Econometric Theory, **26**, 2010, 1088–1114.
doi:10.1017/S0266466609990478

PANEL UNIT ROOT TESTS WITH CROSS-SECTION DEPENDENCE: A FURTHER INVESTIGATION

JUSHAN BAI AND SERENA NG
Columbia University

An effective way to control for cross-section correlation when conducting a panel unit root test is to remove the common factors from the data. However, there remain many ways to use the defactored residuals to construct a test. In this paper, we use the panel analysis of nonstationarity in idiosyncratic and common components (PANIC) residuals to form two new tests. One estimates the pooled autoregressive coefficient, and one simply uses a sample moment. We establish their large-sample properties using a joint limit theory. We find that when the pooled autoregressive root is estimated using data detrended by least squares, the tests have no power. This result holds regardless of how the data are defactored. All PANIC-based pooled tests have nontrivial power because of the way the linear trend is removed.

1. INTRODUCTION

Cross-section dependence can pose serious problems for testing the null hypothesis that all units in a panel are nonstationary. 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 is in fact the consequence of nontrivial 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 panel analysis of nonstationarity in idiosyncratic and common components (PANIC) framework of Bai and Ng (2004) enables the common factors and the idiosyncratic errors to be tested separately, and Moon and Perron (2004) test 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 Moon and Perron (2004) tests (henceforth MP tests) retain the spirit of the original panel unit root test of Levin, Lin, and Chu (2002), which estimates and tests the pooled

We thank three referees, a co-editor, and Benoit Perron for many helpful comments and suggestions. We also acknowledge financial support from the NSF (grants SES-0551275 and SES-0549978). Address correspondence to Jushan Bai, Department of Economics, Columbia University; 1022 IAB, 420 West 118th Street, New York, NY 10027, USA; e-mail: jushan.bai@columbia.edu.

  0266-4666/10 $15.00

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.

Many papers have studied the finite-sample properties of various panel unit root tests. In this paper we try to understand whether the difference in finite-sample properties can be traced to how the pooled autoregressive coefficient is estimated. To this end, we first develop a set of MP type tests using the PANIC residuals, and a panel version of the modified Sargan–Bhargava test (hereafter the PMSB test) that simply uses the sample moments of these residuals but does not estimate the pooled autoregressive coefficient. We then use simulations to show that autoregressive coefficient–based tests have minimal power whenever $\widehat{\rho}$ is constructed using data that are detrended by least squares, irrespective of how the factors are removed. We develop new PANIC-based pooled tests that do not require explicit linear detrending. The three PANIC tests have reasonable power against the trend stationary alternative because they do not involve least squares detrending.

The rest of the paper is organized as follows. In Section 2, we specify the data generating process (DGP), introduce necessary notation, and discuss model assumptions. In Section 3, we consider the PANIC residual-based MP type and PMSB tests. Section 4 discusses issues related to different tests. Section 5 provides finite-sample evidence via Monte Carlo simulations. Concluding remarks are given in Section 6, and the proofs are given in the Appendix.

2. PRELIMINARIES

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, we let $M_z = I - z(z'z)^{-1}z'$ be a matrix that projects on the orthogonal space of z. In particular, the projection matrix M_0 with $z_t = 1$ for all t demeans the data, and M_1 with $z_t = (1, t)'$ demeans and detrends the data. Trivially, M_{-1} is simply an identity matrix.

The DGP is

$$X_{it} = D_{it} + \lambda_i' F_t + e_{it}, \tag{1}$$

$$(1 - L) F_t = C(L) \eta_t,$$

$$e_{it} = \rho_i e_{it-1} + \varepsilon_{it},$$

where F_t is an $r \times 1$ vector of common factors that induce correlation across units, λ_i is an $r \times 1$ vector of factor loadings, e_{it} is an idiosyncratic error, and $C(L)$ is an $r \times r$ matrix consisting of polynomials of the lag operator L, $C(L) = \sum_{j=0}^{\infty} C_j L^j$.

Let $M < \infty$ denote a positive constant that does not depend on T or N. Also let $\|A\| = \text{tr}(A'A)^{1/2}$. We use the following assumptions based on Bai and Ng (2004).

Assumption A.

(a) If λ_i is nonrandom, $\|\lambda_i\| \leq M$; if λ_i is random, $\text{E}\|\lambda_i\|^4 \leq M$;
(b) $N^{-1}\sum_{i=1}^{N}\lambda_i\lambda_i' \xrightarrow{p} \Sigma_\Lambda$, an $r \times r$ positive definite matrix.

Assumption B.

(a) $\eta_t \sim iid(0, \Sigma_\eta)$, $\text{E}\|\eta_t\|^4 \leq M$;
(b) $\text{var}(\Delta F_t) = \sum_{j=0}^{\infty} C_j \Sigma_\eta C_j' > 0$;
(c) $\sum_{j=0}^{\infty} j \left\|C_j\right\| < M$; and
(d) $C(1)$ has rank $r_1, 0 \leq r_1 \leq r$.

Assumption C. for each i, $\varepsilon_{it} = d_i(L)v_{it}$, $v_{it} \sim iid(0,1)$ across i and over t, $\text{E}|v_{it}|^8 \leq M$ for all i and t; $\sum_{j=0}^{\infty} j|d_{ij}| \leq M$ for all i; $d_i(1)^2 \geq c > 0$ for all i and for some $c > 0$.

Assumption D. $\{v_{is}\}$, $\{\eta_t\}$, and $\{\lambda_j\}$ are mutually independent.

Assumption E. $\text{E}\|F_0\| \leq M$, and for every $i = 1, \ldots N$, $\text{E}|e_{i0}| \leq M$.

Assumptions A and B assume that there are r factors. Assumption B allows a combination of stationary and nonstationary factors. Assumption C assumes cross-sectionally independent idiosyncratic errors, which is used to invoke some of the results of Phillips and Moon (1999) for joint limit theory, and for cross-sectional pooling. This assumption is similar to Assumption 2 of Moon and Perron (2004). We point out that many properties of the PANIC residuals derived in the Appendix are not affected by allowing some weak cross-sectional correlations among v_{it}. The variance of v_{it} in the linear process ε_{it} is normalized to be 1; otherwise it can be absorbed into $d_i(L)$. Assumption D assumes that factors, factor loadings, and the idiosyncratic errors are mutually independent. Initial conditions are stated in Assumption E.

For the purpose of this paper we let

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

where Φ_1 is an $r \times r$ matrix. The number of nonstationary factors is determined by the number of unit roots in the polynomial matrix equation, $\Phi(L) = I - \Phi_1 L = 0$. Under (1), X_{it} can be nonstationary 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 nonstationary. An important feature of the DGP given by (1) is that the common and the idiosyncratic components can have different orders of integration. It is only when we reject nonstationarity in both components that we can say that the data are inconsistent with unit root nonstationarity.

Other DGPs have also been considered in the literature on panel unit root tests. The one used in Phillips and Sul (2003) is a special case of (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., $\lambda_i = 1$). A somewhat different DGP is used in Moon and Perron (2004) and Moon, Perron, and Phillips (2007). They let

$$X_{it} = D_{it} + X_{it}^0, \tag{2}$$
$$X_{it}^0 = \rho_i X_{it-1}^0 + u_{it},$$
$$u_{it} = \lambda_i' f_t + \varepsilon_{it},$$

where f_t and ε_{it} are I(0) linear processes, f_t and ε_{it} are independent, and ε_{it} are cross-sectionally independent. Notably, under (2), X_{it} has a unit root if $\rho_i = 1$. This DGP differs from (1) in that it essentially specifies the dynamics of the observed series (e.g., if $D_{it} = 0$, then $X_{it} = X_{it}^0$), whereas (1) specifies the dynamics of unobserved components. Assuming $X_{i0}^0 = 0$ and $\rho_i = \rho$ for all i, (2) can be written in terms of (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 $e_{it} = \rho e_{it-1} + \varepsilon_{it}$, and so both F_t and e_{it} are I(0). Thus the common and idiosyncratic components in (2) are restricted to have the same order of integration. Note that when ρ_i are heterogeneous, (2) cannot be expressed in terms of (1). But under the null hypothesis that $\rho_i = 1$ for all i, (1) covers (2). It follows that the assumptions used for DGP (1) are also applicable to DGP (2). The model considered by Pesaran (2007) is identical to DGP (2) as the dynamics are expressed in terms of the observable variable X_{it}:

$$X_{it} = (1 - \rho_i L)D_{it} + \rho_i X_{it-1} + u_{it}, \tag{3}$$
$$u_{it} = \lambda_i f_t + \varepsilon_{it}.$$

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

$$\sigma_{\varepsilon i}^2 = \mathrm{E}(\varepsilon_{it}^2) = \sum_{j=0}^{\infty} d_{ij}^2, \qquad \omega_{\varepsilon i}^2 = \left(\sum_{j=0}^{\infty} d_{ij}\right)^2, \qquad \lambda_{\varepsilon i} = (\omega_{\varepsilon i}^2 - \sigma_{\varepsilon i}^2)/2,$$

respectively. Throughout, $\omega_{\varepsilon i}^4 = (\omega_{\varepsilon i}^2)^2$ and $\omega_{\varepsilon i}^6 = (\omega_{\varepsilon i}^2)^3$, etc. As in Moon and Perron (2004) we assume that the following limits exist and the first three are strictly positive:

$$\omega_\varepsilon^2 = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \omega_{\varepsilon i}^2, \qquad \sigma_\varepsilon^2 = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \sigma_{\varepsilon i}^2,$$

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

The subscript ε may be dropped when context is clear. For future reference, let

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

be consistent estimates of $\omega_\varepsilon^2, \sigma_\varepsilon^2, \phi_\varepsilon^4$, and λ_ε, respectively. Assumptions necessary for consistent estimation of these long-run and one-sided long-run variances are given in Moon and Perron (2004). These will not be restated here so that we can focus on the main issues we want to highlight here.

