_t. Now

$$\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}(\Delta e_{it})d_i' f_t\|\leq T^{-1/2}\Big(\sum_{i=1}^{N}d_i^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\sum_{t=2}^{T}\Delta e_{it}f_t\|^2\Big)^{1/2}.$$

The above is $T^{-1/2}\Big(\sum_{i=1}^{N}d_i^2\Big)^{1/2}O_p(1)=O_p(\sqrt{N}/T)$, following from $\sum_{i=1}^{N}d_i^2=O_p(N/\min(N^2,T))$. This proves (iii).

Remark 3 *For $p=1$ (linear trend model), Lemmas 1 and 2 hold when e_{it} on the right hand side is replaced by $\widetilde{e}_{it}=e_{it}-\frac{e_{iT}-e_{i1}}{T-1}(t-1)$. However, these results are not directly useful since with the linear trend model, we use the residuals from projecting $\widehat{e}_{it}$ on $[1,t]$.*

Let $\widehat{e}_{it}^{\tau}$ denote the regression residual from regressing $\widehat{e}_{it}$ on a constant and a linear trend. We define e_{it}^{τ} similarly.

Lemma 4 *For $p = 1$, the PANIC residuals $\widehat{e}_{it}^{\tau}$ satisfy, (i)*

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}(\widehat{e}_{it}^{\tau})^2 = \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}(e_{it}^{\tau})^2 + O_p(C_{NT}^{-2})$$

(ii) If $N/T^2 \to 0$,

$$\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{e}_{it-1}^{\tau}\Delta\widehat{e}_{it}^{\tau} = \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it-1}^{\tau}\Delta e_{it}^{\tau} + o_p(1)$$

Using the properties for V_t^{τ} and $\widehat{F}_t^{\tau}$ derived in Bai and Ng (2004), the proof of this lemma is almost identical to that of Lemmas 1 and 2. The details are omitted.

Proof of Theorem 1. For $p = -1, 0$, Lemmas 1 and 2 show that pooling $\widehat{e}_{it}$ is asymptotically the same as pooling the true errors e_{it}. The rest of the proof is similar to that of Levin et al. (2002). For completeness, we provide the key steps of the argument. From $\frac{1}{T}\sum_{t=2}^{T} e_{it-1}\Delta e_{it} \xrightarrow{d} \omega_i^2 Z_i + \lambda_i$, where $Z_i = \int_0^1 W_i(r)dW_i(r)$ and λ_i is one-sided long run variance of $\varepsilon_{it} = \Delta e_{it}$ (*i.e.*, $\lambda_i = [\omega_i^2 - \sigma_i^2]/2$). Note that $EZ_i = 0$ and $\text{var}(Z_i) = 1/2$. By the sequential limit theory in which $T \to \infty$ with N fixed, $\frac{1}{NT}\text{tr}(e_{-1}'\Delta e) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T} e_{it-1}\Delta e_{it} \xrightarrow{p} \bar{\lambda}_N$, where $\bar{\lambda}_N = \frac{1}{N}\sum_{i=1}^{N}\lambda_i$. Furthermore, $\frac{1}{T^2}\sum_{t=2}^{T} e_{it-1}^2 \xrightarrow{d} \omega_i^2 U_i$, where $U_i = \int_0^1 W_i(r)^2 dr$ with $EU_i = 1/2$. Thus $\frac{1}{NT^2}\text{tr}(e_{-1}'e_{-1}) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=2}^{T} e_{it-1}^2 \xrightarrow{p} \frac{1}{2}\bar{\omega}^2$, where $\bar{\omega}_N^2 = \frac{1}{N}\sum_{i=1}^{N}\omega_i^2$. Let ρ^+ (resp. $\widehat{\rho}^+$) be the bias corrected estimator using the true $\bar{\lambda}_N$ (resp. the estimated $\bar{\lambda}_N$), it follows that

$$\sqrt{N}T(\rho^+ - 1) = \sqrt{N}T\frac{\text{tr}\Big(e_{-1}'\Delta e - NT\bar{\lambda}_N\Big)}{\text{tr}(e_{-1}'e_{-1})} = \sqrt{N}\frac{\text{tr}(\frac{1}{NT}e_{-1}'\Delta e) - \bar{\lambda}_N}{\frac{1}{NT^2}\text{tr}(e_{-1}'e_{-1})}$$

