Submitted to Professor Pedroni and Professor Kuttner

NONPARAMETRIC RANK TEST FOR PANELS USING BLOCK BOOTSTRAP

Gea Hyun Shin[1]

In this paper, we consider the use of block bootstrap to improve the performance of the MB rank test introduced by Pedroni, Vogelsang, Wagner, and Westerlund (2010). Asymptotic validity of the booststrap test is proved for T going to infinity. We report the size and power properties of the bootstrap test obtained through Monte Carlo simulations.

Keywords: nonparametric test, block bootstrap, unit root test, cross-sectional dependence.

1. INTRODUCTION

Early generations of panel unit root tests assumed independence across the cross-sectional dimension. However, the assumption of independence across the cross-sectional dimension has faded out in the literature because it induces size distortions, or increased rates of type I errors, in the panel unit root tests when there is a cross-sectional dependency in the data. Thus, the recent literature have focused on identifying the nature and magnitude of cross-sectional dependency in order to develop a proper unit root test accordingly. For instance, Bai and Ng (2004) devised a method called PANIC to identify the number of common stochastic trends and to extract them out of the data in order to create data without common factor type cross-sectional dependency. However, Palm, F., S. Smeekes and J. P. Urbain (2010) point out that such methods involving the specification and estimation of cross-dependency have a fundamental limitation, as they are unable to account for other types of cross-dependency that might be present in the

[1] I would like to thank Professor Pedroni, my thesis advisor, for his support and guidance at every step of the way. I would also like to thank Professor Gentry and Professor Kuttner for their encouragement and suggestions, and Susan Chen '12 for reviewing my draft.

data.

Two methods have been developed which do not require the estimation of cross-sectional dependency. The first method, proposed by Palm et al. (2010), uses a block bootstrap for unit root tests. The authors take a block of panel series and re-sample them to create pseudo-series with the same dependence structure of the original data, and apply modified test statistics of Levin, A., C. F. Lin, and C. S. J. Chu (2002) and Im, K. S., M. H. Pesaran, and Y. Shin (2003). Although their method can deal with a wider variety of cross-sectional dependency than the common-factor approaches or other bootstrap approaches (such as the one by Chang, Y., J. Park and K. Song (2004)), the method is still limited in that it cannot tell exactly how many units in the cross-sectional dimension are stationary. Furthermore, since the test statistics they used are not pivotal, bootstrapping cannot provide asymptotic refinement. The second method was proposed by Pedroni, P., T. Vogelsang, M. Wagner, and J. Westerlund (2010). Their tests, known as nonparametric rank tests, can deal with the most general form of cross-sectional dependency. They show in the paper that short run and long run variances of detrended series both asymptotically converge to distributions with the same nuisance parameter that describes the dependence structure. Hence, by taking the ratio of the two variances, they show that the limiting distribution of the ratio is nuisance-free. Thus, their tests do not require any a priori assumptions about the dependence structure of the series and can deal with the most general form of cross-sectional dependency.

This paper considers applying the block bootstrap method to the tests developed by Pedroni et al. (2010). Since the test statistics by Pedroni et al. (2010) are pivotal, applying the block bootstrap to the rank tests can offer asymptotic refinement. In other words, the distribution arising from the bootstrap test statistic gets closer to the true distribution of the original test statistic faster than the true distribution converging to the asymptotic distribution. At the same time, the block bootstrap rank tests can still account for all types of cross-sectional dependency that the original rank tests can. Thus, combining the block bootstrap method with the tests devised by Pedroni et al. (2010) results in desirable panel unit root tests.

The paper is structured in the following order: Section 2 presents the data generating process (DGP) and the assumptions underlying them. Section 3 presents the construction of the asymptotic test statistic and bootstrap test statistic. Section 4 provides the asymptotic properties of both the asymptotic and bootstrap test statistics. The small sample properties and the issue of block length selection are discussed in Section 5. Finally, Section 6 concludes the paper. Mathematical proofs of the validity of the bootstrap test statistics are presented in the Appendix.

2. ASSUMPTIONS

We follow the data generating process described in Pedroni et al. (2010) which allows for general cross-sectional dependency. We first note that boldfaced symbols represent vectors and matrices, and $\Rightarrow$ implies convergence in distribution as $T \to \infty$. Let $\boldsymbol{y}_t = [y_{1,t}, \ldots, y_{N,t}]'$ (t=1,...,T) be an N-dimensional vector generated by the following process:

$$\boldsymbol{y}_t = \boldsymbol{\alpha}_p \boldsymbol{F}_t^p + \boldsymbol{u}_t, \tag{2.1}$$

where $\boldsymbol{F}_t^p = [1, t, \cdots, t^p]'$, $p \geq -1$, and $\boldsymbol{F}_t^0 = 1$ with $\boldsymbol{\alpha}_p = [\boldsymbol{\alpha}_1', \cdots, \boldsymbol{\alpha}_N']'$ satisfying $\boldsymbol{\alpha}_i \in \mathbb{R}^{(p+1)}$. In words, $\boldsymbol{F}_t^p$ represents a polynomial trend function and $\boldsymbol{\alpha}_p$ is the associated matrix of trend coefficients. Note that $p = -1, 0, 1$ generate the most widely considered situations of $\boldsymbol{y}_t = \boldsymbol{u}_t$, $\boldsymbol{y}_t = \boldsymbol{\alpha}_0 + \boldsymbol{u}_t$ and $\boldsymbol{y}_t = \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 t + \boldsymbol{u}_t$, respectively.

As we have already discussed in the introduction, the rank tests of Pedroni et al. (2010) use the ratio of the properly normalized short run and long run variance of $\boldsymbol{u}_t$. They show that both variances weakly converge to distributions sharing the same nuisance parameter, namely the long run variance of $\Delta \boldsymbol{u}_t$. Note that any such argument requires some form of invariance principle, which requires mild assumptions to be made on the data generating process. The assumptions that we will impose on $\boldsymbol{u}_t$ are necessary both for the validity of the original test and bootstrap test. To state the assumptions and introduce our test statistic in concern, we need to describe $\boldsymbol{u}_t$ in further detail. We follow the steps from Pedroni et al. (2010) again. Suppose that N_1 units of $\boldsymbol{u}_t$ are either stationary or cross-sectionally cointegrated while $N_2 = N - N_1$ units contain independent unit roots. Now, consider an orthogonal matrix $\boldsymbol{A}$ whose first N_1 columns span the cointegrating space of $\boldsymbol{u}_t$. We can express $\boldsymbol{A}$ by its sub-matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ where $\boldsymbol{A}_1$ is composed of the first N_1 columns of $\boldsymbol{A}$ while $\boldsymbol{A}_2$ is composed of the remaining N_2 columns such that $\boldsymbol{A}_1'\boldsymbol{A}_2 = \boldsymbol{A}_2'\boldsymbol{A}_1 = 0$. That is, $\boldsymbol{A} = [\boldsymbol{A}_1, \boldsymbol{A}_2]$. Using this orthogonal matrix $\boldsymbol{A}$, we can extract the stationary portion of the N_1 units and separate them from the N_2 unit-root portion in a following way:

$$\boldsymbol{w}_t = \boldsymbol{A}' u_t = \begin{bmatrix} \boldsymbol{w}_{1,t} \\ \boldsymbol{w}_{2,t} \end{bmatrix}, \tag{2.2}$$

where $\boldsymbol{w}_{1,t} \in \mathbb{R}^{N_1}$ is stationary while $\boldsymbol{w}_{2,t} \in \mathbb{R}^{N_2}$ is unit-root non-stationary. Then, we can obtain a stationary vector by augmenting $\boldsymbol{w}_{1,t}$ with $\Delta \boldsymbol{w}_{2,t}$, the first difference of $\boldsymbol{w}_{2,t}$. Let this new stationary vector be $\boldsymbol{v}_t$. We are now ready to state the assumptions we need for the remainder of this paper.

Assumption 1 *$\boldsymbol{v_t} = \sum_{j=0}^{\infty} \boldsymbol{\psi}_j \boldsymbol{\epsilon}_{t-j}$, $\boldsymbol{\psi}_0 = I$, $\sum_{j=0}^{\infty} j|\boldsymbol{\psi}_j| < \infty$ where $\boldsymbol{\epsilon}_t$ is i.i.d. over t with mean zero and finite positive definite variance $\Sigma_{\boldsymbol{\epsilon}}$. In addition, $det(\sum_{j=0}^{\infty} \boldsymbol{\psi}_j s^j) \neq 0$ for all $\{s \in \mathbb{C} : |s| = 1\}$, $E|\boldsymbol{\epsilon}_t|^{2+\epsilon} < \infty$ for some $\epsilon > 0$, and $\boldsymbol{\psi}_i$ and ϵ_j are independent for all i and j.*

Given the assumption above, we can derive the following invariance principle which will be crucial for the convergence of our test statistic. Let $\boldsymbol{e}_t = \Delta \boldsymbol{w}_t$ and $\boldsymbol{\delta}_t = \Delta \boldsymbol{u}_t$ such that $\boldsymbol{e}_t = \boldsymbol{A}'\boldsymbol{\delta}_t$.