3. PANIC POOLED TESTS

In Bai and Ng (2004) we showed that under (1) testing can still proceed even when both components are unobserved and without knowing a priori whether e_{it} is nonstationary. 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 $\widehat{e}_{it}$). 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 recumulating the estimated factor components. More precisely, when D_{it} in (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 $(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_N)$ and $(\widehat{f}_2, \ldots, \widehat{f}_T)$ and $\widehat{z}_{it}$ for all i and t. Define

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

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 sample mean of ΔX_{it} over t and where $\overline{\Delta F}$ and $\overline{\Delta e_i}$ are similarly defined.

Bai and Ng (2004) provide asymptotically valid procedures for (a) determining the number of stochastic trends in $\widehat{F}_t$, (b) testing if $\widehat{e}_{it}$ are individually I(1) using augmented Dickey–Fuller (ADF) regressions, and (c) testing if the panel is I(1) by pooling the p values of the individual tests. If π_i is the p-value of the ADF test for the ith cross-section unit, the pooled test is

$$P_{\widehat{e}} = \frac{-2\sum_{i=1}^{N} \log \pi_i - 2N}{\sqrt{4N}}. \tag{5}$$

The test is asymptotically standard normal. For a two-tailed 5% test, the null hypothesis is rejected when $P_{\widehat{e}}$ exceeds 1.96 in absolute value. Note that $P_{\widehat{e}}$ does not require a pooled ordinary least squares (OLS) estimate of the AR(1) coefficient in the idiosyncratic errors. Pooling p values has the advantage that more heterogeneity in the units is 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, i.e., $\widehat{e}_{it}$. Specifically, for DGP with $p = -1, 0$, or 1, 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.

The bias-corrected pooled PANIC autoregressive estimator ρ and the test statistics depend on the specification of the deterministic component D_{it}. For $p = -1$ and 0,

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

and the test statistics are

$$P_a = \frac{\sqrt{N}T(\widehat{\rho}^{+} - 1)}{\sqrt{2\widehat{\phi}_{\varepsilon}^{4}/\widehat{\omega}_{\varepsilon}^{4}}}, \tag{6}$$

$$P_b = \sqrt{N}T(\widehat{\rho}^{+} - 1)\sqrt{\frac{1}{NT^2}\mathrm{tr}(\widehat{e}'_{-1}\widehat{e}_{-1})\frac{\widehat{\omega}_{\varepsilon}^{2}}{\widehat{\phi}_{\varepsilon}^{4}}}. \tag{7}$$

For $p = 1$,

$$\widehat{\rho}^{+} = \frac{\mathrm{tr}(\widehat{e}'_{-1}\widehat{e})}{\mathrm{tr}(\widehat{e}'_{-1}\widehat{e}_{-1})} + \frac{3}{T}\frac{\widehat{\sigma}_{\varepsilon}^{2}}{\widehat{\omega}_{\varepsilon}^{2}} = \widehat{\rho} + \frac{3}{T}\frac{\widehat{\sigma}_{\varepsilon}^{2}}{\widehat{\omega}_{\varepsilon}^{2}},$$

and the test statistics are

$$P_a = \frac{\sqrt{N}T(\widehat{\rho}^+ - 1)}{\sqrt{(36/5)\widehat{\phi}_\varepsilon^4 \widehat{\sigma}_\varepsilon^4 / \widehat{\omega}_\varepsilon^8}}, \tag{8}$$

$$P_b = \sqrt{N}T(\widehat{\rho}^+ - 1)\sqrt{\frac{1}{NT^2}\operatorname{tr}(\widehat{e}_{-1}'\widehat{e}_{-1})\frac{5}{6}\frac{\widehat{\omega}_\varepsilon^6}{\widehat{\phi}_\varepsilon^4 \widehat{\sigma}_\varepsilon^4}}, \tag{9}$$

where $\widehat{\lambda}_\varepsilon$, $\widehat{\sigma}_\varepsilon^2$, $\widehat{\omega}_\varepsilon^2$, and $\widehat{\phi}_\varepsilon^4$ are defined in (4). These nuisance parameters are estimated based on AR(1) residuals $\widehat{\varepsilon} = \widehat{e} - \widehat{\rho}\,\widehat{e}_{-1} = [\widehat{\varepsilon}_1, \widehat{\varepsilon}_2, \ldots, \widehat{\varepsilon}_N]$ with $\widehat{\varepsilon}_i$ being $(T-2) \times 1$ for all i.[1]

THEOREM 1. *Let $\widehat{\rho}^+$ be the bias-corrected pooled autoregressive coefficient for the idiosyncratic errors estimated by PANIC. Suppose the data are generated by (1) and Assumptions A–E hold. Then under the null hypothesis that $\rho_i = 1$ for all i, as $N, T \to \infty$ with $N/T \to 0$, $P_a \xrightarrow{d} N(0,1)$ and $P_b \xrightarrow{d} N(0,1)$.*

Jang and Shin (2005) studied the properties of $P_{a,b}$ for $p = 0$ by simulations. Theorem 1 provides the limiting theory for both $p = 0$ and $p = 1$. It shows that the t tests of the pooled autoregressive coefficient in the idiosyncratic errors are asymptotically normal. The convergence holds for N and T tending to infinity jointly with $N/T \to 0$. It is thus a joint limit in the sense of Phillips and Moon (1999). The P_a and P_b are the analogs of t_a and t_b of Moon and Perron (2004), except that (a) the tests are based on PANIC residuals and (b) the method of "defactoring" of the data is different from the method of Moon and Perron (2004). 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 ADF 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 also different from $t_{a,b}$ of Moon and Perron (2004). 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. The usual bias correction made to recenter the numerator of the estimator to zero is no longer appropriate. This is because the deviation of the numerator from its mean, multiplied by $\sqrt{N}$, is still degenerate. However, we can do bias correction to the estimator directly because $T(\widehat{\rho} - 1)$ converges to a constant. In the present case, $T(\widehat{\rho} - 1) \xrightarrow{p} -3\sigma_\varepsilon^2/\omega_\varepsilon^2$. This leads to $\widehat{\rho}^+$ as defined previously for $p = 1$. This definition of $\widehat{\rho}^+$ is crucial for the tests to have power in the presence of incidental trends.

3.1. The Pooled MSB

An important feature that distinguishes stationary from nonstationary processes is that their sample moments require different rates of normalization 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 and 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 e_{it} is stationary, $Z_i = O_p(T^{-1})$. Stock (1990) developed the modified Sargan–Bhargava test (MSB test) to allow $\varepsilon_{it} = \Delta e_{it}$ to be serially correlated with short- and long-run variance $\sigma_{\varepsilon i}^2$ and $\omega_{\varepsilon i}^2$, respectively. In particular, if $\widehat{\omega}_{\varepsilon i}^2$ is an estimate of $\omega_{\varepsilon i}^2$ that is consistent under the null and is bounded under the alternative,[2] $MSB = Z_i/\widehat{\omega}_{\varepsilon i}^2 \Rightarrow \int_0^1 W_i^2(r)\,dr$ under the null and degenerates to zero under the alternative. Thus the null is rejected when the statistic is too small. As shown in Perron and Ng (1996) and Ng and Perron (2001), the MSB has power similar to the ADF test of Said and Dickey (1984) and the Phillips–Perron test developed in Phillips and Perron (1988) for the same method of detrending. An unique feature of the MSB is that it does not require estimation of ρ, which allows us to subsequently assess whether power differences across tests are due to the estimate of ρ. This motivates the following simple panel nonstationarity test for the idiosyncratic errors, denoted the panel PMSB test. Let $\widehat{e}$ be obtained from PANIC. For $p = -1, 0$, the test statistic is defined as

$$PMSB = \frac{\sqrt{N}\left(\mathrm{tr}\left(\frac{1}{NT^2}\widehat{e}'\widehat{e}\right) - \widehat{\omega}_\varepsilon^2/2\right)}{\sqrt{\widehat{\phi}_\varepsilon^4/3}}, \tag{10}$$

where $\widehat{\omega}_\varepsilon^2/2$ estimates the asymptotic mean of $(1/NT^2)\mathrm{tr}(\widehat{e}'\widehat{e})$ and the denominator estimates its standard deviation. For $p = 1$, the test statistic is defined as

$$PMSB = \frac{\sqrt{N}\left(\mathrm{tr}\left(\frac{1}{NT^2}\widehat{e}'\widehat{e}\right) - \widehat{\omega}_\varepsilon^2/6\right)}{\sqrt{\widehat{\phi}_\varepsilon^4/45}}. \tag{11}$$

The variables $\widehat{\omega}_\varepsilon^2$ and $\widehat{\phi}_\varepsilon^4$ are defined in (4), and are estimated from residuals $\widehat{\varepsilon} = \widehat{e} - \widehat{\rho}e_{-1}$, where $\widehat{\rho}$ is the pooled least squares estimator based on $\widehat{e}$. The null hypothesis that $\rho_i = 1$ for all i is rejected for small values of PMSB. We have the following result:

THEOREM 2. *Let PMSB be defined as in (11). Under Assumptions A–E, as* $N, T \to \infty$ *with* $N/T \to 0$, *we have*

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

The convergence result again holds in the sense of a joint limit. But a sequential asymptotic argument 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 $\omega_{\varepsilon i}^2 \int_0^1 V_i(r)^2 dr$ 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 test with P_a and P_b tests when $p = 1$ is of special interest. From Bai and Ng (2004), $\widehat{e}_{it} = e_{it} - e_{i1} - \frac{(e_{iT} - e_{i1})}{T-1}(t-1) + o_p(1)$, which has a time trend component with slope coefficient of $O_p(T^{-1/2})$. Because of the special slope coefficients, detrending is unnecessary when constructing P_a and P_b tests, but suitable bias correction for the autoregressive coefficient is necessary to avoid certain degeneracy (see the discussion of degeneracy following Theorem 1). Detrending is also unnecessary with the PMSB test because the limit of $T^{-\frac{1}{2}}\widehat{e}_{it}$ is simply a Brownian bridge. Not having to detrend $\widehat{e}_{it}$ is key to having tests with good finite-sample properties when $p = 1$.

