ON THE ROBUSTNESS OF THE BLOCK BOOTSTRAP PANEL UNIT ROOT TEST

XIJIA LIU AND JIANXIN WEI

Abstract. Palm, Smeekes and Urbain (2011, PSU) proposed a bootstrap panel unit root test which can deal with a rather general cross-sectional dependency structure, including (but not exclusive to) the popular common factor framework. However, the robustness of the test is not fully investigated in their simulation study. In this article, we did Monte Carlo simulations to study the robustness of the tests proposed by PSU from simple to complex cross-sectional dependence structures. We compare the PSU test with two other representative tests in second generation panel unit root tests, the test proposed by Bai and Ng (2004) and Chang (2004), in terms of model setting, assumptions and small sample performance under different DGPs. We found that the test proposed by PSU exhibits robustness in general. However, it also has size distortions in the cases: 1. when negative moving average coefficients are present in the DGP, and 2. in the factor model, when only one component, either idiosyncratic errors or the common factors, is unit root process and the other component is stationary.

1. INTRODUCTION

The unit root test is an important tool in econometric analysis to test whether a time series is stationary or not. It is known that univariate unit root tests are of low power when the sample size is medium or small. To overcome this problem, testing the unit root in the panel setting was developed and has been intensively studied in the last two decades. These tests utilize the information from the cross-sectional dimension for attempting to increase the power. The so called first generation panel unit root tests, notably the LLC test by Levin et al. (2002) and the IPS test by Im et al. (2003), assume independence along the cross-sectional units. Then it was realized that the cross-sectional independence is an unrealistic assumption and the second generation panel unit root tests were developed in response to the need for cross-sectional correlations. The early approaches in dealing with cross-sectional correlation consist of imposing few or no restrictions on the residuals covariance matrix rather than through the common factor model. This approach is adopted by, among others, Chang (2004), who proposed the use of bootstrap to solve the nuisance parameter problem introduced by cross-sectional dependency. Then, the factor structure model becomes increasingly popular. Some studies treat the factors and factor loadings as nuisance parameter, for example, Moon and Perron (2004) and Pesaran (2007). An influential study by Bai and Ng (2004) considered a more general model in which the nonstationarity can enter into both the common factors part and the idiosyncratic errors part. Based on their PANIC [1] procedure, both common factors and idiosyncratic errors can be consistently estimated, thereby the hypothesis tests of nonstationarity can be done separately in both parts.

Although the cross-sectional dependency is allowed among the cross-sectional units, all of the existing methods have more or less restriction in modelling the dependency. For instance, in most models only the contemporaneous dependency through covariance matrix is allowed for idiosyncratic errors. Besides, the dependence between the common factors and idiosyncratic

Key words and phrases. Panel unit root test, Robustness, Monte Carlo simulation, Bootstrap.

[1]PANIC: Panel Analysis of Nonstationarity in Idiosyncratic and Common components

errors are assumed to be independent. In general, these assumptions are not always satisfied in applications. Considering these limitations, PSU proposed a test based on block bootstrap methodology which can deal with a rather wide range of cross-sectional dependence. In the model proposed by PSU, the factors and the idiosyncratic components are allowed to be dependent. Moreover, a wide array of possible dependencies between the idiosyncratic components are allowed. Therefore, many existing models can be specified as a special case of the model proposed in PSU.

The asymptotic validity of block bootstrap panel unit root tests has been rigorously proved in PSU. However, the small sample properties have not been fully investigated in their simulations. The incompleteness of such studies follows in several ways. First, they only compared their tests with two panel unit root tests from the first generation, but the robustness comparing to the second generation panel unit root tests is not shown. Second, the DGPs in their simulations did not fully explore the generality as their assumptions, for example, they only considered the case in which common factors and idiosyncratic errors are not dependent, neither in dynamic dependence nor the contemporaneous dependence. Third, they did not consider the case in which the common factors are $I(0)$ and idiosyncratic errors are $I(1)$. Fourth, the behavior of the test with multi factors is not studied. Thus in this study, we will do an investigation of the small sample properties of PSU in a more general and complicated DGP by comparison with two other panel unit root tests from the second generation. Specifically, on one hand the PSU test will be applied on the data which are generated by the model considered in Bai and Ng (2004) and Chang (2004) to examine the small sample performance; on the other hand, these three methods will be applied on data with more general cross-sectional dependency than the DGP used in the simulation study of PSU to see the robustness of PSU.

Some other aspects are also investigated in the simulation study. First, Palm et al. (2008) has done a study to compare the different bootstrap unit root tests in the univariate case. In a panel setting, however, no research has been done concerning this issue. In this paper, by comparing the test in Chang (2004) and PSU we try to have a look at the difference between sieve bootstrap and block bootstrap in the panel setting. Furthermore, in a univariate unit root test, it is widely known that the small sample performance can be seriously influenced when there are errors with MA parameters close to -1, see e.g. Perron and Ng (1996), Schwert (1989). This problem, however, was avoided in the research of panel unit root tests. For example, in the recent comparative study of panel unit root tests by Gengenbach et al. (2010), the moving average pattern was present in the DGP, however, the coefficients were restricted between 0.2 and 0.5. In this paper, this problem will also be investigated. Another aspect is that the simulation study of PSU and Chang (2004) did not consider the large N case. PSU considered maximum $N = 50$ and Chang (2004) only considered maximum $N = 10$. The performance of the tests under large $N = 100$ case will be considered in this paper.

The rest of the paper is organized as follows: In section 2, the PSU tests and two other selected tests are discussed in detail. Section 3 and section 4 present the settings and results of simulation studies. Section 5 concludes. In the rest of the paper we use the terms the PSU test, the BN test and the Chang test to denote the tests proposed by PSU, Bai and Ng (2004) and Chang (2004).