Lemma 1 (Invariance Principle) *Let $\boldsymbol{y}_t$ be generated by the data generating process described above and let Assumption 1 hold. Then, as $T \to \infty$,*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} \boldsymbol{e}_t \Rightarrow \boldsymbol{B}(r) = \Omega_{\boldsymbol{e}}^{1/2} \boldsymbol{W}(r),$$

where $\Omega_{\boldsymbol{e}}$ is the positive definite long-run variance matrix of $\boldsymbol{e}_t$ and $\boldsymbol{W}(r)$ is the standard N-dimensional Brownian motion on $r \in [0, 1]$.

We would like to note that Pedroni et al. (2010) assumes more general conditions under which the invariance principle holds. For the purpose of this paper, however, we will assume that $\boldsymbol{v}_t$ follows the linear process condition as the sufficient condition for the invariance principle. This is in order to simplify the proofs required to show the validity of bootstrap. Lastly, we impose an assumption on the growth of the block length as T grows. The following assumption is identical to the one imposed by Paparoditis, E. and D. Politis (2003).

Assumption 2 *As $T \to \infty$, $b \to \infty$ and $b = o(T^{1/2})$.*

The above assumption is crucial for proving the bootstrap version of the invariance principle. The over-differenced series degenerates because the rate of growth of T is faster than that of b, which is guaranteed by the above assumption. We will investigate this issue in detail in the Appendix.

3. BOOTSTRAP RANK TESTS

3.1. *The main test statistic*

The test we will consider tests the null of unit root non-stationarity in $N_2 > 0$ units against the alternative of stationarity across all units. To see how the test works intuitively, consider $\Omega_{\Delta u}$, the long run variance of $\Delta \boldsymbol{u}_t$. Let $r \equiv \text{rank}(\Omega_{\Delta u})$. Using the transformation matrix $\boldsymbol{A}$ we considered in the previous section, we can easily verify the following:

$$\Omega_{\Delta u} = \boldsymbol{A}\Omega_{\Delta w}\boldsymbol{A}' = \boldsymbol{A}_2\Omega_{v_2}\boldsymbol{A}_2',$$

where $\boldsymbol{v}_{2,t} = \Delta \boldsymbol{w}_{2,t}$. This implies that $r = \text{rank}(\Omega_{v_2}) = N_2$. Then, we can use the variance of $\boldsymbol{u}_t$ to construct a test statistic since the variance will capture the information on the presence of unit roots in $\boldsymbol{u}_t$. We now show how to construct the main test statistic of Pedroni et al. (2010), which converges in distribution to a nuisance free distribution under the null of $r = N_2$. The test statistic is called MB. We will stick to their naming scheme. For deeper intuitions on how the test works, refer to Pedroni et al. (2010). To begin with, we need to obtain the estimates of "residuals," $\hat{\boldsymbol{u}}_t$. We can do this using the OLS estimation of (2.1). So,

$$\hat{\boldsymbol{u}}_t = \boldsymbol{y}_t - \left(\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{y}_t(\boldsymbol{F}_t^p)' \left(\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{F}_t^p(\boldsymbol{F}_t^p)' \right)^{-1} \right) \boldsymbol{F}_t^p. \tag{3.1}$$

We now obtain estimated short-run and long-run variances. The short-run, or contemporaneous, variance can be estimated as

$$\hat{\Sigma}_{\hat{u}} = \frac{1}{T}\sum_{t=1}^{T} \hat{\boldsymbol{u}}_t\hat{\boldsymbol{u}}_t'.$$

As for the long-run variance estimate, we state the result from Kiefer, N. M. and T. J. Vogelsang (2000) in which they provide a simplified formula for the estimated long-run variance ($\hat{\Omega}$) using the untruncated Bartlett kernel $k(x) = 1 - \frac{|x|}{T}$.

$$\hat{\Omega}_{\hat{u}} = \frac{2}{T^2}\sum_{t=1}^{T} \hat{\boldsymbol{S}}_t\hat{\boldsymbol{S}}_t',$$

where $\hat{\boldsymbol{S}}_t = \sum_{j=1}^{t} \hat{\boldsymbol{u}}_j$. MB statistic is then given by

$$\text{MB} = \frac{1}{2T}\text{tr}(\hat{\Omega}_{\hat{u}}\hat{\Sigma}_{\hat{u}}^{-1}). \tag{3.2}$$

3.2. *Bootstrap algorithm*

We illustrate the construction of the bootstrap version of the MB statistic. The block bootstrap algorithm we employ is a panel extension of difference-based block bootstrap algorithm by Paparoditis, E. and D. Politis (2003). The algorithm is designed such that the bootstrapped series correctly mimics the behavior of the original series under the null hypothesis. Note that the asterisk mark ($*$) indicates bootstrap variables.

Block Bootstrap Algorithm:

1. Calculate the centered differences of $\boldsymbol{u}_t$[2].

$$\hat{\delta}_{i,t} = u_{i,t} - u_{i,t-1} - \frac{1}{T-1}\sum_{t=2}^{T}(u_{i,t} - u_{i,t-1}), \tag{3.3}$$

2. Choose a positive number b ($< T$) which will be our block length. Let $k = \lfloor (T-2)/b \rfloor + 1$ be our block number. Then, randomly sample $i_0, i_1, \ldots, i_{k-1}$ i.i.d. from uniform distribution on the set $\{1, 2, \ldots, T - b\}$.
3. Set $\boldsymbol{u}_1^* = \boldsymbol{u}_1$ and construct the rest of the bootstrap series $\{\boldsymbol{u}_t^*\}_{t=2}^T$ in the following way:

$$\boldsymbol{u}_t^* = \boldsymbol{u}_{t-1}^* + \boldsymbol{\delta}_t^* \text{ for all } t,$$

where we set $\boldsymbol{\delta}_t^* = \hat{\boldsymbol{\delta}}_{i_m+s}$ with $m = \lfloor (t-2)/b \rfloor$ and $s = t - mb - 1$.
4. Calculate the MB test statistic from the pseudo series $\{\boldsymbol{u}_t^*\}_{t=1}^T$.

$$\text{MB}^* = \frac{1}{2T}\text{tr}(\hat{\Omega}_{\hat{u}}^*(\hat{\Sigma}_{\hat{u}}^{-1})^*).$$

[2]Although our algorithm starts by assuming that we have a detrended series, we need to detrend $\boldsymbol{y}_t$ through (3.1) in practice. And then, in step 4, before we obtain MB^*, we set $\boldsymbol{y}_t^* = \boldsymbol{u}_t^*$ and detrend $\boldsymbol{y}_t^*$ to obtain the new $\boldsymbol{u}_t^*$, which we use to construct MB^*. For more discussions on detrending bootstrap unit root tests, refer to Smeekes, S. (2009). In the mathematical proofs, we assume that there is no deterministic component. However, for the Monte Carlo simulation, we apply these modified steps to account for the $p = 0$ case.

5. Repeat steps 2-4 for however many times one wants. Say that we repeated the steps B times. Then, we have $\{\mathrm{MB}_j^*\}_{j=1}^B$. Under the null of $r = N_2$, the empirical distribution from $\{\mathrm{MB}_j^*\}_{j=1}^B$ provides a consistent approximation to the true asymptotic distribution of the MB statistic. We choose the critical value c_α^* of the bootstrap test as the α-quantile of the ordered $\{\mathrm{MB}_j^*\}_{j=1}^B$. If the MB constructed in (3.2) is smaller than c_α^*, we reject the null of $r = N_2$.

Paparoditis, E. and D. Politis (2003) show that the residual-based block bootstrap (RBB), for which we need to estimate the coefficient in front of $u_{i,t-1}$ in step 1 instead of fixing it as 1, retains higher power than the differenced-based block bootstrap (DBB) in time series. However, Palm et al. (2008) show through simulation that the DBB exhibits smaller size distortion than its residual counterpart. Hence, we adopt the DBB method in this paper for two reasons. First, the DBB does not require the estimation of any coefficient, which is in line with the nonparametric characteristic of our main test statistic in concern. Second, reduction in size distortion is a bigger gain than the loss of power in panel unit root tests. Tests using panel data generally retain high power such that the loss of power using DBB method is acceptable. However, size distortion is often an important issue in panel time series. Thus, the DBB method is suited better for a panel unit root test.

4. ASYMPTOTIC PROPERTIES

In this section, we establish the asymptotic validity of the boostrap test. First, we introduce the asymptotic distribution of the original MB test statistic.