4. THE MP TESTS

The autoregressive coefficient ρ can also be estimated from data in levels

$$X_{it} = (1 - \rho L) D_{it} + \rho X_{it-1} + u_{it}, \qquad u_{it} = \lambda_i' f_t + \varepsilon_{it}.$$

As is standard in the literature on unit roots, there are three models to consider: a base case model (A) that assumes D_{it} is null; a fixed effect model (B) that assumes $D_{it} = a_i$; and an incidental trend model (C) that has $D_{it} = a_i + b_i t$. Note that we use $p = -1, 0, 1$ to represent the DGP and use Models A–C to represent how the trends are estimated. Let $\Lambda = (\lambda_1, \ldots \lambda_N)'$ and X and X_{-1} be $T-1$ by N matrices. Based on the first step estimator $\widehat{\rho} = \frac{\mathrm{tr}(X_{-1}' M_z X)}{\mathrm{tr}(X_{-1}' M_z X_{-1})}$, one computes the residuals $\widehat{u} = M_z X - \widehat{\rho} M_z X_{-1}$, from which a factor model is estimated to obtain $\widehat{\Lambda} = (\widehat{\lambda}_1, \ldots, \widehat{\lambda}_N)'$, where M_z is a projection matrix defined in Section 2. The bias-corrected, defactored, pooled OLS estimator defined in Moon and Perron (2004) is

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

where $M_{\widehat{\Lambda}} = I_N - \widehat{\Lambda}(\widehat{\Lambda}'\widehat{\Lambda})^{-1}\widehat{\Lambda}'$ and $\widehat{\psi}_{\varepsilon}$ is a bias correction term (given subsequently) 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, have the same form as P_a and P_b defined in (6) and (7), with X and X_{-1} replacing $\widehat{e}$ and $\widehat{e}_{-1}$ both in $\widehat{\rho}^{+}$ and in the tests. That is,

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

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

where M_z and the parameters $\widehat{\psi}_{\varepsilon}$, K_a, and K_b are defined as follows. When the data are untransformed (Model A) $M_z = I_{T-2}$, $\widehat{\psi}_{\varepsilon} = \widehat{\lambda}_{\varepsilon}$, $K_a = 2$, and $K_b = 1$.

When the data are demeaned (Model B), then $M_z = M_0$, $\widehat{\phi}_\varepsilon = -\widehat{\sigma}_\varepsilon^2/2$, $K_a = 3$, and $K_b = 2$. When the data are demeaned and detrended (Model C),[3] $M_z = M_1$ and $\widehat{\phi}_\varepsilon = -\widehat{\sigma}_\varepsilon^2/2$, $K_a = 15/4$, and $K_b = 4$. Model A is valid when p is -1 or 0 in the DGP.

There are two important differences between $P_{a,b}$ and $t_{a,b}$. First, our tests explicitly estimate the factors and errors before testing, whereas the MP tests implicitly remove the common factors from the data and thus do not explicitly define $\widehat{e}_{it}$. As a result of this, the bias adjustments are also different. For $p = 1$, we obtain a bias adjustment for $\widehat{\rho}$ directly, whereas Moon and Perron (2004) adjusted the bias for the numerator of $\widehat{\rho}$. Also, Moon and Perron removed the deterministic terms by the least squares estimation of the incidental parameters, whereas PANIC takes the first difference of the data. Note that in the Moon and Perron setup, Models A and B are both valid for $p = 0$ because under the null hypothesis that $\rho = 1$, the intercepts are identically zero. However, the finite-sample properties of the MP test are much better when A is used. Model B (demeaning) gives large size distortions, even though removing the fixed effects seems to be the natural way to proceed.

It should also be remarked that Moon and Perron (2004) estimated the nuisance parameters by averaging over $\widehat{\omega}_{\varepsilon i}^2$ and $\widehat{\sigma}_{\varepsilon i}^2$, where these latter are defined using $\widehat{\varepsilon} = \widehat{u} M_{\widehat{\Lambda}}$. Importantly, $\widehat{\omega}_{\varepsilon i}$ is a function of $\widehat{\rho}$ that 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. This seems to improve the size of the test when $p = 1$ but does not improve power. Simulations show that the MP tests are dominated by $P_{a,b}$ when $p = 1$.

Because the MP tests are applied directly to the observable series, one might infer that t_a and t_b are testing the observed panel of data. It is worth reiterating that after the common factors are controlled for, one must necessarily be testing the properties of the idiosyncratic errors. This is clearly true for (2) because both the common and idiosyncratic terms have the same order of integration. Although less obvious, the statement is also true for model (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. The DGP defined by (1) can be rewritten as

$$X_{it} = \rho_i X_{it-1} + \lambda_i' F_t - \rho_i \lambda_i' F_{t-1} + \varepsilon_{it}. \quad \textbf{(12)}$$

Because the defactored approach will remove the common factors, we can ignore them in the equations. It is then obvious that the $t_{a,b}$ tests (using observable data) will determine if the (weighted) average of ρ_i is unity, where the ρ_i are the autoregressive coefficients for the idiosyncratic error processes. 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 F_t. As such, the covariate augmented Dickey–Fuller (CADF), which takes an average of the t ratios on ρ_i, is also a test of whether the idiosyncratic errors are nonstationary. 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 nonstationary; see Breitung and Das (2008). To make statements concerning nonstationarity for the observed series X_{it}, researchers still have to separately test if the common factors are nonstationary. PANIC presents a framework that can establish if the components are stationary in a coherent manner.

5. FINITE-SAMPLE PROPERTIES

In this section, we report the finite-sample properties of $P_{\widehat{e}}$, PMSB, and autoregressive coefficient–based tests, $P_{a,b}$ and $t_{a,b}$. As $p = -1$ is not usually a case of practical interest, we only report results for $p = 0$ and $p = 1$. For the $t_{a,b}$ tests, we follow Moon and Perron (2004) and use Model A for testing (i.e., no demeaning) instead of B (demeaning) when $p = 0$. To make this clear, we denote the results by $t^A_{a,b}$. For $p = 1$, the $t_{a,b}$ tests are denoted by $t^C_{a,b}$ (with demeaning and detrending).

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 pooling the p values as in Bai and Ng (2004). Gengenbach, Palm, and Urbain (2009) also compared the MP tests with PANIC but also for $p = 0$ only. In addition, 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 four models. Additional results are available on request.

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

Model 1. $\Phi_1 = 1$, $\rho_i = 1$ for all i;
Model 2. $\Phi_1 = 0.5$, $\rho_i \sim U[0.9, 0.99]$;
Model 3. $\Phi_1 = 0.5$, $\rho_i = 1$ for $i = 1, \ldots, N/5$, $\rho_i \sim U[0.9, 0.99]$ otherwise;
Model 4. $\rho_i \sim U[0.9, 0.99]$,

where Φ_1 is the autoregressive 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. The number of replications is 5,000.

We hold the number of factors to the true value when we evaluate the adequacy of the asymptotic approximations because the theory does not incorporate sampling variability due to estimating the number of factors. However, in practice, the number of factors (r) is not known. Bai and Ng (2002) developed procedures that can consistently estimate r. Their simulations showed that when N and T are large, the number of factors can be estimated precisely. However, the number of factors can be overestimated when T or N is small (say, less than 20). In those cases, the authors recommend using BIC_3 (Bai and Ng, 2002, p. 202). Alternatively, classical factor analysis can be used to determine the number of factors for small N or T; see Anderson (1984, Ch. 14). Gengenbach et al. (2009) found that the performance of panel unit root tests can be distorted when the number of common factors is overestimated. It is possible that the BIC_3 can alleviate the problem. Regardless of which criterion to use, the finite-sample distributions of post–model selection estimators typically depend on unknown model parameters in a complicated fashion. As Leeb and Potscher (2008) showed, the asymptotic distribution can be a poor approximation for the finite-sample distributions for certain DGPs. This caveat should be borne in mind.

To illustrate the main point of this paper, namely, tests that explicitly detrend the data first to construct $\widehat{\rho}$ have no power, we additionally report results for two tests that we will denote P_a^D and P_b^D. These tests obtain $\widehat{\rho}^+$ from Model B for DGP with $p = 0$ and Model C when $p = 1$. The former is a regression of $\widehat{e}_{it}$ on $\widehat{e}_{it-1}$, plus an individual specific constant. The latter adds a trend. The tests use the same adjustments as t_a and t_b of Moon and Perron (2004). Although detrending renders the numerator of $\widehat{\rho}^+$ nondegenerate, it also removes whatever power the tests might have, as we now see.

Results are reported in Table 1 for $p = 0$ and Table 2 for $p = 1$. The rejection rates of Model 1 correspond to finite-sample size when the nominal size is 5%. Models 2, 3, and 4 give power. Power is not size adjusted to focus on rejection rates that one would obtain in practice. Table 1 shows that for $p = 0$ PMSB, P_a, and P_b seem to have better size properties. Apart from size discrepancies when T is small, all tests have similar properties.

The difference in performance is much larger when $p = 1$. Table 2 shows that t_a, t_b, P_a^D, and P_b^D are grossly oversized, and both tests use least-squares-detrended data to estimate ρ. The P_a^D and P_b^D have no power. On the other hand, $P_{\widehat{e}}$, PMSB, P_a, and P_b are much better behaved. Importantly, these tests either do not need a pooled estimate of ρ or they do so without linearly detrending $\widehat{e}$. Moon et al. (2007) find that the MP tests have no local power against the alternative of incidental trends. Our simulations suggest that this loss of power arises as a result of detrending the data to construct $\widehat{\rho}$.

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 test is not shown here, the PMSB does appear to have good finite-sample properties. The panel unit root null hypothesis can thus be tested without having to estimate ρ.