which converges, as $T \to \infty$, to $\frac{N^{-1/2}\sum_{i=1}^{N}\omega_i^2 Z_i}{\bar{\omega}_N^2/2}$, which in turn converges to $N(0, \frac{2\phi^4}{\omega^4})$ as $N \to \infty$, where $\phi^4 = \text{plim}\,\frac{1}{N}\sum_{i=1}^{N}\omega_i^4$ and $\omega^2 = \lim_N \bar{\omega}_N^2$. We use the fact that Z_i are iid with zero mean and variance 1/2. The above limit is not changed when $\bar{\lambda}_N$ is replaced by a consistent estimator, see Moon and Perron (2004). This implies that $t_a = \sqrt{N}T(\widehat{\rho}^+ - 1)/\sqrt{2\widehat{\phi}^4/\widehat{\omega}^4} \xrightarrow{d} N(0,1)$. For the t_b test, multiply the first equation on each side by $[\frac{1}{NT^2}\text{tr}(e_{-1}'e_{-1})]^{1/2}$, whose limit is $(\bar{\omega}_N^2/2)^{1/2}$, as $T \to \infty$. We obtain

$$\sqrt{N}T(\rho^+ - 1)\Big(\frac{1}{NT^2}\text{tr}(e_{-1}'e_{-1})\Big)^{1/2} = \sqrt{N}[\text{tr}(\frac{1}{NT}e_{-1}'\Delta e) - \bar{\lambda}_N]\Big(\frac{1}{NT^2}\text{tr}(e_{-1}'e_{-1})\Big)^{-1/2}$$

which converges, as $T \to \infty$, to $\frac{N^{-1/2}\sum_{i=1}^N \omega_i^2 Z_i}{\bar{\omega}_N/\sqrt{2}}$, which in turn converges to $N(0, \phi^4/\omega^2)$, as $N \to \infty$. It follows that $t_b = \sqrt{N}T(\widehat{\rho}^+ - 1)\Big(\frac{1}{NT^2}\text{tr}(e_{-1}'e_{-1})\Big)^{1/2}\sqrt{\widehat{\omega}^2/\widehat{\phi}^4} \xrightarrow{d} N(0,1)$.

For $p = 1$, Bai and Ng (2004, page 1154) show that $\widehat{e}_{it} = e_{it} - e_{i1} - \frac{e_{iT}-e_{i1}}{T-1}(t-1) + o_p(1)$. There exists a linear trend in $\widehat{e}_{it}$. Thus projecting $\widehat{e}_{it}$ on $[1, t]$ will purge the term $e_{i1} + \frac{e_{iT}-e_{i1}}{T-1}(t-1)$, and the resulting projection has the consequence as if the true error e_{it} is projected on $[1, t]$. Lemma 4 shows that pooling $\widehat{e}_{it}^\tau$ is asymptotically the same as pooling e_{it}^τ. Let e^τ denote the matrix of demeaned and detrended series, then $\frac{1}{T}\sum_{t=2}^T e_{it-1}^\tau \Delta e_{it}^\tau \xrightarrow{d} \omega_i^2 Z_i^\tau + \lambda_i$, where $Z_i^\tau = \int_0^1 W^\tau(r)dW(r)$. From $E(Z_i^\tau) = -1/2$, we have $E(\omega_i^2 Z_i^\tau) + \lambda_i = -\sigma_i^2/2$, one half of the short-run variance of ε_{it}. Furthermore, $\frac{1}{T^2}\sum_{t=2}^T (e_{it-1}^\tau)^2 \xrightarrow{d} \omega_i^2 U_i^\tau$, where $U_i^\tau = \int_0^1 W^\tau(r)^2 dr$. Note that $E(U_i^\tau) = 1/15$, see Table 1. Let $\bar{\sigma}_n^2 = \frac{1}{N}\sum_{i=1}^N \sigma_i^2$. Thus,

$$\sqrt{N}T(\rho^+ - 1) = \sqrt{N}T\frac{\text{tr}\Big(e_{-1}^{\tau\prime}\Delta e^\tau + NT\bar{\sigma}_N^2/2\Big)}{\text{tr}(e_{-1}^{\tau\prime}e_{-1}^\tau)} \xrightarrow{d} \frac{N^{-1/2}\sum_{i=1}^N \omega_i^2(Z_i^\tau + 1/2)}{\bar{\omega}_N^2/15}$$