2. Panel unit root tests

This section describes three panel unit root tests in model settings, assumptions and testing procedures. The BN and Chang tests have been chosen in the comparison because their model settings are special cases of the model considered in PSU and these two tests are representative as the second generation panel unit root tests. Bai and Ng (2004) and (2010) are chosen as a representative among the tests with a factor structure for two reasons. First, the framework of BN tests have the closest form to PSU tests. For example, both of them consider a factor model and allow the nonstationarity to enter into the common factor and idiosyncratic errors. Second, the framework of BN tests is a general one among all of the existing methods. For example, as discussed in Bai and Ng (2010), the DGP of Phillips and Sul (2003) can be treated as a special case of BN's DGP; in addition, considering the homogeneous model or under the null hypothesis, the DGPs of Moon and Perron (2004) and Pesaran (2007) can also be expressed as BN's model. The Chang test is selected since it deals with another commonly specified cross-sectional dependency in which the dependency is through the covariance matrix. Apart from this, we choose the Chang test in respect that it also employs the bootstrap methodology to deal with the nuisance parameters introduced by cross-sectional dependency. Rather than the block bootstrap, the Chang test uses the sieve bootstrap and it is interesting to compare these two bootstrap schemes under different DGPs.

2.1. **PSU test.** The PSU test is designed for panels with various types of cross-sectional and temporal dependence. It has two attractive features. First, PSU can deal with a general type of cross-sectional dependency, and almost all forms of cross-sectional dependency in the literature are special cases of the model considered in PSU. To be specific, it allows for dependency between the factors and the idiosyncratic components, which are assumed to be independent in the common factor models considered by, for example, Bai and Ng (2004) and Moon and Perron (2004). In addition, the dependence structure can be through both the long run covariance of the error terms and general ARMA type lag polynomials. Second, the PSU test is easy to apply in practice. As the authors themselves state (p.85) "One does not have to model the dependence (both temporal and cross-sectional) in order to apply it...". In addition, one does not need to estimate the large number of nuisance parameters by using the bootstrap methodology.

2.1.1. *Model and Assumptions.* The model considered in Palm, Smeekes and Urbain (2011, PSU) is

$$y_t = \Lambda F_t + w_t \tag{1}$$

where the observations $y_t = (y_{1,t}, \ldots, y_{N,t})'$, $t = 1, \ldots, T$, the common factor $F_t = (F_{1,t}, \ldots, F_{d,t})'$, with d the number of factors, the non-random factor loadings $\Lambda = (\lambda_1, \ldots, \lambda_N)'$ and w_t are the idiosyncratic components. Both the factors and the idiosyncratic components are of AR(1) form, and they are generated by

$$F_t = \Phi F_{t-1} + f_t \tag{2}$$

$$w_t = \Theta w_{t-1} + v_t \tag{3}$$

where $\Phi = diag(\phi_1, \ldots, \phi_d)$ and $\Theta = diag(\theta_1, \ldots, \theta_N)$. Furthermore, the terms f_t and v_t are constructed as

$$\begin{bmatrix} v_t \\ f_t \end{bmatrix} = \Psi(L)\varepsilon_t = \begin{bmatrix} \Psi_{11}(L) & \Psi_{12}(L) \\ \Psi_{21}(L) & \Psi_{22}(L) \end{bmatrix} \begin{bmatrix} \varepsilon_{v,t} \\ \varepsilon_{f,t} \end{bmatrix}, \tag{4}$$

where the lag polynomial $\Psi(z)_{(N+d)\times(N+d)} = \sum_{j=0}^{\infty} \Psi_j z^j$ with $\Psi_0 = I_{N+d}$, note that each element in $\Psi(L)$ is a scalar lag polynomial. $\Psi_(z)$ and ε_t need to satisfy the following assumptions:

Assumption 1. $\det(\Psi(z)) \neq 0$ for all $\{z \in \mathcal{C} : |z| = 1\}$ and $\sum_{j=0}^{\infty} j|\Psi_j| < \infty$.

Assumption 2. ε_t is i.i.d. with $E\varepsilon_t = 0$, $E\varepsilon_t\varepsilon_t' = \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ and $E|\varepsilon_t|^{2+\epsilon} < \infty$ for some $\epsilon > 0$.

From this model framework, almost all kinds of cross-sectional dependence can be captured. First, the correlation among the idiosyncratic errors are through the covariance matrix Σ_{11} and lag polynomial $\Psi_{11}(L)$. Second, if more factors are included in the model, the factors can also be correlated, this correlation is through Σ_{22} and $\Psi_{22}(L)$. Third, the correlation between idiosyncratic errors and the common factors are also included, which is captured by Σ_{12}, $\Psi_{12}(L)$ and $\Psi_{21}(L)$. PSU consider the null hypothesis $H_0 : Y_{i,t}$ has a unit root for all i. Rather than testing unit roots separately in the common factors and the idiosyncratic components, the PSU test is to test if the panel has a unit root regardless of where the unit root comes from. This makes the method appealing if the practitioner only wants to know if the whole panel has a unit root or not.

2.1.2. *Testing Procedure.* The PSU test can be implemented using two statistics, the group-mean statistic τ_{gm} and the pooled statistic τ_p, based on the auxiliary regression $y_{it} = \rho_i y_{i,t-1} + u_i$. The two statistics are constructed in the form of the Dickey-Fuller coefficient test instead of the augmented Dickey-Fuller t test, which can be viewed as the simplified group-mean test and pooled test in Levin et al. (2002) and Im et al. (2003). This simplification will introduce nuisance parameters into the limiting distribution as a consequence. However, since the whole limiting distribution will be correctly mimicked by the bootstrap distribution asymptotically, this simplification is harmless. The test is very easy to implement in practice with the following steps:
Step 1. Calculate the statistics τ_p or τ_{gm}.

$$\tau_p = T\frac{\sum_{i=1}^{N}\sum_{t=2}^{T} y_{i,t-1}\Delta y_{i,t}}{\sum_{i=1}^{N}\sum_{t=2}^{T} y_{i,t-1}^2},$$
$$\tau_{gm} = \frac{1}{N}\sum_{i=1}^{N} T\frac{\sum_{t=2}^{T} y_{i,t-1}\Delta y_{i,t}}{\sum_{t=2}^{T} y_{i,t-1}^2},$$

where τ_{gm} is the average of the Dickey-Fuller coefficient statistics from the individual regressions of $\Delta y_{i,t}$ on $y_{i,t-1}$ for each $i = 1, \ldots, N$, τ_p is the Dickey-Fuller coefficient statistic from the pooled regression of $y_{i,t}$ on $y_{i,t-1}$.
Step 2. Calculate the bootstrap statistics.
(a) Obtain the recentered residuals

$$\hat{u}_{i,t} = y_{i,t} - \hat{\rho}_i y_{i,t-1} - \frac{1}{T-1}\sum_{t=2}^{T}(y_{i,t} - \hat{\rho}_i y_{i,t-1})$$

where $\hat{\rho}_i = \frac{\sum_{t=2}^{T} y_{i,t-1}y_{i,t}}{\sum_{t=2}^{T} y_{i,t-1}^2}$ for $i = 1, \ldots, N$ is the estimate of ρ_i from the regression $y_{it} = \rho_i y_{i,t-1} + u_{it}$.