THEOREM 1 *Let Assumption 1 hold. Then, under the null hypothesis* H_0 :

$r = N_2 > 0$,

$$\frac{1}{2T}\mathrm{tr}(\hat{\Omega}_{\hat{u}}\hat{\Sigma}_{\hat{u}}^{-1}) \Rightarrow \mathrm{tr}\left(\int_0^1 \boldsymbol{Q}_2^p(s)\boldsymbol{Q}_2^p(s)'ds\left(\int_0^1 \boldsymbol{W}_2^p(r)\boldsymbol{W}_2^p(r)'dr\right)^{-1}\right),$$

where $\boldsymbol{Q}^p(s) = \int_0^s \boldsymbol{W}^p(r)dr$ *and*
$\boldsymbol{W}^p(s) = \boldsymbol{W}(s) - \int_0^1 \boldsymbol{W}(r)(\boldsymbol{F}_r^p)'dr\left(\int_0^1 \boldsymbol{F}_r^p(\boldsymbol{F}_r^p)'dr\right)^{-1}\boldsymbol{F}_s^p$. *Note that the subscript 2 refers to the fact that the distributions are resulted from* $\hat{\boldsymbol{w}}_{2t}$.

Critical values for the MB test are provided in Pedroni et al. (2010). The following shows that the invariance principle holds for the bootstrapped series.

LEMMA 2 (Invariance Principle) *Let Assumption 1 and 2 hold and let* $\boldsymbol{y}_t$ *be generated under* $H_0 : r = N_2 > 0$*. Then, as* $T \to \infty$*,*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT \rfloor} \boldsymbol{e}_t^* \Rightarrow \boldsymbol{B}(r) \text{ in probability.}$$

This lemma and the existence of moments are sufficient to show that the limiting distribution of the bootstrap rank test statistic corresponds to the limiting distribution of the original test statistic.

THEOREM 2 *Let Assumption 1 and 2 hold and let* $\boldsymbol{y}_t$ *be generated under* $H_0 : r = N_2 > 0$*. Then, as* $T \to \infty$*,*

$$MB^* \Rightarrow \mathrm{tr}\left(\int_0^1 \boldsymbol{Q}_2^p(s)\boldsymbol{Q}_2^p(s)'ds\left(\int_0^1 \boldsymbol{W}_2^p(r)\boldsymbol{W}_2^p(r)'dr\right)^{-1}\right) \text{in probability.}$$

This theorem verifies that the MB test statistic constructed using the bootstrapped series is a valid test statistic under the null hypothesis. We now explore the asymptotic properties under the alternative hypothesis. The below asymptotic distribution of TMB is presented in Pedroni et al.(2010)

THEOREM 3 *Let Assumption 1 hold and let $\boldsymbol{y}_t$ be generated under H_A : $r = 0$. Then, as $T \to \infty$,*

$$TMB \Rightarrow \operatorname{tr}\left(\int_0^1 \boldsymbol{B}_1^p(s)\boldsymbol{B}_1^p(s)' ds \Sigma_{v_1 v_1}^{-1}\right),$$

such that MB converges to 0 under $H_A : r = 0$.

In order to satisfy the sufficient conditions for the consistency of the bootstrap test, it is required to show that unlike the original test statistic, the bootstrap test statistic converges to a distribution under the alternative hypothesis. By the discussion in Section 5.2. of Paparoditis, E. and D. Politis (2003), however, the DBB method makes the bootstrap test statistic inconsistent under the alternative hypothesis. The failure of convergence originates from the fact that the invariance principle for the bootstrapped series constructed with the DBB method no longer holds under the alternative hypothesis. As the invariance principle does not hold under the alternative, $H_A : r = 0$, MB^* degenerates as $T \to \infty$. Then, as MB^* degenerates along with MB under the alternative hypothesis, the critical value obtained from the empirical distribution of $\{\mathrm{MB}_j^*\}_{j=1}^B$ could be lower than MB such that we would fail to reject the false null hypothesis. The inconsistency of the bootstrap test statistic under the alternative hypothesis and the resulting loss in power are the costs we must pay in order to reduce the size distortion as we have discussed before.

It is also important to consider other alternative hypotheses. That is, we need to consider the alternative hypotheses of $0 < r < N_2$. Under these alternative hypotheses, the MB statistic converges to the distribution described in Theorem 1 but with a smaller dimension which corresponds to the rank of the long run covariance matrix under the alternative hypotheses. Consequently, the test statistic does not diverge nor degenerate under the alternative hypotheses, indicating that the MB statistic is inconsistent under $H_A : 0 < r < N_2$. Thus, we do not investigate other alternative hypotheses.

We have established that the bootstrap test statistic is partially valid. We will now explore the small sample performance of the bootstrap test under the full rank null hypothesis and under the alternative hypothesis of stationarity across all units.

5. SMALL SAMPLE PERFORMANCE

We investigate the small sample properties of the bootstrap test through Monte Carlo simulations. Both the size and power properties are evaluated under the null hypothesis of unit root across all units against the alternative hypothesis of stationarity across all units. Various types of dynamic and cross-sectional dependency are considered.

5.1. *Monte Carlo setup*

In this section, we describe how our sample is simulated. It is imperative that the simulated samples exhibit various cross-sectional dependence structures. Unless otherwise indicated, the notations we use here are identical to the ones used in Section 2. Without loss of generality, we impose that the stationary units are ordered first. We follow the simulation design described in Pedroni et al. (2010) and modify it slightly to compare our results with both Palm et al. (2010) and Pedroni et al. (2010). Consider the following setup:

$$\boldsymbol{\alpha}_p = 0, \boldsymbol{A}_1 = \begin{bmatrix} \boldsymbol{I}_{N_1} \\ 0 \end{bmatrix}, \boldsymbol{A}_2 = \begin{bmatrix} 0 \\ \boldsymbol{I}_{N_2} \end{bmatrix}$$

$$\begin{bmatrix} \boldsymbol{v}_{1,t} \\ \boldsymbol{v}_{2,t} \end{bmatrix} = \begin{bmatrix} \rho \boldsymbol{I}_{N_1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_{1,t-1} \\ \boldsymbol{v}_{2,t-1} \end{bmatrix} + \boldsymbol{\gamma}_t,$$

where $|\rho| < 1$ and $\boldsymbol{\gamma}_t$ is a randomly generated N-dimensional error vector. Note that the definition of $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ ensures $\boldsymbol{y}_{1,t} = \boldsymbol{v}_{1,t}$ and $\Delta \boldsymbol{y}_{2,t} = \boldsymbol{v}_{2,t}$. The cross-sectional and dynamic dependence will be generated by the error vector $\boldsymbol{\gamma}_t$ in a following way:

$$\boldsymbol{\gamma}_t = \boldsymbol{\Theta}\boldsymbol{\gamma}_{t-1} + \boldsymbol{\epsilon}_t, \;\; \boldsymbol{\epsilon}_t \sim N(0, \boldsymbol{\Sigma})$$
$$\boldsymbol{\Theta} = \mathrm{diag}(\theta_1, \ldots, \theta_N), \;\; \theta_i \sim U(-0.3, 0.3) \text{ or } \theta_i = 0 \; \forall i,$$

where $\boldsymbol{\Sigma}$ is constructed as in Chang, Y. (2004) to ensure that $\boldsymbol{\Sigma}$ is symmetric positive definite. Note that each matrix is $N \times N$.

1. Generate a diagonal matrix $\boldsymbol{D}$ containing eigenvalues $\lambda_1, \ldots, \lambda_N$ as its diagonal elements with $\lambda_1 = r$, $\lambda_i \sim U(r, 1)$ for $i \notin \{1, N\}$, and $\lambda_N = 1$.
2. Generate a random matrix $\boldsymbol{U} \sim U(0, 1)$ and define $\boldsymbol{P} = \boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1/2}$.
3. Finally, let $\boldsymbol{\Sigma} = \boldsymbol{PDP}'$.

The choice of r depends on whether or now we want cross-sectional dependency in our sample. If $r = 1$, the matrix $\boldsymbol{D}$ becomes an identity matrix such that $\boldsymbol{\epsilon}_t$ is independent across the cross-sectional dimension while if $r = 0.1$, $\boldsymbol{\epsilon}_t$ is cross-sectionally dependent. Note that the choice of θ_i governs the serial correlation in our sample. $\boldsymbol{\gamma}_t$ becomes serially uncorrelated when $\theta_i = 0 \; \forall i$ while $\theta_i \sim U(-0.3, 0.3)$ makes $\boldsymbol{\gamma}_t$ become serially correlated. We consider two simple settings.

- Setting 1: Samples generated under the null hypothesis. That is, $N_1 = 0$, $N_2 = N$ with various combinations of dynamic and cross-sectional dependency.
- Setting 2: Samples generated under the alternative hypothesis. That is, $N_1 = N$, $N_2 = 0$ with various combinations of dynamic and cross-sectional dependency.

Since the sets of time and cross-sectional dimensions that were considered in Palm et al. (2010) and Pedroni et al. (2010) are different, we use two different sets of T and N to allow for the direct comparison with each paper. To compare our results with Palm et al. (2010), we consider $T = 25, 50, 100$ and $N = 5, 25$. Furthermore, to compare our results with Pedroni et al.