Incidental parameters clearly create challenging problems for unit root testing using panel data, especially for tests based on an estimate of the pooled autoregressive coefficient. The question arises as to whether alternative methods of detrending might help. In unreported simulations, the finite-sample size and power of P_a^D and P_b^D under generalized least squares detrending are still unsatisfactory. One way of resolving this problem is to avoid detrending altogether. This is the

TABLE 1. Rejection rates when $p = 0$ in DGP

N	T	$P_{\widehat{e}}$	PMSB	P_a	P_b	P_a^D	P_b^D	t_a^A	t_b^A
				Model 1 (Size): $F \sim \text{I}(1), e_{it} \sim \text{I}(1)$					
20	20	0.210	0.004	0.118	0.085	0.099	0.103	0.145	0.105
20	50	0.073	0.017	0.116	0.077	0.072	0.071	0.101	0.066
20	100	0.057	0.020	0.108	0.074	0.063	0.063	0.101	0.063
50	20	0.278	0.007	0.098	0.080	0.129	0.130	0.170	0.147
50	50	0.074	0.022	0.100	0.078	0.078	0.077	0.084	0.058
50	100	0.059	0.031	0.098	0.076	0.067	0.064	0.085	0.064
100	20	0.400	0.007	0.114	0.099	0.163	0.162	0.192	0.178
100	50	0.067	0.020	0.101	0.083	0.082	0.081	0.076	0.059
100	100	0.058	0.034	0.089	0.074	0.070	0.069	0.072	0.058
				Model 2 (Power): $F \sim \text{I}(0), e_{it} \sim \text{I}(0)$					
20	20	0.578	0.160	0.837	0.770	0.105	0.065	0.966	0.938
20	50	0.879	0.971	1.000	0.999	0.025	0.003	1.000	1.000
20	100	1.000	1.000	1.000	1.000	0.005	0.000	1.000	1.000
50	20	0.854	0.478	0.978	0.971	0.117	0.075	0.999	0.998
50	50	0.997	1.000	1.000	1.000	0.020	0.003	1.000	1.000
50	100	1.000	1.000	1.000	1.000	0.003	0.000	1.000	1.000
100	20	0.974	0.752	0.999	0.999	0.130	0.088	1.000	1.000
100	50	1.000	1.000	1.000	1.000	0.012	0.002	1.000	1.000
100	100	1.000	1.000	1.000	1.000	0.001	0.000	1.000	1.000
				Model 3 (Power): $F \sim \text{I}(0), e_{it}$ mixed I(0), I(1)					
20	20	0.479	0.070	0.601	0.516	0.097	0.073	0.880	0.830
20	50	0.738	0.692	0.931	0.895	0.054	0.021	0.969	0.949
20	100	0.995	0.918	0.978	0.962	0.033	0.010	0.981	0.968
50	20	0.751	0.189	0.832	0.795	0.113	0.081	0.980	0.970
50	50	0.970	0.964	0.996	0.995	0.035	0.013	0.999	0.998
50	100	1.000	0.999	1.000	1.000	0.027	0.006	1.000	1.000
100	20	0.929	0.345	0.963	0.956	0.128	0.094	0.998	0.998
100	50	1.000	1.000	1.000	1.000	0.029	0.009	1.000	1.000
100	100	1.000	1.000	1.000	1.000	0.017	0.003	1.000	1.000
				Model 4 (Power): $F \sim \text{I}(0), e_{it} \sim \text{I}(0)$					
20	20	0.506	0.122	0.752	0.679	0.108	0.074	0.818	0.734
20	50	0.784	0.897	0.984	0.976	0.031	0.009	0.987	0.981
20	100	0.996	0.993	0.998	0.998	0.014	0.003	0.999	0.998
50	20	0.776	0.376	0.918	0.895	0.122	0.088	0.933	0.912
50	50	0.959	0.961	0.985	0.981	0.037	0.016	0.988	0.982
50	100	1.000	0.997	0.998	0.997	0.030	0.012	0.998	0.997
100	20	0.925	0.592	0.966	0.962	0.142	0.106	0.968	0.960
100	50	0.994	0.992	0.996	0.995	0.037	0.021	0.997	0.995
100	100	1.000	0.999	1.000	0.999	0.033	0.015	1.000	1.000

Note: $P_{\widehat{e}}$, PMSB, $P_a, P_b, P_a^D, and\, P_b^D$ are tests based on PANIC residuals $\widehat{e}$. The first four are defined in (5), (11), (8), and (9). The tests P_a^D and P_b^D are constructed in the same way as t_a and t_b, but they estimate ρ^+ from an autoregression using $\widehat{e}$ with a constant; t_a^A and t_b^A are MP tests that do not demean the data.

TABLE 2. Rejection rates when $p = 1$ in DGP

N	T	$P_{\hat{e}}$	PMSB	P_a	P_b	P_a^D	P_b^D	t_a^C	t_b^C
				Model 1 (Size): $F \sim \text{I}(1), e_{it} \sim \text{I}(1)$					
20	20	0.367	0.003	0.077	0.056	0.627	0.661	0.931	0.932
20	50	0.079	0.016	0.095	0.067	0.272	0.274	0.366	0.346
20	100	0.063	0.021	0.098	0.062	0.139	0.135	0.152	0.139
50	20	0.582	0.004	0.077	0.063	0.868	0.887	0.998	0.998
50	50	0.069	0.018	0.076	0.058	0.435	0.435	0.588	0.566
50	100	0.054	0.023	0.077	0.058	0.203	0.204	0.238	0.227
100	20	0.790	0.012	0.092	0.085	0.971	0.977	1.000	1.000
100	50	0.068	0.017	0.080	0.068	0.627	0.632	0.833	0.818
100	100	0.062	0.030	0.070	0.058	0.295	0.296	0.368	0.354
				Model 2 (Power): $F \sim \text{I}(0), e_{it} \sim \text{I}(0)$					
20	20	0.382	0.004	0.105	0.075	0.640	0.673	0.980	0.980
20	50	0.171	0.157	0.471	0.386	0.324	0.288	0.708	0.652
20	100	0.644	0.828	0.960	0.933	0.172	0.092	0.332	0.214
50	20	0.623	0.007	0.102	0.082	0.889	0.902	1.000	1.000
50	50	0.256	0.388	0.678	0.632	0.540	0.498	0.915	0.889
50	100	0.924	0.992	0.998	0.997	0.296	0.197	0.692	0.598
100	20	0.838	0.005	0.112	0.101	0.985	0.987	1.000	1.000
100	50	0.444	0.766	0.926	0.914	0.788	0.762	0.994	0.992
100	100	0.998	1.000	1.000	1.000	0.504	0.368	0.915	0.873
				Model 3 (Power): $F \sim \text{I}(0), e_{it}$ mixed I(0), I(1)					
20	20	0.373	0.003	0.092	0.068	0.630	0.662	0.980	0.979
20	50	0.149	0.107	0.379	0.300	0.315	0.285	0.710	0.667
20	100	0.514	0.575	0.814	0.757	0.173	0.109	0.385	0.262
50	20	0.614	0.005	0.079	0.067	0.876	0.894	1.000	1.000
50	50	0.204	0.244	0.521	0.471	0.506	0.477	0.917	0.899
50	100	0.794	0.892	0.963	0.950	0.268	0.196	0.754	0.675
100	20	0.836	0.004	0.090	0.081	0.983	0.987	1.000	1.000
100	50	0.320	0.481	0.752	0.719	0.735	0.711	0.995	0.993
100	100	0.975	0.995	0.999	0.999	0.404	0.318	0.947	0.927
				Model 4 (Power): $F \sim \text{I}(0), e_{it} \sim \text{I}(0)$					
20	20	0.367	0.006	0.114	0.085	0.656	0.686	0.939	0.939
20	50	0.143	0.117	0.390	0.314	0.299	0.274	0.380	0.324
20	100	0.506	0.639	0.844	0.796	0.149	0.094	0.148	0.087
50	20	0.610	0.010	0.122	0.102	0.885	0.898	0.998	0.997
50	50	0.198	0.265	0.528	0.478	0.476	0.449	0.608	0.563
50	100	0.747	0.846	0.903	0.891	0.239	0.183	0.245	0.182
100	20	0.823	0.016	0.156	0.144	0.975	0.979	1.000	1.000
100	50	0.332	0.559	0.762	0.737	0.696	0.676	0.824	0.796
100	100	0.936	0.943	0.959	0.955	0.388	0.313	0.407	0.322

Note: $P_{\hat{e}}$, PMSB, P_a, P_b, P_a^D, *and* P_b^D are tests based on PANIC residuals $\hat{e}$. The first four are defined in (5), (11), (8), and (9). The tests P_a^D and P_b^D are constructed in the same way as t_a and t_b, but they estimate ρ^+ from an autoregression using $\hat{e}$ with a constant and a linear trend. The tests t_a^C and t_b^C are MP tests that detrend the observable data.

key behind the drastic difference in the properties of P_a and P_b on the one hand, and P_a^D, P_b^D, t_a, and t_b on the other. Still, the $P_{\hat{e}}$, PMSB, P_a, and P_b tests require that N or T not be too small, or else the test is oversized. Overall, tests of nonstationarity in panel data with incidental trends are quite unreliable without N and T being reasonably large.

In a recent paper, Westerlund and Larsson (2009) 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); they also provide a small-sample bias correction. 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, P_b and PMSB are also PANIC-based pooled tests. Our simulations show that all PANIC-based pooled tests have good finite-sample properties.

6. CONCLUSION

In this paper, we (a) develop a PANIC-based estimate of the pooled autoregressive coefficient and (b) develop a PMSB test that does not rely on the pooled autoregressive coefficient. Upon comparing their finite-sample properties, we find that tests based on the autoregressive coefficient have no power against incidental trends whenever linear detrending is performed before estimating the pooled autoregressive parameter. The PMSB test, the original PANIC pooled test of the p values, and the new P_a and P_b tests all have satisfactory properties. None of these tests require a projection $\hat{e}$ on time trends.

It is worth emphasizing that tests that control cross-section correlation only permit hypotheses 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 hardly meaningful.

NOTES

1. A kernel estimate based on $\Delta\hat{e}_{it}$, although consistent under the null hypothesis that all units are nonstationary, is degenerate under the specific alternative that all units are stationary. Accordingly, nuisance parameters are estimated using $\hat{\varepsilon}_{it}$ instead of $\Delta\hat{e}_{it}$.
2. Estimation of $\omega_{\varepsilon i}^2$ is discussed in Perron and Ng (1998).
3. As long as the data are demeaned and detrended, the projection matrix M_1 must be used even if the true DGP is given by Model A.

REFERENCES

Anderson, T.W. (1984) *An Introduction to Multivariate Statistical Analysis*. Wiley.
Bai, J. (2003) Inferential theory for factor models of large dimensions. *Econometrica* 71, 135–172.

Bai, J. & S. Ng (2002) Determining the number of factors in approximate factor models. *Econometrica* 70, 191–221.