(b) Choose a block length b. Draw the block index $i_0, \ldots, i_{k-1}$ i.i.d. from the discrete uniform distribution on $\{1, 2, \ldots, T-b\}$, the number of blocks $k = \lfloor (T-2)/b \rfloor + 1$.

(c) Construct the bootstrap errors $u_2^*, \ldots, u_T^*$ as

$$u_t^* = \hat{u}_{i_m+s},$$

where $m = \lfloor (t-2) \rfloor$ and $s = t - mb - 1$.

(d) Let $y_1^* = y_1$ and construct y_t^* for $t \geq 2$ recursively as $y_t^* = y_{t-1}^* + u_t^*$.

(e) Calculate the bootstrap versions of τ_p and τ_{gm}

$$\tau_p^* = T\frac{\sum_{i=1}^N \sum_{t=2}^T y_{i,t-1}^* \Delta y_{i,t}^*}{\sum_{i=1}^N \sum_{t=2}^T y_{i,t-1}^{*2}},$$

$$\tau_{gm}^* = \frac{1}{N}\sum_{i=1}^N T\frac{\sum_{t=2}^T y_{i,t-1}^* \Delta y_{i,t}^*}{\sum_{t=2}^T y_{i,t-1}^{*2}}.$$

Step 3. Conduct the bootstrap test.

Taking τ_p for example, repeat steps 2(a)-(e) B times, obtaining B bootstrap test statistics $\tau_{pi}^*, i = 1, 2, ..., B$ and reject the unit root null hypothesis if

$$\frac{\sum_{i=1}^B I\{\tau_{pi}^* < \tau_p\}}{B} < \lambda,$$

where $I(\cdot)$ is the indicator function, λ is the level of the test.

Remark 1. *In step 2(b), the block length b should satisfy $b = o(T^{1/2})$ as $T \to \infty$. The selection of b is more difficult than in the case of the stationary time series and still an open question. One can either use the minimum volatility method and calibration method proposed by Politis et al. (1999) or fix b as $b = CT^{1/k}$.*

2.2. **PANIC.** The main theoretical results of Bai and Ng (2004) indicate that in the PANIC procedure, the factors and factor loadings can be consistently estimated by the principal components method whenever the idiosyncratic errors are $I(0)$ or $I(1)$. The consistent estimation of common factors and idiosyncratic errors makes the separate tests on both parts feasible. Thereby, the original aim of the PANIC procedure is to distinguish the sources of nonstationarity. As pointed out by Bai and Ng (2004), if nonstationarity is introduced by common factors, then the nonstationarity of each cross section unit is due to a common source; if unit roots are detected from idiosyncratic errors, we say that nonstationarity is due to a series-specific source. In addition, Bai and Ng (2004) and Bai and Ng (2010) also provide a further pooled test of nonstationarity of the idiosyncratic part. Next, we give the model and assumptions of PANIC, then the main ideas will be intuitively illustrated, and the testing procedure is listed at last.

2.2.1. *Model and Assumptions.* The model applied by Bai and Ng (2004) can be expressed by equations (1) to (4) with the following extra assumptions:

Assumption 3. (i) For non-random λ_i, $\|\lambda_i\| < M$;.

Assumption 4. (i) $\Sigma_{12} = \Sigma_{21} = 0$; (ii) Σ_{11} is a diagonal matrix; (iii) $E\|\varepsilon_{f,t}\|^4 < M$;

Assumption 5. (i) $\Psi_{12} = \Psi_{21} = 0$; (ii) Ψ_{11} is a diagonal matrix; (iii) $\Psi_{22}(1)$ has rank r_1, $0 \le r_1 \le r$; (iv) $Var(\varepsilon_{f,t}) = \sum_{j=0}^{\infty} \psi_{22j}\Sigma_{22}\psi_{22j}' > 0$.

Remark 2. *Assumption 4 (i) and 4 (ii) guarantee that the common factors and idiosyncratic errors are independent. Assumption 4 (ii) and Assumption 5 (ii) avoid the contemporaneous and dynamic dependence present in idiosyncratic errors.*

Remark 3. *As discussed before, PANIC considered both the testing of unit root for each units separately and the testing the joint null hypothesis. If one only wants to get consistent estimation and individually test nonstationarity of common factors, then Assumption 4 (ii) can be removed, i.e. the common factors and factor loadings still can be consistently estimated when the contemporaneous dependence is present in the idiosyncratic part.*

2.2.2. *Testing procedure.* Different from the classical estimating procedure, PANIC can provide consistent estimation of factors and factor loadings ignoring the integrated order of idiosyncratic errors. The main idea is that the principal components method is applied to the first differenced data, then the consistent estimation of factors and idiosyncratic errors is obtained by the cumulative sum. Based on these results, the different tests of nonstationarity can be applied. Regarding the idiosyncratic part, the ADF test can be performed on the estimated idiosyncratic errors of each units. Then, using a combined p-values statistic, by the justification and more rigorous proof of Westerlund and Larsson (2009), the statistic of combining the p-value of the individual ADF test converges to the standard normal distribution as $N \to \infty$. Furthermore, Bai and Ng (2010) proposed two other tests for nonstationarity of idiosyncratic errors. Considering the small sample performance which is reported in that paper, we only chose the pooling statistics which follow the spirit of Moon and Perron (2004). As their theoretical results, the MP type test statistics are asymptotic normal distributed as $N, T \to \infty$ with $N/T \to 0$. Regarding the common factors part, Bai and Ng (2004) distinguish two situations. If the number of common factors is equal to 1, then we can do an ADF test based on the estimates of common factors. If the number of common factors is larger than 1, then the nonstationarity of the common factors part will be tested by the common trends tests, i.e. the number of stochastic trends in the common factors will be identified. When the common factors share one or more stochastic trends, it implies that the common factors are $I(1)$, otherwise, they are $I(0)$. Furthermore, if one is interested in concluding the nonstationarity of the panel, then the joint inference should be made from these two parts. The details of the whole estimating and testing procedure are as follows:

Step 1. Estimating the common factors and idiosyncratic errors.
Let $x_{it} = \Delta X_{it}$, $f_t = \Delta F_t$, and $z_{it} = \Delta e_{it}$, then the model in first-differenced form is $x_{it} = \lambda_i' f_t + z_{it}$. Applying the method of principal components [2] to x yields r estimated factors $\widehat{f}_t$ and the associated loadings $\widehat{\lambda}_i$, and the estimated residuals, $\widehat{z}_{it} = x_{it} - \widehat{\lambda}_i' \widehat{f}_t$. Then by applying cumulative sum, we have

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

Step 2. Testing the nonstationarity of the idiosyncratic errors part.
(a) The combining p-values test is as follows: Let $ADF_{\widehat{e}}(i)$ be the t statistic for testing $d_{i0} = 0$ in the univariate augmented auto-regression (with no deterministic terms)

$$\Delta \widehat{e}_{it} = d_{i0} \widehat{e}_{it-1} + d_{i1} \Delta \widehat{e}_{it-1} + \cdots + d_{ip} \Delta \widehat{e}_{it-p} + error.$$

Let $P_{\widehat{e}}(i)$ be the p-values associated with $ADF_{\widehat{e}}(i)$, then the combining p-values statistic is defined as

$$P_{\widehat{e}} = -N^{-1/2} \left(\sum_{i=1}^{N} \log P_{\widehat{e}}(i) + N \right).$$

(b) The MP type test is as follows: Pooling OLS estimation of the model

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

[2]For the details of principle components method, see Bai and Ng (2002)

we have $\widehat{\rho} = tr\left(\widehat{e}'_{-1}\widehat{e}\right)/tr\left(\widehat{e}'_{-1}\widehat{e}_{-1}\right)$ where $\widehat{e} = \{\widehat{e}_{it}\}$ and $\widehat{e}_{-1} = \{\widehat{e}_{it-1}\}$ are $(T-2)\times N$ matrices. Define the short run, long run and one side variance of ϵ_{it} as

$$\sigma^2_{\epsilon i} = E\left(\epsilon^2_{it}\right) = \sum_{j=0}^{\infty} D^2_{ij},\ \varpi^2_{\epsilon i} = \left(\sum_{j=0}^{\infty} D_{ij}\right)^2,\ \lambda_{\epsilon i} = \left(\varpi^2_{\epsilon i} - \sigma^2_{\epsilon i}\right)/2$$

Assuming that the following limits exist and that the first three are strictly positive:

$$\varpi^2_{\epsilon} = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\varpi^2_{\epsilon i},\ \sigma^2_{\epsilon} = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\sigma^2_{\epsilon i},$$

$$\phi^4_{\epsilon} = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\varpi^4_{\epsilon i},\ \lambda_{\epsilon} = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\lambda_{\epsilon i},$$

let

$$\widehat{\varpi}^2_{\epsilon} = N^{-1}\sum_{i=1}^{N}\widehat{\varpi}^2_{\epsilon i},\ \widehat{\sigma}^2_{\epsilon} = N^{-1}\sum_{i=1}^{N}\widehat{\sigma}^2_{\epsilon i},\ \widehat{\phi}^4_{\epsilon} = N^{-1}\sum_{i=1}^{N}\widehat{\varpi}^4_{\epsilon i},\ \widehat{\lambda}_{\epsilon} = N^{-1}\sum_{i=1}^{N}\widehat{\lambda}_{\epsilon i}$$

be consistent estimates of ϖ^2_{ϵ}, σ^2_{ϵ}, ϕ^4_{ϵ} and λ_{ϵ}, respectively. Additional assumptions are necessary for the consistent estimation of the long-run and one-sided long-run variances which are given in Moon and Perron (2004). Then the bias-corrected pooled PANIC autoregressive estimator ρ can be defined as

$$\widehat{\rho}^{+} = \frac{tr\left(\widehat{e}'_{-1}\widehat{e}\right) - NT\widehat{\lambda}_{\epsilon}}{tr\left(\widehat{e}'_{-1}\widehat{e}_{-1}\right)}$$

and two test statistics are

$$P_a = \frac{\sqrt{N}T\left(\widehat{\rho}^{+}-1\right)}{\sqrt{2\widehat{\phi}^4_{\epsilon}/\widehat{\omega}^4_{\epsilon}}} \text{ and } P_b = \sqrt{N}T\left(\widehat{\rho}^{+}-1\right)\sqrt{N^{-1}T^{-2}tr\left(\widehat{e}'_{-1}\widehat{e}_{-1}\right)\frac{\widehat{\omega}^2_{\epsilon}}{\widehat{\phi}^4_{\epsilon}}}.$$

.

Step 3. Testing nonstationarity of the common factors part.

If $r = 1$, then we apply the ADF test to the common factor. Let $ADF^c_{\widehat{F}}$ be the t statistic for testing $\delta_0 = 0$ in the univariate augmented auto-regression (with an intercept):

$$\Delta\widehat{F}_t = c + \delta_0\widehat{F}_{t-1} + \delta_1\Delta\widehat{F}_{t-1} + \cdots + \delta_p\Delta\widehat{F}_{t-p} + error.$$

If $r > 1$, then we do a sequence of common trends tests to the common factors to determine the number of common stochastic trends among the common factors. Specifically, demean $\widehat{F}_t$ and define $\widehat{F}^c_t = \widehat{F}_t - \overline{\widehat{F}}$, where $\overline{\widehat{F}} = (T-1)^{-1}\sum_{t=2}^{T}\widehat{F}_t$. Start with $m = r$:

(a) Let $\widehat{\beta}_{\perp}$ be the m eigenvectors associated with the m largest eigenvalues of $T^{-2}\sum_{t=2}^{T}\widehat{F}^c_t\widehat{F}^{c\prime}_t$. Let $\widehat{Y}^c_t = \widehat{\beta}'_{\perp}\widehat{F}^c_t$.