(2010), we consider $T = 100, 200$ and $N = 10$, and consider the DGP with constant ($p = 0$). As in Palm et al. (2010), the fixed block length b is chosen as $b = \lceil 1.75T^{1/3} \rceil$. The results are based on 1000 Monte Carlo simulations and 400 bootstrap replications with $\alpha = 0.05$. Note that we first generate 100 presample values for the errors of each unit with $\boldsymbol{\epsilon}_0 = \mathbf{0}$ which are fixed throughout the simulation. Also, note that the size and power values for the tests from Palm et al. (2010) and Pedroni et al. (2010) are taken from their respective papers.

5.2. *Simulation results*

The size results of the bootstrap test under the Setting 1 are summarized in Table I and the power results are reported in Table II. The following provides a summary of the information we have gathered from these two tables.

1. As expected with bootstrapping a pivotal statistic, the new bootstrap test exhibits good size. Regardless of the choice of T and of the existence of cross-sectional dependence, the size is close to 0.05 for a small N. But when N gets larger, the size gets smaller than 0.05, indicating that the new test has some under-size issues. However, the under-size problem is more desirable than the over-size problem present in many asymptotic tests that are considered in the simulation study of Palm et al. (2010). The good size property can be explained by the use of the DBB method as opposed to the RBB method and by the asymptotic refinement provided by the bootstrap.
2. Unlike the good size property, the new test retains relatively poor power. However, power of the test does improve and converge to 1 as N and T get larger and as ρ gets smaller[3]. It might seem surprising

[3]Recall that ρ determines the "degree" of stationarity within the stationary units. If ρ gets closer to 1, the test could fail to distinguish the stationary series from the unit

TABLE I

Size at $\alpha = 5\%$

T	N	MB^{boot}	τ_p	τ_g
No dependence				
25	5	0.040	0.024	0.020
50	5	0.046	0.031	0.024
100	5	0.054	0.032	0.032
100	25	0.001	0.009	0.014
Contemporaneous dependence				
25	5	0.054	0.026	0.022
50	5	0.062	0.033	0.028
100	5	0.050	0.040	0.031
100	25	0.002	0.013	0.015
T	N	MB^{boot}	MB	
Contemporaneous but no dynamic				
100	10	0.025	0.059	
200	10	0.027	0.055	
Contemporaneous and dynamic				
100	10	0.016	0.063	
200	10	0.036	0.053	

TABLE II

Power at $\alpha = 5\%$

$\rho = 0.9$				
T	N	MB^{boot}	τ_p	τ_g
No dependence				
25	5	0.148	0.507	0.354
50	5	0.419	0.757	0.810
100	5	0.747	0.929	0.974
100	25	0.990	1.000	1.000
Contemporaneous dependence				
25	5	0.142	0.508	0.357
50	5	0.371	0.630	0.648
100	5	0.767	0.943	0.985
100	25	0.992	1.000	1.000
$\rho = 0.95$				
T	N	MB^{boot}	MB	
Contemporaneous and dynamic				
100	10	0.648	0.836	
200	10	1.000	1.000	

that the new bootstrap test exhibits reasonable power when the test is asymptotically inconsistent under the alternative hypothesis. The test still works in practice since the invariance principle under the alternative hypothesis actually holds for $k \to \infty$, although it does not hold for $T \to \infty$, as explored by Paparoditis, E. and D. Politis (2003) for a time series case. They explain that the unit root tests based on the DBB method can still work in practice even though the power is a bit worse than that of the tests based on the RBB method.

3. The presence of serial or cross-sectional correlation does not seem to alter the size and power of the new bootstrap test.

root series as the rank of the covariance matrix $\Omega_{\Delta u}$ will get closer to full.

Table I suggests that the size property of the new bootstrap test is more desirable than the size property of τ_p and τ_{gm} introduced by Palm et al. (2010). In fact, the new test retains a good size property even when $T = 25$ for which the tests considered by Palm et al. (2010) show poor size. They comment that the poor size of their tests arises from the property of BIC criteria, which tend to select too many lags when $T = 25$. Unlike their tests, however, the rank test does not require any lag-selection process so that it is free of the size distortion when T is small. The asymptotic rank test by Pedroni et al. (2010) generally retains good size. Like most asymptotic tests, however, its size increases above 0.05 when N increases[4]. As mentioned above, size of the new bootstrap test actually decreases as N gets larger.

As discussed in Palm et al. (2010), the choice of block length b can have a big impact on the performance of any test based on block bootstrap method. This is due to the fact that the growth of b is essential for the bootstrap validity. If the block length is too small, the bootstrapped panel series would fail to capture the entire dependence structure present in the original panel series. On the other hand, if the block length is too big, the bootstrap test statistics might not mimic the true distribution of the original test statistic as there would not be enough variations within the bootstrap test statistics. Palm et al. (2010) proposes the Warp-Speed calibration method for choosing a block length. However, they remark that the block length chosen by the calibration method is not optimal with respect to size or power. Hence, we used the simple fixed block length, which depends on T, in our simulation study.

[4]Although the values are not reported in the tables here, the size of the MB test increases from 0.063 to 0.081 when N increases from 10 to 20 for $T = 100$ with the presence of both contemporaneous and dynamic dependence.

6. CONCLUSION

This paper considers the use of block bootstrap to improve the performance of the MB rank test developed by Pedroni et al. (2010). We have shown that the new bootstrap test based on the rank test is asymptotically valid under the null hypothesis. Furthermore, even if the test is inconsistent under the alternative hypothesis, the Monte Carlo simulation results support that the test actually retains reasonable power. The simulation results also support that the new bootstrap test has a better size property than the original rank test.

Like the original rank test, the bootstrap rank test can deal with various types of cross-sectional dependency. In fact, the size and power properties of the new test are robust to the existence of cross-sectional or dynamic dependence. Hence, the new test is ideal when the size distortion is a big concern due to the presence of cross-sectional dependency and large N. The lack of power of the new test can be overcome by either increasing N or T, and neither would have significant impact on the good size property of the new test.

APPENDIX A: MATHEMATICAL PROOFS

Proof of Lemma 1: By Assumption 1, it follows from the standard invariance principle result that (Phillips and Solo (1992) and Phillips and Moon (1999)),

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT\rfloor}\boldsymbol{\epsilon}_t \Rightarrow \Sigma_{\epsilon}^{1/2}\boldsymbol{W}(r).$$

We now decompose $\Delta\boldsymbol{w}_t$ into over-differenced and correctly-differenced parts. Recall that $\boldsymbol{\psi}_i = [\boldsymbol{\psi}_{1,i}, \boldsymbol{\psi}_{2,i}]'$ where $\boldsymbol{\psi}_{1,i}$ contains the coefficients for the first N_1 units while $\boldsymbol{\psi}_{2,i}$ contains the coefficients for the latter N_2 units. Note that

$\boldsymbol{\psi}_{1,i}$ and $\boldsymbol{\psi}_{2,i}$ satisfy Assumption 1.Let $\boldsymbol{\psi}_i^+ = [\boldsymbol{\psi}_{1,i}, 0]'$. Then, by definition,

$$\Delta \boldsymbol{w}_t = \sum_{j=0}^{\infty} \left(\boldsymbol{\psi}_j \boldsymbol{\epsilon}_{t-j} - \boldsymbol{\psi}_j^+ \boldsymbol{\epsilon}_{t-j-1} \right).$$

By Theorem 3.15 of Phillips and Solo (1992),

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} \left(\sum_{j=0}^{\infty} \boldsymbol{\psi}_j \boldsymbol{\epsilon}_{t-j} \right) \Rightarrow \boldsymbol{\psi}(1) \Sigma_\epsilon^{1/2} \boldsymbol{W}(r),$$

and

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} \left(\sum_{j=0}^{\infty} \boldsymbol{\psi}_j^+ \boldsymbol{\epsilon}_{t-j-1} \right) \Rightarrow \boldsymbol{\psi}^+(1) \Sigma_\epsilon^{1/2} \boldsymbol{W}(r),$$

such that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} \Delta \boldsymbol{w}_t \Rightarrow \left(\boldsymbol{\psi}(1) \Sigma_\epsilon^{1/2} - \boldsymbol{\psi}^+(1) \Sigma_\epsilon^{1/2} \right) \boldsymbol{W}(r).$$

Let $\widetilde{\boldsymbol{\psi}}_i = [0, \boldsymbol{\psi}_{2,i}]'$. Then,

$$\left(\boldsymbol{\psi}(1) \Sigma_\epsilon^{1/2} - \boldsymbol{\psi}^+(1) \Sigma_\epsilon^{1/2} \right) \boldsymbol{W}(r) = \widetilde{\boldsymbol{\psi}}(1) \Sigma_\epsilon^{1/2} \boldsymbol{W}(r).$$