Bai, J. & S. Ng (2004) A PANIC attack on unit roots and cointegration. *Econometrica* 72, 1127–1177.

Breitung, J. & S. Das (2008) Testing for unit root in panels with a factor structure. *Econometric Theory* 24, 88–108.

Chang, Y. (2002) Nonlinear IV unit root tests in panels with cross-section dependency. *Journal of Econometrics* 110, 261–292.

Chang, Y. & W. Song (2002) Panel Unit Root Tests in the Presence of Cross-Section Heterogeneity. Manuscript, Rice University.

Choi, I. (2006) Combination unit root tests for cross-sectionally correlated panels. In D. Corbae, S.N. Durlauf, & B.E. Hansen (eds.), *Econometric Theory and Practice: Frontiers of Analysis and Applied Research: Essays in Honor of Peter C.B. Phillips*, pp. 311–333. Cambridge University Press.

Gengenbach, C., F. Palm, & J.P. Urbain (2009) Panel unit root tests in the presence of cross-sectional dependencies: Comparison and implications for modelling. *Econometric Review*, forthcoming.

Jang, M.J. & D. Shin (2005) Comparison of panel unit root tests under cross-sectional dependence. *Economics Letters* 89, 12–17.

Leeb, H. & B. Potscher (2008) Can one estimate the unconditional distribution of post-model-selection estimators? *Econometric Theory* 24, 338–376.

Levin, A., C.F. Lin, & J. Chu (2002) Unit root tests in panel data: Asymptotic and finite sample properties. *Journal of Econometrics* 98, 1–24.

Maddala, G.S. & S. Wu (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. & B. Perron (2004) Testing for a unit root in panels with dynamic factors. *Journal of Econometrics* 122, 81–126.

Moon, R., B. Perron, & P. Phillips (2007) Incidental trends and the power of panel unit root tests. *Journal of Econometrics* 141, 416–459.

Ng, S. & P. Perron (2001) Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69, 1519–1554.

O'Connell, P. (1998) The overvaluation of purchasing power parity. *Journal of International Economics* 44, 1–19.

Perron, P. & S. Ng (1996) Useful modifications to unit root tests with dependent errors and their local asymptotic properties. *Review of Economic Studies* 63, 435–465.

Perron, P. & S. Ng (1998) An autoregressive spectral density estimator at frequency zero for nonstationarity tests. *Econometric Theory* 14, 560–603.

Pesaran, H. (2007) A simple unit root test in the presence of cross-section dependence. *Journal of Applied Economics* 22, 265–312.

Phillips, P.C.B. & R. Moon (1999) Linear regression limit theory for nonstationary panel data. *Econometrica* 67, 1057–1111.

Phillips, P.C.B. & P. Perron (1988) Testing for a unit root in time series regression. *Biometrika* 75, 335–346.

Phillips, P.C.B. & W. Ploberger (2002) Optimal Testing for Unit Roots in Panel Data. Mimeo, University of Rochester.

Phillips, P.C.B. & D. Sul (2003) Dynamic panel estimation and homogeneity testing under cross-section dependence. *Econometrics Journal* 6, 217–259.

Said, S.E. & D.A. Dickey (1984) Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71, 599–607.

Sargan, J.D. & A. Bhargava (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, Department of Economics, Harvard University.

Westerlund, J. & R. Larsson (2009) A note on the pooling of individual PANIC unit root tests. *Econometric Theory* 25, 1851–1868.

APPENDIX

Assumptions A–E are assumed when analyzing the properties of PANIC residuals $\widehat{e}_{it}$.

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 of Lemma 1. From Bai and Ng (2004, p. 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 preceding expression 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

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{e}_{it}^2 = \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}^2 + 2\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}A_{it} + \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}A_{it}^2 = I + II + III. \quad \textbf{(A.1)}$$

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 Bai and Ng (2004),

$$\begin{aligned} III &= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}A_{it}^2 \le 3\frac{1}{T}\left(\frac{1}{N}\sum_{i=1}^{N}e_{i1}^2\right) \\ &\quad + 3\left(\frac{1}{T^2}\sum_{t=1}^{T}\|V_t\|^2\right)\left(\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i H^{-1}\|^2\right) \\ &\quad + \left(\frac{1}{N}\sum_{i=1}^{N}\|d_i\|^2\right)\frac{1}{T^2}\sum_{t=1}^{T}\|\widehat{F}_t\|^2 \\ &= O_p(T^{-1}) + O_p(N^{-1}) + O_p\left(\left[\min[N^2, T]\right]^{-1}\right) = O_p(C_{NT}^{-2}). \end{aligned}$$

Next, consider II (ignoring the factor of 2),

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

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

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

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

By (A.4) of Bai and Ng (2004, p. 1157), $N^{-1/2} \left(\frac{1}{T^2} \sum_{t=1}^{T} \|H^{-1} V_t\|^2 \right)^{1/2} = O_p(N^{-1}) = O_p(C_{NT}^{-2})$. 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:

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

$$= O_p(T^{-1/2}) \left(\frac{1}{T} \sum_{t=1}^{T} \left[\frac{1}{N} \sum_{i=1}^{N} e_{it} d_i \right]^2 \right)^{1/2}.$$

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

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

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 that for each t,

$$\mathrm{E} \left(\frac{1}{N} \sum_{i=1}^{N} e_{it} \frac{1}{T} \sum_{s=1}^{T} f_s \varepsilon_{is} \right)^2 = O(T^{-1}) + O(N^{-1}), \tag{A.3}$$

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 preceding expression is proved for the case of $t = T$ subsequently (see the proof of (A.7)); the argument is exactly the same for every t. Thus $II = O_p(C_{NT}^{-2})$, and the lemma follows. ■

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

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

Proof of Lemma 2. 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 Lemma 2 is a consequence of Lemma 3, which follows. ■

LEMMA 3. *If $N/T^2 \to 0$, then the PANIC residuals $\widehat{e}_{it}$ satisfy, for $p=-1,0$,*

(i) $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}(\widehat{e}_{i1}^2 - e_{i1}^2) = O_p(\sqrt{N}/T),$

(ii) $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}(\widehat{e}_{iT}^2 - e_{iT}^2) = O_p(\sqrt{N}/T) + O_p(C_{NT}^{-1}),$

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

Proof of Lemma 3. Proof of (i). Because $\widehat{e}_{i1}$ is defined to be zero, it follows that the left-hand side of (i) is $(\sqrt{N}/T)\left(\frac{1}{N}\sum_{i=1}^{N}e_{i1}^2\right) = o_p(1)$ if $\sqrt{N}/T \to 0$.

Proof of (ii). From $\widehat{e}_{iT} = e_{iT} + A_{iT}$, it is sufficient to show that (a) $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}A_{iT}^2 = O_p(\sqrt{N}/T)$ and that (b) $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}e_{iT}A_{iT} = O_p(\sqrt{N}/T) + O_p\left(C_{NT}^{-1}\right)$. 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).

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}e_{iT}A_{iT} = \frac{-1}{\sqrt{NT}}\sum_{i=1}^{N}e_{iT}e_{i1} + \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}e_{iT}\lambda_i' H^{-1}V_T - \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}e_{iT}d_i'\widehat{F}_T. \tag{A.4}$$

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

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

(see Bai and Ng, 2004, p. 1157). For the last term of (A.4), from $T^{-1/2}\widehat{F}_T = O_p(1)$, we need to bound

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}e_{iT}d_i = \left(\frac{N}{T}\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}e_{iT}d_i\right). \tag{A.5}$$

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 (A.5) is $o_p(1)$ if $N/T \to 0$. But $N/T \to 0$ is not necessary. To see this, first use d_i in (A.2) to obtain

$$\frac{1}{N}\sum_{i=1}^{N}e_{iT}d_i = \frac{1}{N}\sum_{i=1}^{N}e_{iT}\frac{1}{T}\sum_{s=1}^{T}f_s\varepsilon_{is} + O_p(T^{1/2})O_p(C_{NT}^{-2}). \tag{A.6}$$

We shall show

$$\mathrm{E}\left(\frac{1}{N}\sum_{i=1}^{N}e_{iT}\frac{1}{T}\sum_{s=1}^{T}f_s\varepsilon_{is}\right)^2 = O(T^{-1}) + O(N^{-1}). \tag{A.7}$$

From $e_{iT} = \sum_{t=1}^{T} \varepsilon_{it}$, the left-hand side of (A.7) is

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{T^2}\sum_{s,k,t,h}^{T}\mathrm{E}(f_s f_k \varepsilon_{is}\varepsilon_{it}\varepsilon_{jk}\varepsilon_{jh}). \qquad \textbf{(A.8)}$$

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

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

because $T^{-1}\sum_{s,k}^{T}|\mathrm{E}(f_s f_k)| \le M$ under weak correlation for f_s. If ε_{it} is serially correlated, then the sum in (A.8) for $i \neq j$ is bounded by

$$\begin{aligned}&\frac{1}{N^2}\sum_{i\neq j}^{N}\left(\frac{1}{T^2}\sum_{s,k}^{T}|\mathrm{E}(f_s f_k)|\right)\left(\max_s \sum_{t=1}^{T}|\mathrm{E}(\varepsilon_{is}\varepsilon_{it})|\right)\left(\max_k \sum_{h=1}^{T}|\mathrm{E}(\varepsilon_{jk}\varepsilon_{jh})|\right)\\ &\quad \le \frac{1}{N^2}\sum_{i\neq j}^{N}\left(\frac{1}{T}\sum_{s,k}^{T}|\mathrm{E}(f_s f_k)|\right)\left(\sum_{\ell=0}^{\infty}|\gamma_i(\ell)|\right)\left(\sum_{\ell=0}^{\infty}|\gamma_j(\ell)|\right),\end{aligned}$$

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 $i = j$. Because $\frac{1}{T^2}\sum_{s,t,k,h}^{T}\mathrm{E}(f_s f_k \varepsilon_{is}\varepsilon_{it}\varepsilon_{ik}\varepsilon_{ih}) = O(1)$, we have $\frac{1}{N^2}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{s,t,k,h}^{T}\mathrm{E}(f_s f_k \varepsilon_{is}\varepsilon_{it}\varepsilon_{ik}\varepsilon_{ih}) = O(N^{-1})$, proving (A.7). Combining (A.5)–(A.7),

$$\begin{aligned}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} e_{iT}d_i &= \left(\frac{N}{T}\right)^{1/2}\left[O_p(T^{-1/2}) + O_p(N^{-1/2}) + T^{1/2}O_p(C_{NT}^{-2})\right]\\ &= O_p(\sqrt{N}/T) + O_p\left(C_{NT}^{-1}\right).\end{aligned}$$

This proves (b). Combining (a) and (b) yields (ii).