(b) Let $K(j) = 1 - j/(J+1)$, $j = 0, 1, ..., J$:

(i) Let $\widehat{\xi}^c_t$ be the residuals from estimating a first-order VAR in $\widehat{Y}^c_t$, and let

$$\widehat{\Sigma}^c_1 = \sum_{j=1}^{J} k(j)\left(T^{-1}\sum_{t=2}^{T}\widehat{\xi}^c_{t-j}\widehat{\xi}^{c\prime}_t\right).$$

(ii) Let $\nu^c_c(m)$ be the smallest eigenvalue of

$$\widehat{\Phi}^c_c(m) = .5\left[\sum_{t=2}^{T}\left(\widehat{Y}^c_t\widehat{Y}^{c\prime}_{t-1} + \widehat{Y}^c_{t-1}\widehat{Y}^{c\prime}_t\right) - T\left(\widehat{\Sigma}^c_1 + \widehat{\Sigma}^{c\prime}_1\right)\right]\left(\sum_{t=2}^{T}\widehat{Y}^c_{t-1}\widehat{Y}^{c\prime}_{t-1}\right)^{-1}$$

(iii) Define $MQ_c^c(m) = T[\nu_c^c(m) - 1]$.

(c) If $H_0 : r_1 = m$ is rejected, set $m = m - 1$ and return to step A. Otherwise, $\widehat{r}_1 = m$ and stop.

Remark 4. *For an empirical study, there should be an additional step before the PANIC procedure, i.e. the number of common factors should be estimated. Bai and Ng (2002) proposed several criteria to determine the number of common factors. However, we skip this step and fix the number of common factors as a known value for two reasons. First, it is possible that one could wrongly estimate the number of common factors as 0 and the PANIC procedure can not be applied in this case. Second, Gengenbach et al. (2010) did a comparison study about the performance of different panel unit root tests. In that study, they also fixed the number of common factors as a known value, and investigated the influence of wrong model specification by fixing the number of common factors as a wrong number. The result in their study indicates that if the number of common factors is misspecified then all statistics show strong size distortion. In this study, we will skip this step and the corresponding influence.*

Remark 5. *In fact, Bai and Ng (2004) also suggest another common trends test MQ_f^c for the case in which the delay parameter of the VAR model is fixed. However we skip it in this study for two reasons. First, as pointed out by Bai and Ng (2004), the MQ_f^c test is valid only when the common trends can be represented as finite order $AR(p)$ process, and the MQ_c^c test is relatively more general. Second, we found that there is no significant difference in the performance of these two tests in our simulations. This is also pointed out by Bai and Ng (2004) and Gengenbach et al. (2010). Furthermore, Bai and Ng (2004) also considered Johansen's trace test in their simulations, however, in this study we also skip it.*

Remark 6. *Regarding the pooled tests on the idiosyncratic errors, Bai and Ng (2010) also proposed an MSBt[3] type test based on their PANIC residuals, however, by considering its relative poor small sample performance we did not select it in our study.*

Remark 7. *We apply the same method to determine J and p as Bai and Ng (2004), i.e. p and J are all determined by $4\lceil\min(N,T)/100\rceil^{1/4}$, where $\lceil\cdot\rceil$ denotes the smallest integer greater or equal to $\cdot$.*

2.3. **Chang test.** Chang (2004) developed a bootstrap panel unit root test which also allows for cross-sectional dependency. Chang (2004) considered a panel model without common factors and the cross-sectional dependency is only through the long run covariance matrix. Like the test proposed by PSU, the test statistics in Chang (2004) involve several nuisance parameters and the bootstrap methodology is used to mimic the distribution of the statistics. In contrast to the test proposed by PSU, Chang (2004) utilizes the sieve bootstrap instead of block bootstrap. It was found in Palm et al. (2008) that the sieve bootstrap ADF test performs better than the block bootstrap DF test. It is interesting to know if this is still the case for panel unit root tests.

2.3.1. *Model and Assumptions.* The model considered in Chang (2004) can be expressed by equations (1) to (4) with the following assumptions.

Assumption 6. All of the components involving common factors are zero.

Assumption 7. $\Psi_{11}(L) = diag(\psi_1(L), \ldots, \psi_N(L))$, i.e. each $v_{it} = \psi_i(L)\varepsilon_{v,t}$ is a general linear process with $\psi_i(z) \neq 0$ for all $|z| \leq 1$, and $\sum_{j=0}^{\infty} |k|^s |\psi_{i,j}| < \infty$ for some $s \geq 1$, for all $i = 1, \ldots, N$.

Assumption 8. $\varepsilon_{v,t}$ is i.i.d. with $E\varepsilon_{v,t} = 0$, $E\varepsilon_{v,t}\varepsilon_{v,t}' = \Sigma_{11}$ and $E|\varepsilon_{v,t}|^r < \infty$ for some $r \geq 4$.

Heterogeneity is allowed in individual serial correlation structures, i.e. $\psi_i(L)$ is different across i. The cross-sectional dependency is through the covariance matrix Σ_{11}, which can be any positive definite matrix.

---

[3]The pooled MSB type test was originally proposed by Sargan and Bhargava (1983)

2.3.2. *Testing Procedure.* The Chang test is constructed from the estimation of the entire N cross-sectional units. In order to do so, each time series y_{it} is approximated by a finite order autoregressive process of order p_i increasing with T and all of the N regression equations are stacked in one regression equation.

Step 1. Stack all of the observations in one regression equation

(a) Fit each time series y_i by an augmented autoregression,

$$\Delta y_{it} = \alpha_i y_{i,t-1} + \sum_{k=1}^{p_i} \alpha_{i,k} \Delta y_{i,t-k} + \varepsilon_{it}^{p_i}, \tag{5}$$

or

$$\Delta y_{it} = \sum_{k=1}^{p_i} \alpha_{i,k} \Delta y_{i,t-k} + \varepsilon_{it}^{p_i}. \tag{6}$$

The lag order p_i is selected using information criteria.