as $T \to \infty$. The above term then converges to $N(0, \frac{15\phi^4}{4\omega^4})$ as $N \to \infty$ because $Z_i^\tau + 1/2$ are iid zero mean and variance $1/60$. It follows that $t_a \xrightarrow{d} N(0,1)$. Next for t_b,

$$\sqrt{N}T(\rho^+ - 1)[\frac{1}{NT^2}\text{tr}(e_{-1}^{\tau\prime}e_{-1}^\tau)]^{1/2} \quad \xrightarrow{d} \quad \frac{N^{-1/2}\sum_{i=1}^N \omega_i^2(Z_i^\tau + 1/2)}{(\bar{\omega}_N^2/15)^{1/2}}$$

as $T \to \infty$. The above converges to $N(0, \frac{\phi^4}{4\omega^2})$, as $N \to \infty$. This implies $t_b \xrightarrow{d} N(0,1)$, proving Theorem 1. For Theorem 2, we need

Lemma 5 *(i) For $p = -1, 0$, assume $N/T^2 \to 0$, then we have*

$$\frac{1}{\sqrt{N}T^2}\sum_{i=1}^N\sum_{t=1}^T(\widehat{e}_{it})^2 = \frac{1}{\sqrt{N}T^2}\sum_{i=1}^N\sum_{t=1}^T e_{it}^2 + o_p(1).$$

(ii) For $p = 1$, assume $N/T^2 \to 0$, then we have

$$\frac{1}{\sqrt{N}T^2}\sum_{i=1}^N\sum_{t=1}^T(\widehat{e}_{it})^2 = \frac{1}{\sqrt{N}T^2}\sum_{i=1}^N\sum_{t=1}^T(\widetilde{e}_{it})^2 + o_p(1)$$

where $\widetilde{e}_{it} = e_{it} - (e_{iT} - e_{i1})\frac{t-1}{T-1}$.

Proof of (i). This follows from Lemma 1 by multiplying $N^{1/2}$ on each side of the equation, and noting $\sqrt{N}O_p(C_{NT}^{-2}) = o_p(1)$ if $N/T^2 \to 0$.

Proof of (ii). Lemma 1 continues to hold for $p = 1$ as long as e_{it} is replaced by $\widetilde{e}_{it}$ defined above. See Remark 3. Again, multiplying $\sqrt{N}$ as in (i), we obtain (ii).

Proof of Theorem 2. For fixed i and N, $\frac{1}{T^2}\sum_{t=2}^T e_{it}^2 \xrightarrow{d} \omega_i^2 U_i$, where $U_i = \int_0^1 W(r)^2 dr$ with $E\int_0^1 W(r)^2 dr = 1/2$ and $\mathrm{var}(U_i) = 1/3$. Thus, the sequential limit theorem implies

$$\sqrt{N}[\mathrm{tr}(\frac{1}{NT^2}e'e) - \bar{\omega}_N^2/2] \xrightarrow{d} \frac{1}{\sqrt{N}}\sum_{i=1}^N \omega_i^2(U_i - \frac{1}{2}), \quad \text{as } T \to \infty$$

for fixed N. By Lemma 5 (i), replaced e by $\widehat{e}$ for $p = -1$ or $p = 0$,

$$\sqrt{N}[\mathrm{tr}(\frac{1}{NT^2}\widehat{e}'\widehat{e}) - \bar{\omega}_N^2/2] \xrightarrow{d} \frac{1}{\sqrt{N}}\sum_{i=1}^N \omega_i^2(U_i - \frac{1}{2}), \quad \text{as } T \to \infty$$

Since U_i are independent over i, taking the limit with N, the above converges to $N(0, \phi^4/3)$. In addition, $\sqrt{N}(\widehat{\omega}^2 - \bar{\omega}_N^2) = o_p(1)$, if $N/T \to 0$, see Moon and Perron (2004). The limit is not affected when $\bar{\omega}_N^2$ is replaced by its estimate. This implies the theorem for $p = -1, 0$.