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 Bai and Ng (2004, p. 1158), $T^{-1}\sum_{t=2}^{T}a_{it}^2 = O_p(C_{NT}^{-2})$; thus the second term on the right-hand side of the preceding expression is bounded by $\sqrt{N}O_p(C_{NT}^{-2})$. Consider the first term

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

By the Cauchy–Schwartz inequality,

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

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

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

The preceding expression is $T^{-1/2}\left(\sum_{i=1}^{N} d_i^2\right)^{1/2} O_p(1) = O_p(\sqrt{N}/T)$, because $\sum_{i=1}^{N} d_i^2 = O_p(N/\min(N^2,T))$. Thus (iii) is equal to $\sqrt{N}O_p(C_{NT}^{-2}) + O_p(C_{NT}^{-1}) + O_p(\sqrt{N}/T) = O_p(C_{NT}^{-1}) + O_p(\sqrt{N}/T)$. ■

LEMMA 4. *For $p = 1$, the PANIC residuals satisfy, with $N, T \to \infty$,*

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

(ii) $$\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{e}_{it-1}\Delta\widehat{e}_{it} = \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{e}_{it-1}\Delta\widetilde{e}_{it} + O_p(\sqrt{N}/T) + O_p(C_{NT}^{-1}),$$

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

Proof of Lemma 4. This is an argument almost identical to that in the proof of Lemmas 1 and 2. The details are omitted. Note that when $p = 1$, the PANIC residuals $\widehat{e}_{it}$ are estimating $\widetilde{e}_{it}$; see Bai and Ng (2004). ■

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 5. *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) $$\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{e}_{it-1}^{\tau}\Delta\widehat{e}_{it}^{\tau} = \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it-1}^{\tau}\Delta e_{it}^{\tau} + O_p(\sqrt{N}/T) + O_p(C_{NT}^{-1}).$$

Proof of Lemma 5. 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. ■

LEMMA 6. *Suppose that Assumption C holds. Under $\rho_i = 1$ for all i, we have, as $N, T \to \infty$,*

(i) $\frac{1}{NT^2} \sum_{i=1}^N \sum_{t=1}^T e_{it-1}^2 \xrightarrow{p} \frac{1}{2}\omega_\varepsilon^2$,

(ii) with $N/T \to 0$, $\sqrt{N}\left(\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T e_{it-1}\varepsilon_{it} - \bar{\lambda}_N\right) \xrightarrow{d} N\left(0, \frac{1}{2}\phi_\varepsilon^4\right)$,

(iii) $\frac{1}{NT^2} \sum_{i=1}^N \sum_{t=1}^T \left(e_{it-1}^\tau\right)^2 \xrightarrow{p} \frac{1}{15}\omega_\varepsilon^2$,

(iv) with $N/T \to 0$, $\sqrt{N}\left(\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T e_{it-1}^\tau\varepsilon_{it} + \frac{1}{2}\bar{\sigma}_N^2\right) \xrightarrow{d} N\left(0, \frac{1}{60}\phi_\varepsilon^4\right)$,

(v) $\frac{1}{NT^2} \sum_{i=1}^N \sum_{t=1}^T (\widetilde{e}_{it})^2 \xrightarrow{p} \frac{1}{6}\omega_\varepsilon^2$,

where e_{it}^τ is the demeaned and detrended version of e_{it} and $\widetilde{e}_{it} = e_{it} - e_{i1} - e_{iT} - e_{i1}/(T-1)(t-1)$, $\bar{\lambda}_N = \frac{1}{N}\sum_{i=1}^N \lambda_{\varepsilon i}$, $\bar{\sigma}_N^2 = \frac{1}{N}\sum_{i=1}^N \sigma_{\varepsilon i}^2$, and $\lambda_{\varepsilon i}$, $\sigma_{\varepsilon i}^2$, ω_ε^2, and ϕ_ε^4 are all defined in Section 2.

Proof of Lemma 6. Parts (i) and (ii) are from Lemma A.2 of Moon and Perron (2004). Parts (iii), (iv), and (v) can be proved using similar arguments as in Moon and Perron (2004). These are all joint limits. To provide intuition and justification for the parameters involved, we give a brief explanation of the preceding results using sequential argument. Consider (i). For each i, as $T \to \infty$, $\frac{1}{T^2}\sum_{t=2}^T e_{it-1}^2 \xrightarrow{d} \omega_{\varepsilon i}^2 U_i$, where $U_i = \int_0^1 W_i(r)^2 dr$ with $\mathrm{E}U_i = \frac{1}{2}$. Thus, for fixed N, $\frac{1}{N}\sum_{i=1}^N \frac{1}{T^2}\sum_{t=2}^T e_{it-1}^2 \xrightarrow{p} \frac{1}{2}\bar{\omega}_N^2$, where $\bar{\omega}_N^2 = \frac{1}{N}\sum_{i=1}^N \omega_{\varepsilon i}^2$. Part (i) is obtained by letting N go to infinity. Consider (ii). For each i, $\frac{1}{T}\sum_{t=2}^T e_{it-1}\Delta e_{it} \xrightarrow{d} \omega_{\varepsilon i}^2 Z_i + \lambda_{\varepsilon i}$, where $Z_i = \int_0^1 W_i(r)\,dW_i(r)$ and $\lambda_{\varepsilon i}$ is one-sided long-run variance of $\varepsilon_{it} = \Delta e_{it}$ (i.e., $\lambda_{\varepsilon i} = [\omega_{\varepsilon i}^2 - \sigma_{\varepsilon i}^2]/2$). Thus for fixed N, $\sqrt{N}\left(\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T e_{it-1}\varepsilon_{it} - \bar{\lambda}_N\right) \xrightarrow{d} \frac{1}{\sqrt{N}}\sum_{i=1}^N \omega_{\varepsilon i}^2 Z_i$. Because Z_i are independent and identically distributed (i.i.d.), zero mean, and $\mathrm{var}(Z_i) = \frac{1}{2}$, we obtain (ii) by the central limit theorem as $N \to \infty$. Similar to (i), (iii) follows from $\frac{1}{T^2}\sum_{t=2}^T (e_{it-1}^\tau)^2 \xrightarrow{d} \omega_{\varepsilon i}^2 U_i^\tau$, where $U_i^\tau = \int_0^1 W^\tau(r)^2 dr$ with $\mathrm{E}(U_i^\tau) = 1/15$. For (iv), $\frac{1}{T}\sum_{t=2}^T e_{it-1}^\tau \Delta e_{it}^\tau \xrightarrow{d} \omega_{\varepsilon i}^2 Z_i^\tau + \lambda_{\varepsilon i}$, where $Z_i^\tau = \int_0^1 W^\tau(r)\,dW(r)$. From $\mathrm{E}(Z_i^\tau) = -\frac{1}{2}$, we have $\mathrm{E}(\omega_{\varepsilon i}^2 Z_i^\tau) + \lambda_{\varepsilon i} = -\sigma_{\varepsilon i}^2/2$. Thus, $\sqrt{N}\left(\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T e_{it-1}^\tau\varepsilon_{it} + \frac{1}{2}\bar{\sigma}_N^2\right) \xrightarrow{d} \frac{1}{\sqrt{N}}\sum_{i=1}^N \omega_{\varepsilon i}^2 Z_i^\tau$. Because $\mathrm{var}(Z_i^\tau) = \frac{1}{60}$, (iv) is obtained by the central limit theorem when $N \to \infty$. For (v), $\frac{1}{T^2}\sum_{t=1}^T (\widetilde{e}_{it})^2 \xrightarrow{d} \omega_{\varepsilon i}^2 V_i$ with $V_i = \int_0^1 B_i^2(r)\,dr$, where B_i is a Brownian bridge. Part (v) follows from $\mathrm{E}(V_i) = 1/6$. The proofs of these results under joint limits are more involved; the details are omitted because of similarity to the arguments of Moon and Perron (2004). Note that joint limits in (iii) and (iv) require $N/T \to 0$. Also see the detailed proof for the joint limit in Lemma 8. ■

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}. Let ρ^+ (resp. $\widehat{\rho}^+$) be the bias-corrected estimator based on the true idiosyncratic error matrix e and $\bar{\lambda}_N$ (resp. $\widehat{e}$ and the estimated $\bar{\lambda}_N$). Under the null of $\rho_i = 1$ for all i, we have

$$\sqrt{N}T(\rho^+ - 1) = \sqrt{N}T\frac{\mathrm{tr}\left(e_{-1}'\Delta e\right) - NT\bar{\lambda}_N}{\mathrm{tr}(e_{-1}'e_{-1})} = \frac{\sqrt{N}\left[\mathrm{tr}\left(\frac{1}{NT}e_{-1}'\Delta e\right) - \bar{\lambda}_N\right]}{\frac{1}{NT^2}\mathrm{tr}(e_{-1}'e_{-1})}. \qquad \textbf{(A.9)}$$

By Lemma 6(i) and (ii), (A.9) converges in distribution to $N\left(0, \frac{2\phi_\varepsilon^4}{\omega_\varepsilon^4}\right)$ as $N, T \to \infty$ with $N/T \to 0$. This limiting distribution does not change when $\bar{\lambda}_N$ is replaced by its estimate $\widehat{\lambda}_N$ because $\sqrt{N}(\widehat{\lambda}_N - \bar{\lambda}_N) = o_p(1)$; see Moon and Perron (2004). By Lemmas 1 and 2 the limiting distribution continues to hold when e is replaced by $\widehat{e}$. That is, $\sqrt{N}T(\widehat{\rho}^+ - 1) \xrightarrow{d} N\left(0, \frac{2\phi_\varepsilon^4}{\omega_\varepsilon^2}\right)$. Thus, $P_a = \sqrt{N}T(\widehat{\rho}^+ - 1)/\sqrt{2\widehat{\phi}_\varepsilon^4/\widehat{\omega}_\varepsilon^4} \xrightarrow{d} N(0,1)$, as $N, T \to \infty$ with $N/T \to 0$.