(b) Stack the augmented autoregression in the matrix form

$$\begin{pmatrix} \Delta y_1 \\ \vdots \\ \Delta y_N \end{pmatrix} = \begin{pmatrix} y_{l,1} & & 0 \\ & \ddots & \\ 0 & & y_{l,N} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix} + \begin{pmatrix} X_1^{p_1} & & 0 \\ & \ddots & \\ 0 & & X_N^{p_N} \end{pmatrix} \begin{pmatrix} \beta_1^{p_1} \\ \vdots \\ \beta_N^{p_N} \end{pmatrix} + \begin{pmatrix} \varepsilon_1^{p_1} \\ \vdots \\ \varepsilon_N^{p_N} \end{pmatrix}$$

or more compactly

$$\Delta y = Y_l \alpha + X_p \beta_p + \varepsilon_p. \tag{7}$$

where for all $i = 1, \ldots, N$, $y_{l,i} = (y'_{i,0}, \ldots, y'_{i,T-1})'$ is $NT \times 1$, $\beta_i^{p_i} = (\alpha_{i,1}, \ldots, \alpha_{i,p_i})'$ and $X_i^{p_i} = (x_{i1}^{p_i}, \ldots, x_{iT}^{p_i})'$ is $NT \times \sum p_i$ with $x_{it}^{p_i\prime} = (\Delta y_{i,t-1}, \ldots, \Delta y_{i,t-p_i})$.

To test the null hypothesis H_0: $\alpha = (\alpha_1, \ldots, \alpha_N) = 0$, i.e y_{it} is unit root process for all i, an F-type test $\hat{\alpha}'(var(\hat{\alpha}))^{-1}\hat{\alpha}$ is used.

Step 2. Calculate the test statistics.

(a) Calculate the estimates of the covariance matrix Σ. Letting $\tilde{\varepsilon}_t^p = (\tilde{\varepsilon}_{1t}^{p_1}, \ldots, \tilde{\varepsilon}_{Nt}^{p_N})'$, Σ is estimated as $\hat{\Sigma} = T^{-1} \sum_{t=1}^{T} \tilde{\varepsilon}_t^p \tilde{\varepsilon}_t^{p\prime}$.

(b) Calculate the OLS-based F-type statistics $F_{OT} = \hat{\alpha}'_{OT}(var(\hat{\alpha}_{OT}))^{-1}\hat{\alpha}_{OT} = A'_{OT} M_{OT}^{-1} A_{OT}$, where

$$A_{OT} = Y_l'(I - X_p(X_p'X_p)^{-1}X_p')\Delta y,$$

$$\begin{aligned} M_{OT} =& Y_l'(\tilde{\Sigma} \otimes I_T)Y_l - Y_l'X_p(X_p'X_p)^{-1}X_p'(\tilde{\Sigma} \otimes I_T)Y_l - Y_l'(\tilde{\Sigma} \otimes I_T)X_p(X_p'X_p)^{-1}X_p'Y_l \\ &+ Y_l'X_p(X_p'X_p)^{-1}X_p'(\tilde{\Sigma} \otimes I_T)X_p(X_p'X_p)^{-1}X_p'Y_l. \end{aligned}$$

The F-type test is a two sided test which rejects the null $\alpha_i = 0$ for all i when $\alpha \neq 0$, i.e. they reject the unit roots not only against the stationary but also against the explosive process, with $\alpha > 0$ for some i. This will have a negative effect on the results. Chang provided a K test which is a modification of the F-type test by adjusting the positive estimates of α to zero,

$$K_{OT} = A'_{OT} \cdot I(A'_{OT} \leq 0) M_{OT}^{-1} A_{OT} \cdot I(A_{OT} \leq 0),$$

where $I(\cdot)$ is the indicator function. Chang (2004) also considered the GLS-based F-type test and K test, where $\hat{\alpha}$ in equation (7) is obtained by generalized least square estimation. As mentioned in the simulation study of Chang (2004), F_{OT} often performs better than F_{GT} in finite samples, thereby we only consider the OLS-based F_{OT} and K_{OT} tests in our study for simplicity.

The distribution of the F_{OT} is non-standard with several nuisance parameters and the bootstrap method is utilized to mimic the distribution of the statistic. The bootstrap procedure used in the Chang test is the residual based sieve bootstrap and the cross-sectional dependency is preserved by resampling the errors as a vector. The bootstrap procedure is implemented in the following steps:

Step 3. Calculate the bootstrap test statistics.

(a) Sample the bootstrap residuals from the recentered residuals from $(\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_T)$, where $\tilde{\varepsilon}_i = (\tilde{\varepsilon}_{1i}^{p_1}, \dots, \tilde{\varepsilon}_{Ni}^{p_N})'$ for $i = 1, \dots, T$.

(b) Construct the bootstrap sample y_{it}^*.

$$v_{it}^* = \tilde{\alpha}_{i,1}^{p_i} v_{i,t-1}^* + \cdots + \tilde{\alpha}_{i,p_i}^{p_i} v_{i,t-p_i}^* + \varepsilon_{it}^*$$

where $(\tilde{\alpha}_{i,1}^{p_i}, \cdots, \tilde{\alpha}_{i,p_i}^{p_i})$ are the coefficient estimates from the restricted fitted regression (6). In the end, obtain the bootstrap sample y_{it}^* under the unit root restriction by taking partial sums of u_{it}^* by $y_{it}^* = y_{i0}^* + \sum_{k=1}^{t} v_{ik}^*$, start from $y_{i0}^* = 0$.