The same result can be derived using $\widetilde{\boldsymbol{\psi}}_i$ and $\boldsymbol{\psi}_i^+$. First, note that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} \Delta \boldsymbol{\epsilon}_t = \frac{1}{\sqrt{T}} \left(\boldsymbol{\epsilon}_{\lfloor rT \rfloor} + \boldsymbol{\epsilon}_0 \right) = O_p\left(\frac{1}{\sqrt{T}} \right) \to 0.$$

Then, the result directly follows as below:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} \Delta \boldsymbol{w}_t = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} \left(\sum_{j=0}^{\infty} \widetilde{\boldsymbol{\psi}}_j \boldsymbol{\epsilon}_{t-j} + \sum_{j=0}^{\infty} \boldsymbol{\psi}_j^+ \Delta \boldsymbol{\epsilon}_{t-j} \right) \Rightarrow \widetilde{\boldsymbol{\psi}}(1) \Sigma_\epsilon^{1/2} \boldsymbol{W}(r).$$

We now need to verify that $\widetilde{\boldsymbol{\psi}}(1) \Sigma_\epsilon^{1/2} = \Omega_{\Delta w}^{1/2}$. Recall that by definition,

$$\Omega_{\Delta w} = \lim_{T \to \infty} \frac{1}{T} E \left[\left(\sum_{t=1}^{T} \Delta \boldsymbol{w}_t \right) \left(\sum_{t=1}^{T} \Delta \boldsymbol{w}_t \right)' \right].$$

Observe the following:

$$
\begin{aligned}
\Omega_{\Delta w} &= \lim_{T\to\infty} \frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} E[\Delta \boldsymbol{w}_s \Delta \boldsymbol{w}_t'] \\
&= \lim_{T\to\infty} \frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} E\left[\sum_{i=0}^{\infty} \left(\boldsymbol{\psi}_i \boldsymbol{\epsilon}_{s-i} - \boldsymbol{\psi}_i^{+} \boldsymbol{\epsilon}_{s-i-1} \right) \left(\sum_{j=0}^{\infty} \left(\boldsymbol{\psi}_j \boldsymbol{\epsilon}_{t-j} - \boldsymbol{\psi}_j^{+} \boldsymbol{\epsilon}_{t-j-1} \right) \right)' \right] \\
&= \lim_{T\to\infty} \frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} (\boldsymbol{\psi}_i E[\boldsymbol{\epsilon}_{s-i} \boldsymbol{\epsilon}_{t-j}'] \boldsymbol{\psi}_j' - \boldsymbol{\psi}_i^{+} E[\boldsymbol{\epsilon}_{s-i} \boldsymbol{\epsilon}_{t-j}'] \boldsymbol{\psi}_j' \\
&\quad - \boldsymbol{\psi}_i E[\boldsymbol{\epsilon}_{s-i} \boldsymbol{\epsilon}_{t-j}'] (\boldsymbol{\psi}_j^{+})' + \boldsymbol{\psi}_i^{+} E[\boldsymbol{\epsilon}_{s-i} \boldsymbol{\epsilon}_{t-j}'] \boldsymbol{\psi}_j') \\
&= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \left(\boldsymbol{\psi}_i \Sigma_\epsilon \boldsymbol{\psi}_j' - \boldsymbol{\psi}_i^{+} \Sigma_\epsilon \boldsymbol{\psi}_j' - \boldsymbol{\psi}_i \Sigma_\epsilon (\boldsymbol{\psi}_j^{+})' + \boldsymbol{\psi}_i^{+} \Sigma_\epsilon (\boldsymbol{\psi}_j^{+})' \right) \\
&= \boldsymbol{\psi}(1) \Sigma_\epsilon \boldsymbol{\psi}(1)' - \boldsymbol{\psi}^{+}(1) \Sigma_\epsilon \boldsymbol{\psi}(1)' - \boldsymbol{\psi}(1) \Sigma_\epsilon (\boldsymbol{\psi}(1)^{+})' \\
&\quad + \boldsymbol{\psi}^{+}(1) \Sigma_\epsilon (\boldsymbol{\psi}(1)^{+})' \\
&= \left(\boldsymbol{\psi}(1) - \boldsymbol{\psi}^{+}(1) \right) \Sigma_\epsilon \left(\boldsymbol{\psi}(1) - \boldsymbol{\psi}^{+}(1) \right)',
\end{aligned}
$$

which verifies that $\Omega_{\Delta w}^{1/2} = \widetilde{\boldsymbol{\psi}}(1) \Sigma_\epsilon^{1/2}$. *Q.E.D.*

Similarly, we can show that $\Sigma_{\Delta w} = \sum_{j=0}^{\infty} (\boldsymbol{\psi}_j \Sigma_\epsilon \boldsymbol{\psi}_j' - \boldsymbol{\psi}_{j+1}^{+} \Sigma_\epsilon \boldsymbol{\psi}_j' - \boldsymbol{\psi}_j \Sigma_\epsilon \times (\boldsymbol{\psi}_{j+1}^{+})' + \boldsymbol{\psi}_j^{+} \Sigma_\epsilon (\boldsymbol{\psi}_j^{+})')$. In order to prove Theorem 2, we first need to prove Lemma 2. The following lemmas, which are the modified versions of the lemmas from Palm et al. (2010), build upon each other to establish the proof of Lemma 2.

Lemma A.1 *Let Assumptions 1 and 2 hold. Then, if*

$$
\frac{1}{\sqrt{T}} \sum_{m=0}^{\lfloor (k-1) r \rfloor} \sum_{s=1}^{b} (\boldsymbol{e}_{i_m+s} - E^*[\boldsymbol{e}_{i_m+s}]) \Rightarrow \boldsymbol{B}(r) \text{ in probability,}
$$

then,

$$
\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} \boldsymbol{e}_t^* \Rightarrow \boldsymbol{B}(r) \text{ in probability.}
$$

PROOF OF LEMMA A.1: For simplicity, let $\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT\rfloor}\boldsymbol{e}_t^*$ be $S_T^*(r)$. We first obtain $\boldsymbol{e}_t^*$ from $\boldsymbol{\delta}_t^*$. Recall the definition of $\boldsymbol{\delta}_t^*$.

$$\boldsymbol{\delta}_t^* = \hat{\boldsymbol{\delta}}_{i_m+s}.$$

Applying the orthogonal matrix $\boldsymbol{A}'$ to $\hat{\boldsymbol{\delta}}_{i_m+s}$, we get,

$$\boldsymbol{e}_t^* = \boldsymbol{A}'\hat{\boldsymbol{\delta}}_{i_m+s} = \hat{\boldsymbol{e}}_{i_m+s}.$$

Then, we can express $S_T^*(r)$ in terms of $\hat{\boldsymbol{e}}_{i_m+s}$.

$$S_T^*(r) = \frac{1}{\sqrt{T}}\boldsymbol{w}_1 + \frac{1}{\sqrt{T}}\sum_{m=0}^{M_r}\sum_{s=1}^{B}\hat{\boldsymbol{e}}_{i_m+s},$$

where $M_r = \lfloor(\lfloor rT\rfloor - 2)/b\rfloor\rfloor$ and $B = \min\{b, \lfloor rT\rfloor - mb - 1\}$. Note that the way M_r and B are determined follows directly from the bootstrap algorithm. We can simplify the above summation as the following:

$$S_T^*(r) = O_p(\frac{1}{\sqrt{T}}) + \frac{1}{\sqrt{T}}\sum_{m=0}^{M_r}\sum_{s=1}^{b}\hat{\boldsymbol{e}}_{i_m+s} - \frac{1}{\sqrt{T}}\sum_{s=B+1}^{b}\hat{\boldsymbol{e}}_{i_{M_r}+s}. \tag{A.1}$$

Paparoditis, E. and D. Politis (2003) show the following:

$$\sup_{0\le r\le 1}\left|\frac{1}{\sqrt{T}}\sum_{s=B+1}^{b}\hat{e}_{j,i_{M_r}+s}\right| = O_p(\frac{1}{\sqrt{k}}),$$

for each individual j. Then, in order to investigate the asymptotic behavior of $S_T^*(r)$, our only concern is the second term in the equation (A.1). Using the definitions of $\hat{\boldsymbol{\delta}}_t$ in (3.3) and $\boldsymbol{e}_t$, we can rewrite the second term into a more meaningful expression:

$$\frac{1}{\sqrt{T}}\sum_{m=0}^{M_r}\sum_{s=1}^{b}\hat{\boldsymbol{e}}_{i_m+s} = \frac{1}{\sqrt{T}}\sum_{m=0}^{M_r}\sum_{s=1}^{b}\left(\boldsymbol{e}_{i_m+s} - \frac{1}{T-1}\sum_{h=2}^{T}\boldsymbol{e}_h\right).$$