For the P_b test, multiply equation (A.9) on each side by $[(1/NT^2)\mathrm{tr}(e'_{-1}e_{-1})]^{1/2}$, whose limit is $(\omega_\varepsilon/2)^{1/2}$ by part (i) of Lemma 6. We obtain

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

which converges to $N(0, \phi_\varepsilon^4/\omega_\varepsilon^2)$, as $N, T \to \infty$ with $N/T \to \infty$. It follows that $P_b = \sqrt{N}T(\widehat{\rho}^+ - 1)\left((1/NT^2)\mathrm{tr}(\widehat{e}'_{-1}\widehat{e}_{-1})\right)^{1/2}\sqrt{\widehat{\omega}_\varepsilon^2/\widehat{\phi}_\varepsilon^4} \xrightarrow{d} N(0,1)$.

For $p = 1$, recall that $\widehat{e}_{it}$ estimates $\widetilde{e}_{it} = e_{it} - e_{i1} - e_{iT} - e_{i1}/(T-1)(t-1)$. Let $\widetilde{e}$ be the matrix consisting of elements $\widetilde{e}_{it}$. Note that $\widetilde{e}_{iT} \equiv 0$ for all i. And $\sum_{t=1}^T \widetilde{e}_{it-1}\Delta\widetilde{e}_{it} = \frac{1}{2}\widetilde{e}_{iT}^2 - \frac{1}{2}\sum_{t=1}^T(\Delta\widetilde{e}_{it})^2 = -\frac{1}{2}\sum_{t=1}^T(\Delta\widetilde{e}_{it})^2$. But $\Delta\widetilde{e}_{it} = \varepsilon_{it} - \bar{\varepsilon}_i$. Thus $\frac{1}{T}\sum_{t=1}^T \widetilde{e}_{it-1}\Delta\widetilde{e}_{it} = -\frac{1}{2T}\sum_{t=1}^T(\varepsilon_{it} - \bar{\varepsilon}_i)^2 \xrightarrow{p} -\frac{1}{2}\sigma_{\varepsilon i}^2$. Thus $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \widetilde{e}_{it-1}\Delta\widetilde{e}_{it} \xrightarrow{p} -\frac{1}{2}\sigma_\varepsilon^2$. Together with Lemma 6(v),

$$\frac{\frac{1}{NT}\mathrm{tr}(\widetilde{e}'_{-1}\Delta\widetilde{e})}{\frac{1}{NT^2}\mathrm{tr}(\widetilde{e}'_{-1}\widetilde{e}_{-1})} \xrightarrow{p} \frac{-\sigma_\varepsilon^2/2}{\omega_\varepsilon^2/6} = -3(\sigma_\varepsilon^2/\omega_\varepsilon^2).$$

For given N, the preceding limit is $-3\bar{\sigma}_N^2/\bar{\omega}_N^2$, where $\bar{\sigma}_N^2 = \frac{1}{N}\sum_{i=1}^N \sigma_{\varepsilon i}^2$ and $\bar{\omega}_N^2 = \frac{1}{N}\sum_{i=1}^N \omega_{\varepsilon i}^2$. Again, let ρ^+ denote the bias-corrected estimator based on $\widetilde{e}$ and the true parameters $(\bar{\sigma}_N^2, \bar{\omega}_N^2)$, i.e.,

$$\rho^+ = \frac{\mathrm{tr}(\widetilde{e}'_{-1}\widetilde{e})}{\mathrm{tr}(\widetilde{e}'_{-1}\widetilde{e}_{-1})} + \frac{3}{T}\frac{\bar{\sigma}_N^2}{\bar{\omega}_N^2}.$$

Then

$$T(\rho^+ - 1) = \frac{\frac{1}{NT}\mathrm{tr}(\widetilde{e}'_{-1}\Delta\widetilde{e})}{(1/NT^2)\mathrm{tr}(\widetilde{e}'_{-1}\widetilde{e}_{-1})} + 3\frac{\bar{\sigma}_N^2}{\bar{\omega}_N^2} = \frac{A}{B} + \frac{\bar{\sigma}_N^2/2}{\bar{\omega}_N^2/6},$$

where $A = \frac{1}{NT}\mathrm{tr}(\widetilde{e}'_{-1}\Delta\widetilde{e})$ and $B = (1/NT^2)\mathrm{tr}(\widetilde{e}'_{-1}\widetilde{e}_{-1})$. It follows that

$$\sqrt{N}T(\rho^+ - 1) = \frac{1}{B}\sqrt{N}\left(A + \bar{\sigma}_N^2/2\right) + \frac{3}{B}\frac{\bar{\sigma}_N^2}{\bar{\omega}_N^2}\sqrt{N}\left(B - \bar{\omega}_N^2/6\right).$$

Note

$$\sqrt{N}(A + \bar{\sigma}_N^2/2) = -\frac{1}{2}\frac{1}{\sqrt{N}T}\sum_{i=1}^N\sum_{t=1}^T[(\varepsilon_{it} - \bar{\varepsilon}_i)^2 - \sigma_{\varepsilon i}^2] = O_p\left(\frac{1}{\sqrt{T}}\right).$$

By Lemma 9, as $N, T \to \infty$ with $N/T \to 0$,

$$\sqrt{N}(B - \bar{\omega}_N^2/6) = \sqrt{N}\left[\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}(\widetilde{e}_{it})^2 - \bar{\omega}_N^2/6\right] \xrightarrow{d} N(0, \phi_\varepsilon^4/45).$$

From $B \xrightarrow{p} \omega_\varepsilon^2/6$ by Lemma 6(v), we have

$$\sqrt{N}T(\rho^+ - 1) \xrightarrow{d} 18\frac{\sigma_\varepsilon^2}{\omega_\varepsilon^4} N\left(0, \frac{\phi_\varepsilon^4}{45}\right) = N\left(0, \frac{36}{5}\frac{\phi_\varepsilon^4\sigma_\varepsilon^4}{\omega_\varepsilon^8}\right).$$

By Lemma 4, the result continues to hold when $\widetilde{e}_{it}$ is replaced by $\widehat{e}_{it}$ and $(\bar{\sigma}_N^2, \bar{\omega}_N^2)$ is replaced by $(\widehat{\sigma}_\varepsilon^2, \widehat{\omega}_\varepsilon^2)$ because $\sqrt{N}(\widehat{\sigma}_\varepsilon^2 - \bar{\sigma}_N^2) = o_p(1)$ and $\sqrt{N}(\widehat{\omega}_\varepsilon^2 - \bar{\omega}_N^2) = o_p(1)$; see Moon and Perron (2004). That is,

$$\sqrt{N}T(\widehat{\rho}^+ - 1) \xrightarrow{d} N\left(0, \frac{36}{5}\frac{\phi_\varepsilon^4\sigma_\varepsilon^4}{\omega_\varepsilon^8}\right).$$

Thus

$$P_a = \sqrt{N}T(\widehat{\rho}^+ - 1)\Big/\sqrt{\frac{36}{5}\frac{\widehat{\phi}_\varepsilon^4\widehat{\sigma}_\varepsilon^4}{\widehat{\omega}_\varepsilon^8}} \xrightarrow{d} N(0, 1).$$

For P_b, using $[(1/NT^2)\mathrm{tr}(\widehat{e}_{-1}'\widehat{e}_{-1}]^{1/2} \xrightarrow{p} (\omega_\varepsilon^2/6)^{1/2}$, we have

$$\sqrt{N}T(\widehat{\rho}^+ - 1)\left[\frac{1}{NT^2}\mathrm{tr}(\widehat{e}_{-1}'\widehat{e}_{-1})\right]^{1/2} \xrightarrow{d} N\left(0, \frac{6}{5}\frac{\phi_\varepsilon^4\sigma_\varepsilon^4}{\omega_\varepsilon^6}\right).$$

Normalizing leads to $P_b \xrightarrow{d} N(0, 1)$. This completes the proof of Theorem 1. ■

Remark. If demeaning and detrending are performed when $p = 1$, the following analysis will be applicable. The bias-corrected estimator is

$$\sqrt{N}T(\rho^+ - 1) = \sqrt{N}T\frac{\mathrm{tr}\left(e_{-1}^{\tau\prime}\Delta e^\tau\right) + NT\bar{\sigma}_N^2/2}{\mathrm{tr}\left(e_{-1}^{\tau\prime}e_{-1}^\tau\right)} = \frac{\sqrt{N}\left(\frac{1}{NT}\mathrm{tr}(e_{-1}^{\tau\prime}\Delta e^\tau) + \bar{\sigma}_N^2/2\right)}{\frac{1}{NT^2}\mathrm{tr}\left(e_{-1}^{\tau\prime}e_{-1}^\tau\right)}.$$

By Lemma 6(iii) and (iv), the preceding expression converges to $N\left(0, \frac{15\phi_\varepsilon^4}{4\omega_\varepsilon^4}\right)$, as $N, T \to \infty$ with $N/T \to 0$. Replacing e^τ by $\widehat{e}^\tau$ and replacing $\bar{\sigma}_N^2$ by $\widehat{\sigma}_N^2$ do not change the limit because of $\sqrt{N}\left(\widehat{\sigma}_N^2 - \bar{\sigma}_N^2\right) = o_p(1)$; see Moon and Perron (2004) and Lemma 5. This implies

$$\sqrt{N}T(\widehat{\rho}^+ - 1) \xrightarrow{d} N\left(0, \frac{15\phi_\varepsilon^4}{4\omega_\varepsilon^4}\right)$$

as $N, T \to \infty$ with $N/T \to 0$. Thus, $P_a = \sqrt{N}T(\widehat{\rho}^+ - 1)\Big/\sqrt{\frac{15\widehat{\phi}_\varepsilon^4}{4\widehat{\omega}_\varepsilon^4}} \xrightarrow{d} N(0, 1)$. From the limit for $\sqrt{N}T(\widehat{\rho}^+ - 1)$ and Lemma 6(iii), $\sqrt{N}T(\widehat{\rho}^+ - 1)\left((1/NT^2)\mathrm{tr}(\widehat{e}_{-1}^{\tau\prime}\widehat{e}_{-1}^\tau)\right)^{1/2} \xrightarrow{d} N\left(0, \frac{\phi_\varepsilon^4}{4\omega_\varepsilon^2}\right)$. It follows that $P_b = \sqrt{N}T(\widehat{\rho}^+ - 1)\left((1/NT^2)\mathrm{tr}(\widehat{e}_{-1}^{\tau\prime}\widehat{e}_{-1}^\tau)\right)^{1/2}\sqrt{4\widehat{\omega}_\varepsilon^2/\widehat{\phi}_\varepsilon^4} \xrightarrow{d} N(0, 1)$.