(c) Calculate

$$\Delta y_{it}^* = \alpha_i y_{i,t-1}^* + \sum_{k=1}^{p_i} \alpha_{i,k} \Delta y_{i,t-k}^* + \varepsilon_{it}^*, \tag{8}$$

and write these in a matrix form as

$$\Delta y^* = Y_l^* \alpha + X_p^* \beta_p + \varepsilon^*.$$

(d) Calculate the bootstrap statistics F_{OT}^*, K_{OT}^* (F^*, K^*) as

$$F_{OT}^* = A_{OT}^*{}' M_{OT}^{*-1} A_{OT}^*,$$
$$K_{OT}^* = A_{OT}^*{}' \cdot I(A_{OT}^*{}' \leq 0) M_{OT}^{*-1} A_{OT}^*{}' \cdot I(A_{OT}^*{}' \leq 0).$$

Step 4. Conduct the bootstrap test

Taking the F_{OT} test for example, repeat Step 3 B times to obtain the empirical distribution of the bootstrap test statistics $F_{OTi}^*, i = 1, ..., B$. The null hypothesis is rejected if

$$\frac{\sum_{i=1}^{B} I\{F_{OT}^* > F_{OT}\}}{B} < \lambda.$$

Remark 8. *In step 1(a), the regression can be fitted either with restriction or without restriction. This will not affect the consistency of the test. However, it is believed that, when the data is from the alternative hypothesis, obtaining the residuals from the regression without restriction will affect the power of the tests.*

Remark 9. *In step 1(b), the initial values are not important, we can set $y_{i,p_{i-1}}, \dots, y_0$ as zero.*

Remark 10. *When generating the bootstrap, initial values need to be considered in two places. First, when generating the error term v_{it}^*, we may use the first p_i values of v_{it}, however, this will not ensure stationarity. Alternatively, we may generate extra M values of v_{it}^* and discard the first M values of v_{it}^* to make sure that v_{it}^* which we used is stationary. In this case, the initial values do not matter and we may simply set them to zero. Second, when generating the bootstrap sample y_{it}^*, the initial values y_{i0}^* do not matter and are set to zero as with y_{i0}.*

Remark 11. *We can use the estimates of the covariance matrix $\hat{\Sigma}^*$ from step 2(a) as $\hat{\Sigma}^*$ when calculating the bootstrap statistics, since this is the population variance in the probability space of bootstrapping.*

Remark 12. *In step 3(c), when estimating the bootstrap version $\hat{\alpha}^*$ in equation (8), the lagged order p_i is the same as in the autoregression (5) or (6).*

Remark 13. *Since the Chang test is an F type test, the larger the statistic is the more evidence there is against the null hypothesis. Hence in step 4, the null hypothesis is rejected if the proportion of $F^* > F$ is less than the level of the test.*

3. Design of the Monte Carlo Studies

In this section, we conduct Monte Carlo simulations to examine the robustness of the PSU test from two aspects. First, considering the PSU test's flexibility in dealing with various cross-sectional dependencies, it is questioned that the PSU test will sacrifice the performance on particular DGPs in order to obtain the robustness in general. To investigate this issue, we choose DGPs designed for the model in Chang (2004) and Bai and Ng (2004) and only compare the relevant test with the PSU test. Second, we examine the robustness of the PSU test on a more general DGP than that used in the simulation study of PSU. This DGP is misspecified for the model of the Chang test and the BN test, we include both of these two tests as references for the robustness of the PSU test.

As mentioned before, the models proposed in Chang (2004) and Bai and Ng (2004) are special cases of the model in PSU. Hence, the corresponding DGP can also be described as special cases in the general framework of the PSU's DGP. We consider the following PSU's DGP corresponding to Model (1) to (4) and Assumption 2.

1. Generate $\varepsilon_t = \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \sim N(0, \Sigma)$ where $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$.
2. Generate
$$\begin{pmatrix} v_t \\ f_t \end{pmatrix} = A \begin{pmatrix} v_{t-1} \\ f_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} + B \begin{pmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{pmatrix}, \tag{9}$$
where $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$, $B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$.
3. Generate
$$F_{i,t} = \phi_i F_{i,t-1} + f_{i,t} \quad \text{with} \quad f_{i,0} = 0, \tag{10}$$
$$w_{i,t} = \theta_i w_{i,t-1} + v_{i,t} \quad \text{with} \quad v_{i,0} = 0. \tag{11}$$
$$\tag{12}$$
4. Construct
$$y_t = \Lambda F_t + w_t.$$

As in PSU and Chang (2004), the dynamic structure in the model is captured by an ARMA process with parameters A and B. By making restrictions to A, B and Σ, all of the three DGPs in the simulation study fit into this framework. Table 1 summarizes how the corresponding parameters are restricted for different DGP, which will be introduced separately in the following.

DGP 1: Chang (2004)

The model considered in Chang (2004) does not have common factors and hence all of the factor

Table 1. Summary of three DGPs

		Chang	BNG		PSU Gen1		PSU Gen2	
			$d=1$	$d=2$	$d=1$	$d=2$	$d=1$	$d=2$
Σ	Σ_{11}	No Restr.	I_N		No Restr.			
	Σ_{22}	0	1	No Restr.				
	Σ_{12}	0	0	0				
A	A_{11}	$diag(a_1,\dots,a_N)$	$diag(a_1,\dots,a_N)$		No Restr.		No Restr.	
	A_{22}	0	a_2	$diag(a_{21},a_{22})$	a_2	No Restr.		
	A_{12},A_{21}	0	0		0			
B	B_{11}	$diag(b_1,\dots,b_N)$	$diag(b_1,\dots,b_N)$		No Restr.		No Restr.	
	B_{22}	0	b_2	$diag(b_{21},b_{22})$	b_2	No Restr.		
	B_{12},B_{21}	0	0		0			

relevant elements in A, B, and Σ are simply zero, accordingly the DGP is simplified as

$$
\begin{aligned}
y_t &= w_t,\\
w_{i,t} &= \theta_i w_{i,t-1} + v_{i,t}, \quad i = 1, ..., N,\\
v_{i,t} &= a_i v_{i,t-1} + \varepsilon_{i,t} + b_i \varepsilon_{i,t-1} \quad \text{with} \quad \varepsilon_t \sim N(0, \Sigma),
\end{aligned}
$$

where Σ is generated as in Chang (2004) [4]. Note that A_{11} and B_{11} in equation (9) become diagonal matrices $diag(a_1, ..., a_N)$ and $diag(b_1, ..., b_N)$ since dynamic dependence structure is not allowed. As we mentioned before, the performance of the unit root test is influenced by the negative moving average parameters. In order to investigate how different ARMA parameter values affect the results, we choose three different settings of the ARMA parameters a_i and b_i,

(1) $a_i \sim U[-0.8, 0.8]$, $b_i \sim U[-0.8, 0.8]$, for all i,
(2) $a_i = 0$, $b_i \sim U[-0.8, -0.4]$, for all i,
(3) $a_i = 0$, $b_i \sim U[0.4, 0.8]$, for all i,

The ARMA parameter in the first setting can be considered as no restriction as long as it guarantees $v_{i,t}$ a stationary and invertible ARMA process. The second setting considers the case with negative MA coefficients in which we have a particular interest. The third setting takes positive MA coefficients to make a comparison with the negative one.