Observe that $\frac{1}{T-1}\sum_{h=2}^{T}\boldsymbol{e}_h$ is a consistent estimator of the mean of the difference of $\boldsymbol{w}_t$. Then, since $\boldsymbol{e}_{i_m+s}$'s are randomly drawn from the centered differences of $\boldsymbol{w}_t$, the estimator should converge, uniformly in r, to the expectation of $\boldsymbol{e}_{i_m+s}$ as $T \to \infty$. That is,

$$\frac{1}{\sqrt{T}} \sum_{m=0}^{M_r} \sum_{s=1}^{b} \left(E^*\left[\boldsymbol{e}_{i_m+s}\right] - \frac{1}{T-1} \sum_{h=2}^{T} \boldsymbol{e}_h \right) \to 0.$$

Finally, note that using $\lfloor (k-1)r \rfloor$ where $k = \lfloor (T-2)/b \rfloor + 1$ as the upper limit of the summation is asymptotically equivalent to using M_r as the upper limit of the summation. Hence, if

$$\frac{1}{\sqrt{T}} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} (\boldsymbol{e}_{i_m+s} - E^*[\boldsymbol{e}_{i_m+s}]) \Rightarrow \boldsymbol{B}(r) \text{ in probability,}$$

then,

$$\begin{aligned} S_T^*(r) &= O_p(\frac{1}{\sqrt{T}}) + \frac{1}{\sqrt{T}} \sum_{m=0}^{M_r} \sum_{s=1}^{b} \left(\boldsymbol{e}_{i_m+s} - \frac{1}{T-1} \sum_{h=2}^{T} \boldsymbol{e}_h \right) + O_p(\frac{1}{\sqrt{k}}) \\ &\Rightarrow \boldsymbol{B}(r) \text{ in probability.} \end{aligned}$$

Q.E.D.

The above Lemma A.1 implies that it is sufficient to investigate the asymptotic behavior of $\frac{1}{\sqrt{T}} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} (\boldsymbol{e}_{i_m+s} - E^*[\boldsymbol{e}_{i_m+s}])$ instead of $S_T^*(r)$ itself.

LEMMA A.2 *Let* $H_m^* = \frac{1}{\sqrt{b}} \sum_{s=1}^{b} (\boldsymbol{\epsilon}_{i_m+s} - E^*[\boldsymbol{\epsilon}_{i_m+s}])$ *and let Assumptions 1 and 2 hold. Then,*

1. $E^*[H_m^*] = 0$,
2. $E^*[H_m^* H_m^{*\prime}] = \Sigma_\epsilon + o_p(1)$,
3. *As* $T \to \infty$,

$$\frac{1}{\sqrt{T}} \sum_{k=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} (\boldsymbol{\epsilon}_{i_m+s} - E^*[\boldsymbol{\epsilon}_{i_m+s}]) \Rightarrow \Sigma_\epsilon^{1/2} W(r) \text{ in probability.}$$

PROOF OF LEMMA A.2: Refer to the proofs of Lemma A.3 and Lemma A.4 of Palm et al. (2008). *Q.E.D.*

LEMMA A.3 *Let Assumptions 1 and 2 hold and let* $\boldsymbol{y}_t$ *be generated under* $H_0 : r = N_2 > 0$*. Then, as* $T \to \infty$,

$$\frac{1}{\sqrt{T}} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} (\boldsymbol{e}_{i_m+s} - E^*[\boldsymbol{e}_{i_m+s}]) \Rightarrow \widetilde{\boldsymbol{\psi}}(1) \Sigma_\epsilon^{1/2} W(r) \text{ in probability.}$$

PROOF OF LEMMA A.3: We begin by decomposing $\boldsymbol{e}_t$ into $\boldsymbol{e}_t^I$ and $\boldsymbol{e}_t^{II}$. Let $\boldsymbol{e}_t^I = [\boldsymbol{e}_{1,t}, 0]'$ and $\boldsymbol{e}_t^{II} = [0, \boldsymbol{e}_{2,t}]'$ where $\boldsymbol{e}_{1,t} \in \mathbb{R}^{N_1}$ and $\boldsymbol{e}_{2,t} \in \mathbb{R}^{N_2}$. Note that each vector is still N-dimensional. Then,

$$\begin{aligned}
&\frac{1}{\sqrt{T}} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} (\boldsymbol{e}_{i_m+s} - E^*[\boldsymbol{e}_{i_m+s}]) \\
&= \frac{1}{\sqrt{T}} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} \left(\boldsymbol{e}^I_{i_m+s} - E^*[\boldsymbol{e}^I_{i_m+s}]\right) \\
&\qquad + \frac{1}{\sqrt{T}} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} \left(\boldsymbol{e}^{II}_{i_m+s} - E^*[\boldsymbol{e}^{II}_{i_m+s}]\right) = C_T^* + D_T^*.
\end{aligned}$$

We analyze C_T^* and D_T^* separately. The convergence of D_T^* to $\boldsymbol{B}(r)$ follows from the proof of Lemma A.5. of Palm et al. (2008). For the ease of the verification, we present the adapted version of the proof. We first decompose D_T^* using Beveridge-Nelson decomposition (BN) which is possible under Assumption 1.

$$\begin{aligned}
&\frac{1}{\sqrt{T}} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} \left(\boldsymbol{e}^{II}_{i_m+s} - E^*[\boldsymbol{e}^{II}_{i_m+s}]\right) \\
&= \frac{1}{\sqrt{T}} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} \widetilde{\boldsymbol{\psi}}(1) \left(\boldsymbol{\epsilon}_{i_m+s} - E^*[\boldsymbol{\epsilon}_{i_m+s}]\right) \\
&- \frac{1}{\sqrt{T}} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \left((\overline{\boldsymbol{\psi}}(L)\boldsymbol{\epsilon}_{i_m+b} - E^*[\overline{\boldsymbol{\psi}}(L)\boldsymbol{\epsilon}_{i_m+b}])(\overline{\boldsymbol{\psi}}(L)\boldsymbol{\epsilon}_{i_m} - E^*[\overline{\boldsymbol{\psi}}(L)\boldsymbol{\epsilon}_{i_m}]) \right),
\end{aligned}$$

where $\overline{\boldsymbol{\psi}}_j = \sum_{h=j+1}^{\infty} \widetilde{\boldsymbol{\psi}}_h$. We first show that the second term is $o_p^*(1)$. By the Markov inequality, for any given $\epsilon > 0$,

$$P^* \left[\left| \frac{1}{\sqrt{T}} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \left(\overline{\boldsymbol{\psi}}(L)\boldsymbol{\epsilon}_{i_m+\eta} - E^*[\overline{\boldsymbol{\psi}}(L)\boldsymbol{\epsilon}_{i_m+\eta}] \right) \right| > \epsilon \right] \\ \leq \frac{1}{\epsilon^2} E^* \left| \frac{1}{\sqrt{T}} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \left(\overline{\boldsymbol{\psi}}(L)\boldsymbol{\epsilon}_{i_m+\eta} - E^*[\overline{\boldsymbol{\psi}}(L)\boldsymbol{\epsilon}_{i_m+\eta}] \right) \right|^2,$$

for $\eta = 0, b$. Using the fact that the blocks are independent, we get

$$E^* \left| \frac{1}{\sqrt{T}} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \left(\overline{\boldsymbol{\psi}}(L)\boldsymbol{\epsilon}_{i_m+\eta} - E^*[\overline{\boldsymbol{\psi}}(L)\boldsymbol{\epsilon}_{i_m+\eta}] \right) \right|^2 \\ = \frac{1}{T} \sum_{m_1=0}^{\lfloor (k-1)r \rfloor} \sum_{m_2=0}^{\lfloor (k-1)r \rfloor} E^* \left(\sum_{j=0}^{\infty} \overline{\boldsymbol{\psi}}_j (\boldsymbol{\epsilon}_{i_{m_1}+\eta-j} - E^*[\boldsymbol{\epsilon}_{i_{m_1}+\eta-j}]) \right)' \\ \times \left(\sum_{j=0}^{\infty} \overline{\boldsymbol{\psi}}_j (\boldsymbol{\epsilon}_{i_{m_2}+\eta-j} - E^*[\boldsymbol{\epsilon}_{i_{m_2}+\eta-j}]) \right) \\ = \frac{1}{T} \sum_{m=0}^{\lfloor (k-1)r \rfloor} E^* \left| \sum_{j=0}^{\infty} \overline{\boldsymbol{\psi}}_j (\boldsymbol{\epsilon}_{i_m+\eta-j} - E^*[\boldsymbol{\epsilon}_{i_m+\eta-j}]) \right|^2.$$