LEMMA 7. *The PANIC residuals satisfy, as $N, T \to \infty$ with $N/T^2 \to 0$,*

(i) for $p = -1, 0$,

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

(ii) for $p = 1$,

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

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

Proof of Lemma 7. Proof of (i). This follows from Lemma 1 upon multiplying by $N^{1/2}$ 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). This follows from Lemma 4(i) upon multiplying by $\sqrt{N}$ on each side and noting $\sqrt{N}O_p(C_{NT}^{-2}) = o_p(1)$. ■

LEMMA 8. *Under Assumption C and $\rho_i = 1$ for all i, as $N, T \to \infty$ with $N/T \to 0$,*

$$\sqrt{N}\left[\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}^2 - \bar{\omega}_N^2\Big/2\right] \xrightarrow{d} N(0, \phi_\varepsilon^4/3),$$

where $\bar{\omega}_N^2 = \frac{1}{N}\sum_{i=1}^{N}\omega_{\varepsilon i}^2$ and $\phi_\varepsilon^4 = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\omega_{\varepsilon i}^4$.

Proof of Lemma 8. We first give a sequential argument, which provides a useful intuition. For each i, $\frac{1}{T^2}\sum_{t=2}^{T}e_{it}^2 \xrightarrow{d} \omega_{\varepsilon i}^2 U_i$, where $U_i = \int_0^1 W_i(r)^2 dr$ with $\mathrm{E}\int_0^1 W_i(r)^2\,dr = \frac{1}{2}$ and $\mathrm{var}(U_i) = \frac{1}{3}$. Thus, the sequential limit theorem implies, for fixed N,

$$\sqrt{N}\left[\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}^2 - \bar{\omega}_N^2/2\right] \xrightarrow{d} \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega_{\varepsilon i}^2\left(U_i - \frac{1}{2}\right), \quad \text{as } T \to \infty.$$

Because U_i are i.i.d. with mean $\frac{1}{2}$ and variance $\frac{1}{3}$, from the central limit theorem over the cross sections, the right-hand side of the preceding expression converges in distribution to $N(0, \phi_\varepsilon^4/3)$, as $N \to \infty$.

The argument for a joint limiting theory is more involved. From the Beveridge–Nelson decomposition,

$$\varepsilon_{it} = d_i(1)v_{it} + \varepsilon_{it-1}^* - \varepsilon_{it}^*,$$

where $\varepsilon_{it}^* = \sum_{j=0}^{\infty}d_{ij}^* v_{it-j}$ with $d_{ij}^* = \sum_{k=j+1}^{\infty}d_{ik}$. The assumption on $d_i(L)$ ensures that $\mathrm{E}[(\varepsilon_{it}^*)^2]$ is bounded. Let $S_{it} = \sum_{s=1}^{t}v_{is}$. The cumulative sum of the preceding expression gives

$$e_{it} = d_i(1)S_{it} + \varepsilon_{i0}^* - \varepsilon_{it}^*.$$

Taking the square on each side and then summing over i and t,

$$\frac{1}{\sqrt{N}T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}^2=\frac{1}{\sqrt{N}T^2}\sum_{i=1}^{N}d_i(1)^2\sum_{t=1}^{T}S_{it}^2+\frac{1}{\sqrt{N}T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}d_i(1)S_{it}(\varepsilon_{i0}^*-\varepsilon_{it}^*)$$
$$+\frac{\sqrt{N}}{T}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}(\varepsilon_{i0}^*-\varepsilon_{it}^*)^2.$$

The last term is $o_p(1)$ if $\sqrt{N}/T\to 0$. By the Cauchy–Schwarz inequality, the middle term on the right-hand side is bounded by

$$(N/T)^{1/2}\frac{1}{N}\sum_{i=1}^{N}\left[|d_i(1)|\left(\frac{1}{T^2}\sum_{t=1}^{T}S_{it}^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}(\varepsilon_{i0}^*-\varepsilon_{it}^*)^2\right)^{1/2}\right]=(N/T)^{1/2}O_p(1),$$

which is $o_p(1)$ if $N/T\to 0$. Thus, if $N/T\to 0$, we have

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

Let $Y_{iT}=\frac{1}{T^2}\sum_{t=1}^{T}(S_{it}^2-\mathrm{E}S_{it}^2)$, where $\sum_{t=1}^{T}\mathrm{E}S_{it}^2=\frac{1}{2}(T+1)T$. Notice that $\omega_{\varepsilon i}^2=d_i(1)^2$. We have

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\frac{1}{T^2}\sum_{t=1}^{T}e_{it}^2-\bar{\omega}_N^2/2\right]=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega_{\varepsilon i}^2Y_{iT}+\frac{\sqrt{N}}{2T}\bar{\omega}_N^2+o_p(1).$$

The variables Y_{iT} are i.i.d. over i, having zero mean and finite variance. Furthermore, $Y_{iT}\xrightarrow{d}U_i-\frac{1}{2}$. Direct calculation shows that $\mathrm{E}Y_{iT}^2\to\frac{1}{3}$, which is equal to $\mathrm{E}(U_i-\frac{1}{2})^2$. This implies that Y_{iT} is uniformly integrable over T. The rest of the conditions of Theorem 3 of Phillips and Moon (1999) are satisfied under our assumptions. Thus by their Theorem 3, as $N,T\to\infty$ jointly,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega_{\varepsilon i}^2Y_{iT}\xrightarrow{d}N(0,\phi_\varepsilon^4/3).$$

This completes the proof of Lemma 8. ∎

LEMMA 9. *Under Assumption C and $\rho_i=1$ for all i, and as $N,T\to\infty$ with $N/T\to 0$, we have*

$$\sqrt{N}\left[\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}(\widetilde{e}_{it})^2-\bar{\omega}_N^2/6\right]\xrightarrow{d}N(0,\phi_\varepsilon^4/45),$$

where $\widetilde{e}_{it}=e_{it}-(e_{iT}-e_{i1})(t-1)/(T-1)$ and $\bar{\omega}_N^2$ and ϕ_ε^4 are defined in Lemma 8. ∎

Proof of Lemma 9. Again, we first consider a sequential argument. For each fixed i, $\frac{1}{T^2}\sum_{t=2}^{T}\widetilde{e}_{it}^2\xrightarrow{d}\omega_{\varepsilon i}^2V_i$, as $T\to\infty$, where $V_i=\int_0^1B_i(r)^2dr$ with B_i a Brownian bridge, and so $\mathrm{E}V_i=1/6$ and $\mathrm{var}(V_i)=1/45$. Thus, as $T\to\infty$ with fixed N,

$$\sqrt{N}\left[\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{e}_{it}^2-\bar{\omega}_N^2/6\right]\xrightarrow{d}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega_{\varepsilon i}^2\left(V_i-\frac{1}{6}\right).$$

Letting N go to infinity and by the central limit theorem over i, we obtain the limiting distribution as stated in the lemma. The proof for the joint limit follows the same argument as in Lemma 8. The details are omitted. ■

Proof of Theorem 2. Consider the case for $p = -1, 0$. By Lemma 7(i),

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\frac{1}{T^2}\sum_{t=1}^{T}(\widehat{e}_{it})^2 - \widehat{\omega}_\varepsilon^2/2\right] = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\left(\frac{1}{T^2}\sum_{t=1}^{T}e_{it}^2\right) - \bar{\omega}_N^2/2\right]$$
$$-\sqrt{N}(\widehat{\omega}_\varepsilon^2 - \bar{\omega}_N^2)/2 + o_p(1).$$

By Lemma 8, the first term on the right-hand side converges in distribution to $N(0, \phi_\varepsilon^4/3)$ jointly as $N, T \to \infty$ with $N/T \to 0$. In addition, $\sqrt{N}(\widehat{\omega}^2 - \bar{\omega}_N^2) = o_p(1)$, if $N/T \to 0$ (see Moon and Perron, 2004). It follows that

$$PMSB = \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\frac{1}{T^2}\sum_{t=1}^{T}(\widehat{e}_{it})^2 - \widehat{\omega}_\varepsilon^2/2\right]}{\sqrt{\widehat{\phi}_\varepsilon^4/3}} \xrightarrow{d} N(0,1).$$

Next consider $p = 1$. By Lemma 7(ii),

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\frac{1}{T^2}\sum_{t=1}^{T}(\widehat{e}_{it})^2 - \widehat{\omega}_\varepsilon^2/6\right] = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\frac{1}{T^2}\sum_{t=1}^{T}(\widetilde{e}_{it})^2 - \bar{\omega}_N^2/6\right]$$
$$-\sqrt{N}(\widehat{\omega}_\varepsilon^2 - \bar{\omega}_N^2)/6 + o_p(1).$$

By Lemma 9, the first term on the right-hand side converges in distribution to $N(0, \phi_\varepsilon^4/45)$ jointly as $N, T \to \infty$ with $N/T \to 0$. The second term, $\sqrt{N}(\widehat{\omega}_\varepsilon^2 - \bar{\omega}_N^2)$, is again $o_p(1)$ with $N/T \to 0$. It follows that

$$PMSB = \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\frac{1}{T^2}\sum_{t=1}^{T}(\widehat{e}_{it})^2 - \widehat{\omega}_\varepsilon^2/6\right]}{\sqrt{\widehat{\phi}_\varepsilon^4/45}} \xrightarrow{d} N(0,1),$$

as $N, T \to \infty$ with $N/T \to 0$. This completes the proof of Theorem 2. ■