Consider $p = 1$. From $\frac{1}{T^2}\sum_{t=2}^T \widetilde{e}_{it}^2 \xrightarrow{d} \omega_i^2 V_i$, as $T \to \infty$, where $\widetilde{e}_{it} = e_{it} - (e_{iT} - e_{i1})\frac{t-1}{T-1}$, V_i is a Brownian bridge with $EV_i = 1/6$ and $\mathrm{var}(V_i) = 1/45$. Thus as $T \to \infty$ with fixed N,

$$\sqrt{N}[\mathrm{tr}(\frac{1}{NT^2}\widetilde{e}'\widetilde{e}) - \bar{\omega}_N^2/6] \xrightarrow{d} \frac{1}{\sqrt{N}}\sum_{i=1}^N \omega_i^2(V_i - \frac{1}{6}).$$

Lemma 5(ii) shows that the unobserved $\widetilde{e}_{it}$ can be replaced by $\widehat{e}_{it}$. Since V_i are independent, letting N go to infinity, the right-hand side above converges to $N(0, \phi^4/45)$. Replacing $\bar{\omega}_N^2$ by its estimate and rescaling, we prove the theorem.

References

Bai, J. 2003, Inferential Theory for Factor Models of Large Dimensions, *Econometrica* **71:1**, 135–172.

Bai, J. and Ng, S. 2002, Determining the Number of Factors in Approximate Factor Models, *Econometrica* **70:1**, 191–221.

Bai, J. and Ng, S. 2004, A PANIC Attack on Unit Roots and Cointegration, *Econometrica* **72:4**, 1127–1117.

Breitung, J. and Das, S. 2008, Testing for Unit Root in Panels with a Factor Structure, *Econometric Theory.*

Chang, Y. 2002, Nonlinear IV Unit Root Tests in Panels with Cross-Section Dependencecy, *Journal of Econometrics* **110:2**, 261–292.

Chang, Y. and Song, W. 2002, Panel Unit Root Tests in the Presence of Cross-Section Heterogeneity, unpublished manuscript, Rice University.

Choi, I. 2006, Combination Unit Root Tests for Cross-Sectionally Correlated Panels, *Econometric Theory and Practice: Frontiers Analysis and Applied Research: Essays in Horor of Peter C.B. Phillips*, Cambridge University Press, pp. 311–333.

Gengenbach, C., Palm, F. and Urbain, J. P. 2006, Panel Unit Root Tests in the Presence of Cross-Sectional Dependencies: Comparison and Implications for Modelling, University of Masstricht.

Jang, M. J. and Shin, D. 2005, Comparison of Panel Unit Root Tests Under Cross-Sectional Dependence, *Economics Letters* **89**, 12–17.

Levin, A., Lin, C. F. and Chu, J. 2002, Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties, *Journal of Econometrics* **98:1**, 1–24.

Maddala, G. S. and Wu, S. 1999, A Comparative Study of Unit Root Tests with Panel data and a New Simple Test, *Oxford Bulletin of Economics and Statistics* **61**, 631–652.

Moon, R. and Perron, B. 2004, Testing For a Unit Root in Panels with Dynamic Factors, *Journal of Econometrics* **122:1**, 81–126.

Moon, R., Perron, B. and Phillips, P. 2005, Incidental Trends and the Power of Panel Unit Root Tests, *Journal of Economics.*

O'Connell, P. 1998, The Overvaluation of Purchasing Power Parity, *Journal of International Economics* **44**, 1–19.

Pesaran, H. 2007, A Simple Unit Root Test in the Presence of Cross-Section Dependence, *Journal of Applied Economics* **22:2**, 265–312.

Phillips, P. and Moon, R. 1999, Linear Regression Limit Theory for Nonstationary Panel Data, *Econometrica* **67**, 1057–1111.

Phillips, P. and Sul, D. 2003, Dynamic Panel Estimation and Homogeneity Testing Under Cross-Section Dependence, *Econometrics Journal* **6:1**, 217–259.

Phillips, P. C. B. and Ploberger, W. 2002, Optimal Testing for Unit Roots in Panel Data. mimeo, University of Rochester.

Sargan, J. D. and Bhargava, A. 1983, Testing for Residuals from Least Squares Regression Being Generated by Gaussian Random Walk, *Econometrica* **51**, 153–174.

Stock, J. H. 1990, A Class of Tests for Integration and Cointegration, mimeo.

Westerlund, J. and Larsson, R. 2007, A Note on the Pooling of Individual PANIC Unit Root Tests, unpublished manuscript.