By Minkowski's inequality,

$$\frac{1}{T} \sum_{m=0}^{\lfloor (k-1)r \rfloor} E^* \left| \sum_{j=0}^{\infty} \overline{\boldsymbol{\psi}}_j (\boldsymbol{\epsilon}_{i_m+\eta-j} - E^*[\boldsymbol{\epsilon}_{i_m+\eta-j}]) \right|^2 \\ \leq \frac{1}{T} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \left(\sum_{j=0}^{\infty} \left| \overline{\boldsymbol{\psi}}_j \right| (E^* \left[\left| \boldsymbol{\epsilon}_{i_m+\eta-j} - E^*[\boldsymbol{\epsilon}_{i_m+\eta-j}] \right|^2 \right])^{1/2} \right)^2 \\ \leq \frac{4k}{T} \left(\sum_{j=0}^{\infty} \left| \overline{\boldsymbol{\psi}}_j \right| \right)^2 \max_j \frac{1}{T-b} \sum_{t=1}^{T-b} |\boldsymbol{\epsilon}_{t+\eta-j}|^2,$$

which holds uniformly in r and for $\eta = 0, b$. Note that the second inequality was derived directly by taking the maximum of the average of $\boldsymbol{\epsilon}$ with a change of variable. By Assumption 1,

$$\sum_{j=0}^{\infty} \left|\overline{\boldsymbol{\psi}}_j\right| < \infty,$$

$$\frac{1}{T-b}\sum_{t=1}^{T-b} |\boldsymbol{\epsilon}_{t+\eta-j}| = O_p(1),$$

where the first inequality holds as $\sum_{j=0}^{\infty} j \left|\widetilde{\boldsymbol{\psi}}_j\right| < \infty$ which is shown in Phillips and Solo (1992). Thus, for $\eta = 0, b$,

$$\frac{1}{\epsilon^2}E^*\left|\frac{1}{\sqrt{T}}\sum_{m=0}^{\lfloor (k-1)r\rfloor}\left(\overline{\boldsymbol{\psi}}(L)\boldsymbol{\epsilon}_{i_m+\eta} - E^*[\overline{\boldsymbol{\psi}}(L)\boldsymbol{\epsilon}_{i_m+\eta}]\right)\right|^2 = O_p(b^{-1}),$$

which holds uniformly in r. Hence,

$$\frac{1}{\sqrt{T}}\sum_{m=0}^{\lfloor (k-1)r\rfloor}\left((\overline{\boldsymbol{\psi}}(L)\boldsymbol{\epsilon}_{i_m+b} - E^*[\overline{\boldsymbol{\psi}}(L)\boldsymbol{\epsilon}_{i_m+b}])(\overline{\boldsymbol{\psi}}(L)\boldsymbol{\epsilon}_{i_m} - E^*[\overline{\boldsymbol{\psi}}(L)\boldsymbol{\epsilon}_{i_m}])\right)$$
$$= o_p^*(1).$$

Then, by the third result of Lemma A.2,

$$D_T^* \Rightarrow \widetilde{\boldsymbol{\psi}}(1)\Sigma_{\epsilon}^{1/2}W(r) \text{ in probability.}$$

We conclude our proof by showing that C_t^* degenerates to 0 as $T \to \infty$. Note first that we can rewrite $\boldsymbol{e}_t^I$ as $\Delta\boldsymbol{w}_{1,t}$. Then,

$$\begin{aligned}
C_T^* &= \frac{1}{\sqrt{T}}\sum_{m=0}^{\lfloor (k-1)r\rfloor}\sum_{s=1}^{b}(\Delta\boldsymbol{w}_{1,i_m+s} - E^*[\Delta\boldsymbol{w}_{1,i_m+s}]) \\
&= \frac{1}{\sqrt{T}}\sum_{m=0}^{\lfloor (k-1)r\rfloor}(\boldsymbol{w}_{1,i_m+b} - \boldsymbol{w}_{1,i_m} - E^*[\boldsymbol{w}_{1,i_m+b} - \boldsymbol{w}_{1,i_m}]) \\
&= \frac{1}{\sqrt{T}}\sum_{m=0}^{\lfloor (k-1)r\rfloor}(\boldsymbol{w}_{1,i_m+b} - E^*[\boldsymbol{w}_{1,i_m+b}]) \\
&\quad - \frac{1}{\sqrt{T}}\sum_{m=0}^{\lfloor (k-1)r\rfloor}(\boldsymbol{w}_{1,i_m} - E^*[\boldsymbol{w}_{1,i_m}]) \\
&= C_{T,b}^* + C_{T,0}^*.
\end{aligned}$$

By the definition of $\boldsymbol{w}_{1,t}$ and by the above analysis, for $\eta = 0, b$,

$$C^*_{T,\eta} = \frac{1}{\sqrt{T}} \sum_{m=0}^{\lfloor (k-1)r \rfloor} (\boldsymbol{\psi}^+(L)\boldsymbol{\epsilon}_{i_m+\eta} - E^*[\boldsymbol{\psi}^+(L)\boldsymbol{\epsilon}_{i_m+\eta}]).$$

By the proof of Lemma 8.5 of Paparoditis, E. and D. Politis (2003), we have

$$\frac{\sqrt{T}}{\sqrt{k}} C^*_{T,\eta} = O^*_p(1) \text{ as } k \to \infty,$$

indicating that the left-hand side is bounded in the limit as $k \to \infty$ in probability[5]. By the definition of b, the above implies that uniformly in r,

$$C^*_{T,\eta} = O^*_p(\frac{1}{\sqrt{b}}) \text{ as } b \to \infty,$$

which holds for $\eta = 0, b$. Then, by Assumption 2, for $\eta = 0, b$

$$C^*_{T,\eta} \Rightarrow 0 \text{ in probability.}$$

Hence,

$$\frac{1}{\sqrt{T}} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} (\boldsymbol{e}_{i_m+s} - E^*[\boldsymbol{e}_{i_m+s}]) \Rightarrow \widetilde{\boldsymbol{\psi}}(1)\Sigma_\epsilon^{1/2} W(r) \text{ in probability.}$$

Q.E.D.

The above result implies that increasing block length as T grows larger is crucial in order for $C^*_{T,\eta}$ to degenerate. As Palm et al. (2008) discusses the issue in detail, in finite samples, we expect $C^*_{T,\eta}$ term, which represents the stationary portion, to affect the covariance matrix of the Brownian motion in the limiting distribution. Thus, the test will perform well only if we have moderately large sample size and the block length.

[5] In fact, by Lemma 8.5 of Paparoditis, E. and D. Politis (2003), $\frac{\sqrt{T}}{\sqrt{k}} C^*_T$ weakly converges to a Brownian motion involving the variance of $\boldsymbol{w}_{1,t}$ as $k \to \infty$.

Proof of Lemma 2: Let Assumptions 1 and 2 hold. By Lemma A.3 and by the proof of Lemma 1,

$$\frac{1}{\sqrt{T}}\sum_{m=0}^{\lfloor (k-1)r\rfloor}\sum_{s=1}^{b}(\boldsymbol{e}_{i_m+s}-E^*[\boldsymbol{e}_{i_m+s}])\Rightarrow \boldsymbol{B}(r) \text{ in probability.}$$

Then, by Lemma A.1,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT\rfloor}\boldsymbol{e}_t^*\Rightarrow \boldsymbol{B}(r) \text{ in probability.}$$

Q.E.D.

We have verified that $\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT\rfloor}\boldsymbol{e}_t^*$ converges to a Brownian motion involving the long run variance matrix of $\boldsymbol{e}_t=\Delta\boldsymbol{w}_t$. However, we have not verified that the variances constructed by $\boldsymbol{e}_t^*$ agree with the original variances. The next lemma shows that variances of $\boldsymbol{e}_t^*$ correctly mimic the original variances.

Lemma A.4 *Let Assumption 1 and 2 hold and let $\boldsymbol{y}_t$ be generated under $H_0: r=N_2>0$. Then, as $T\to\infty$,*

1. $\Omega^*_{\Delta w}=\Omega_{\Delta w}+o_p(1)$,
2. $\Sigma^*_{\Delta w}=\Sigma_{\Delta w}+o_p(1)$.

Proof of Lemma A.4: For notational simplicity, we again use $\boldsymbol{e}_t$ instead of $\Delta\boldsymbol{w}_t$. By definition,

$$\Omega_e^*=\frac{1}{T}\left(E^*\left[\left(\sum_{t=1}^{T}\boldsymbol{e}_t^*\right)\left(\sum_{t=1}^{T}\boldsymbol{e}_t^*\right)'\right]-E^*\left[\sum_{t=1}^{T}\boldsymbol{e}_t^*\right]E^*\left[\sum_{t=1}^{T}\boldsymbol{e}_t^*\right]'\right).$$

We would like to analyze $\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{e}_t^*$ first. By the proof of Lemma A.1,

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{e}_t^*=\frac{1}{T}\sum_{m=0}^{k-1}\sum_{s=1}^{b}(\boldsymbol{e}_{i_m+s}-E^*[\boldsymbol{e}_{i_m+s}])+o_p^*(1).$$

Thus, when we put the expected value sign in front of it, the first term disappears and only the second little o-notation in probability stays. Hence,

$$\Omega_e^* = \frac{1}{T}E^*\left[\left(\sum_{m=0}^{k-1}\sum_{s=1}^{b}(\boldsymbol{e}_{i_m+s} - E^*[\boldsymbol{e}_{i_m+s}])\right)\right.$$
$$\left.\times\left(\sum_{m=0}^{k-1}\sum_{s=1}^{b}(\boldsymbol{e}_{i_m+s} - E^*[\boldsymbol{e}_{i_m+s}])\right)'\right] + o_p(1).$$

We now analyze the term $\sum_{m=0}^{k-1}\sum_{s=1}^{b}(\boldsymbol{e}_{i_m+s} - E^*[\boldsymbol{e}_{i_m+s}])$. Using the Beveridge-Nelson decomposition and following the argument used in Lemma A.3, we arrive at the equality (the terms that are similar to $C_{T,\eta}^*$ can be grouped together in the little o-notation),

$$\frac{1}{\sqrt{T}}\sum_{m=0}^{k-1}\sum_{s=1}^{b}(\boldsymbol{e}_{i_m+s} - E^*[\boldsymbol{e}_{i_m+s}])$$
$$= \frac{1}{\sqrt{T}}\widetilde{\boldsymbol{\psi}}(1)\sum_{m=0}^{k-1}\sum_{s=1}^{b}(\boldsymbol{\epsilon}_{i_m+s} - E^*[\boldsymbol{\epsilon}_{i_m+s}]) + o_p^*(1).$$

Plugging the right-hand side into the long-run variance formula, we get

$$\Omega_e^* = \frac{1}{T}\widetilde{\boldsymbol{\psi}}(1)E^*\left[\left(\sum_{m=0}^{k-1}\sum_{s=1}^{b}(\boldsymbol{\epsilon}_{i_m+s} - E^*[\boldsymbol{\epsilon}_{i_m+s}])\right)\right.$$
$$\left.\times\left(\sum_{m=0}^{k-1}\sum_{s=1}^{b}(\boldsymbol{\epsilon}_{i_m+s} - E^*[\boldsymbol{\epsilon}_{i_m+s}])\right)'\right]\widetilde{\boldsymbol{\psi}}(1)' + o_p(1).$$

By the second result of Lemma A.2 and using the fact that the blocks are independent (such that the product of terms from different blocks disappears in expectation),

$$\frac{1}{T}\widetilde{\boldsymbol{\psi}}(1)E^*\left[\left(\sum_{m=0}^{k-1}\sum_{s=1}^{b}(\boldsymbol{\epsilon}_{i_m+s} - E^*[\boldsymbol{\epsilon}_{i_m+s}])\right)\right.$$
$$\left.\times\left(\sum_{m=0}^{k-1}\sum_{s=1}^{b}(\boldsymbol{\epsilon}_{i_m+s} - E^*[\boldsymbol{\epsilon}_{i_m+s}])\right)'\right]\widetilde{\boldsymbol{\psi}}(1)' + o_p(1)$$
$$= \widetilde{\boldsymbol{\psi}}(1)\Sigma_\epsilon\widetilde{\boldsymbol{\psi}}(1)' + o_p(1).$$

Thus, we confirm that $\Omega_e^* = \Omega_e + o_p(1)$.

Next, we want to confirm the equivalence of the contemporaneous variances. By definition,

$$\Sigma_e^* = \frac{1}{T}\sum_{t=1}^{T}\left(E^*\left[\boldsymbol{e}_t^*(\boldsymbol{e}_t^*)'\right] - E^*\left[\boldsymbol{e}_t^*\right]E^*\left[\boldsymbol{e}_t^*\right]'\right).$$

By the proof of A.1, similar to the above analysis,

$$\Sigma_e^* = \frac{1}{T}\sum_{m=0}^{k-1}\sum_{s=1}^{b} E^*\left[(\boldsymbol{e}_{i_m+s} - E^*[\boldsymbol{e}_{i_m+s}])(\boldsymbol{e}_{i_m+s} - E^*[\boldsymbol{e}_{i_m+s}])'\right] + o_p(1).$$

As in the proof of Lemma A.3, let $\boldsymbol{e}_t = \boldsymbol{e}_t^I + \boldsymbol{e}_t^{II}$. Once we expand the above expression using this decomposition and apply the proof of Lemma A.6 of Palm et al. (2008), we finally confirm that,

$$\Sigma_e^* = \Sigma_e + o_p(1).$$

Q.E.D.

We have shown that the bootstrap invariance principle holds and that the bootstrapped moments are equivalent to the asymptotic moments. We can then easily prove the convergence in probability of the bootstrap MB statistic to the asymptotic distribution introduced in Theorem 1.

PROOF OF THEOREM 2: The proof of the theorem is identical to the proof of the convergence of MB statistic provided by Pedroni et al. (2010). Lemma 2 and Lemma A.4 provide the sufficient conditions for their proof.

LEMMA A.5 *Let Assumption 1 and 2 hold and let $\boldsymbol{y}_t$ be generated under $H_A : r = 0$. Then, as $T \to \infty$,*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT \rfloor} \boldsymbol{e}_t^* \Rightarrow 0 \text{ in probability.}$$

Proof of Lemma A.5: The proof is identical to the analysis of C_T^* term in the proof of Lemma A.3. The only difference is that we use $\boldsymbol{\epsilon}_t$ instead of $\boldsymbol{\epsilon}_t^I$, and use $\boldsymbol{\psi}(L)$ instead of $\boldsymbol{\psi}^+(L)$. *Q.E.D.*

The above lemma concludes that the bootstrapped MB test statistic under the alternative degenerates.

Q.E.D.

REFERENCES

Andrews, D. (2004): "The Block-Block Bootstrap: Improved Asymptotic Refinements," *Econometrica*, 72(3), 673-700.

Bai, J. and S. Ng (2004): "A Panic Attack on Unit Roots and Cointegration," *Econometrica*, 72(4), 1127-1177.

Chang, Y. (2004): "Bootstrap Unit Root Tests in Panels with Cross-Sectional Dependency," *Journal of Econometrics*, 120, 263-293.

Chang, Y., J. Park and K. Song (2006): "Bootstrapping Cointegrating Regressions," *Journal of Econometrics*, 133, 703-739.

Davidson, R. and N. MacKinnon (2006): "The Power of Bootstrap and Asymptotic Tests," *Journal of Econometrics*, 133, 421-441.

Gengenbach, C., J.P. Urbain and F. Palm (2004): "Panel Unit Root Tests in the Presence of Cross-Sectional Dependencies: Comparison and Implications for Modelling," *METEOR*, Research Memoranda 040.

Herwartz, H. and M. Neumann (2005): "Bootstrap Inference in Systems of Single Equation Error Correction Models," *Journal of Econometrics*, 128, 165-193.

Im, K. S., M. H. Pesaran, and Y. Shin (2003): "Testing for Unit Roots in Heterogeneous Panels," *Journal of Econometrics*, 115, 53-74.

Kiefer, N. M., and T. J. Vogelsang (2000): "Heteroskedasticity-Autocorrelation Robust Standard Errors Using the Bartlett Kernel without Truncation," *Econometrica*, 70, 2093-2095.

Levin, A., C.F. Lin, and C.S. J. Chu (2002): "Unit Root Tests in Panel Data: Asymptotic and Finite-sample Properties," *Journal of Econometrics*, 108, 124.

Palm, F., S. Smeekes and J.P. Urbain (2010): "Cross-Sectional Dependence Robust Block Bootstrap Panel Unit Root Tests," *Journal of Econometrics* (forthcoming).

Palm, F., S. Smeekes and J.P. Urbain (2008): "Bootstrap Unit Root Tests: Comparison and Extensions," *Journal of Time Series Analysis*, 29, 371-401.

Paparoditis, E. and D. Politis (2003): "Residual-Based Block Bootstrap for Unit Root Testing," *Econometrica*, 71(3), 813-855.

Parker, C., E. Paparoditis and M. Politis (2006): "Unit Root Testing Via the Stationary Bootstrap," *Journal of Econometrics*, 133, 601-638.

Pedroni, P., T. Vogelsang, M. Wagner, and J. Westerlund (2010): "Nonparametric Rank Tests for Non-stationary Panels," Mimeo.

Pedroni, P. (2010): Lecture Notes for Econ 464.

Phillips, P. C. B., and V. Solo (1992): "Asymptotics for Linear Processes," *Annals of Statistics*, 20(2), 971-1001.

Phillips, P. C. B., and Hyungsik R. Moon (1999): "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67(5), 1057-1111.

Politis, D. N. and J. P. Romano (1994): "Large Sample Confidence Regions Based on Subsamples Under Minimal Assumptions," *Annals of Statistics*, 22, 20312050.

Smeekes, S. (2009): "Detrending Bootstrap Unit Root Tests," *METEOR*, Research Memoranda 056.

Smeekes, S. (2006): "Bootstrapping Unit Root Tests," *Medium Econometrische Toepassingen*, 14 (4), 24-28.