The simulation under this DGP considers all combinations of $T = 25, 50, 100$, and $N = 10, 25, 50, 100$. A small $N = 10$ is included because the Chang test has good performance especially when N is small.[5].

Both size and power properties are studied with two settings of parameter θ_i.

I. Unit root panel: $\theta_i = 1$ for all i.
II. Stationary panel: $\theta_i \sim U[0.8, 1]$ for all i.

[4] Σ is generated via steps:

1. Generate an $(N \times N)$ matrix $U \sim [U, 1]$,
2. Construct from U an orthogonal matrix $H = U(U'U)^{-1/2}$,
3. Generate a set of N eigenvalues, $\lambda_1, ..., \lambda_N$. Let $\lambda_1 = 0.1$, $\lambda_N = 1$ and draw $\lambda_2, ..., \lambda_{N-1}$ from $U[0.1, 1]$,
4. Form a diagonal matrix $\Lambda = diag(\lambda_1, ..., \lambda_N)$,
5. Construct the covariance matrix Σ as $\Sigma = H \Lambda H'$.

[5] In the simulation study of Chang (2004), they only considered small N, i.e. $N = 5, 10$, and performance of the Chang test when N is large is not known, our simulation study makes a contribution to this aspect

DGP 2: Bai and Ng (2004)

Common factors are included in the DGP under Bai and Ng (2004), but the factors are assumed to be independent with the idiosyncratic errors. Hence, the corresponding elements $\Sigma_{12}, A_{12}, A_{21}, B_{12}$ and B_{21} are set to zero and the factors and the idiosyncratic components can be generated separately. For the idiosyncratic component,

$$\begin{aligned} w_{i,t} &= \theta_i w_{i,t-1} + v_{i,t}, \\ v_{i,t} &= a_i v_{i,t-1} + \varepsilon_{1i,t} + b_i \varepsilon_{1i,t-1}. \end{aligned}$$

The pooled test in BN's procedure requires independence among the idiosyncratic errors, thereby $\varepsilon_{1,t} \sim N(0, I_N)$ and A_{11}, B_{11} are diagonal as $A_{11} = diag(a_{11}, ..., a_{1N})$, $B_{11} = diag(b_{11}, ..., b_{1N})$ as in DGP 1. For the ARMA parameter values a_{1i} and b_{1i} , we choose the same three settings as in DGP 1.

For the factor part, both a single factor and two factors are considered. For a single factor, the ARMA parameters A_{22} and B_{22} are scalars, denoted by a_2, b_2,

$$\begin{aligned} F_t &= \phi F_{,t-1} + f_t, \\ f_t &= a_2 f_{t-1} + \varepsilon_{2,t} + b_2 \varepsilon_{2,t-1}, \end{aligned}$$

where $\varepsilon_{2,t} \sim N(0, 1)$, a_2 and b_2 are taken in accordance with the setting for the idiosyncratic components, for example, if $a_{1i} \sim U[-0.8, 0.8]$, then the same for a_2.

For two factors,

$$\begin{aligned} F_{i,t} &= \phi_i F_{i,t-1} + f_{i,t}, \quad i = 1, 2 \\ f_t &= \begin{pmatrix} f_{1,t} \\ f_{2,t} \end{pmatrix} = \begin{pmatrix} a_{21} & 0 \\ 0 & a_{22} \end{pmatrix} \begin{pmatrix} f_{1,t-1} \\ f_{2,t-1} \end{pmatrix} + \varepsilon_{2,t} + \begin{pmatrix} b_{21} & 0 \\ 0 & b_{22} \end{pmatrix} \varepsilon_{2,t-1}. \end{aligned}$$

where $\varepsilon_{2,t} \sim N(0, \Sigma_{22})$ is unrestricted, i.e. dependency between the two factors is allowed through the covariance matrix. However, the dynamic dependency structure between factors is not allowed, correspondingly $A_{22} = diag(a_{21}, a_{22})$, $B_{22} = diag(b_{21}, b_{22})$, and a_{2i}, b_{2i}, are taken from the same distribution as a_{1j}, b_{1j}.

Considering the unit root can enter both in the common factors and the idiosyncratic components, all of the four possibilities are studied.

I. F1, e1: unit root for all common factors and all idiosyncratic components. $\phi_j = 1$ for all j, $\theta_i = 1$ all i.
II. F0, e1: stationary common factors, unit root for all idiosyncratic components. $\phi_j = 0.95$ for all j, $\theta_i = 1$ for all i.
III. F1, e0: unit root for all common factors, stationary idiosyncratic components. $\phi_j = 1$ for all j, $\theta_i \sim U[0.8, 1]$ for all i.
IV. F0, e0: stationary common factors and idiosyncratic components. $\phi_j = 0.95$ for all j, $\theta_i \sim U[0.8, 1]$ for all i.

In all settings, the factor loading $\lambda_i \sim U[-1, 3]$. All combinations of $T = 25, 50, 100$, and $N = 25, 50, 100$ are considered.

DGP 3: PSU with correlation between idiosyncratic components and common factors

This DGP is to study the robustness of the PSU test under more general correlation structure. Based on the DGP considered in the simulation study in PSU, we make further generations in two aspects. First, correlations between the idiosyncratic components and the common factors are allowed. Second, two factors are considered, and consequently the correlation within factors is also considered. Taking account of the correlation between common factors and idiosyncratic components can either be through the covariance matrix Σ or the dynamic dependency A and B, and we distinguish two DGPs for different extent of the correlation. We denote these two DGPs as Gen 1 and Gen 2.

Gen 1: First we allow for correlation between the factors and the idiosyncratic component only through the covariance matrix Σ. To be specific, the corresponding element Σ_{12} in Σ is not zero, and hence Σ is generated as a whole without restriction. $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are generated together as,

$$\varepsilon_t = \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \sim N(0, \Sigma).$$

In addition, we use a different method to generate Σ by (1). generating an $(N+d) \times (N+d)$ matrix $U \sim U[0,1]$, and (2). $\Sigma = U'U$. This method is different from the one used in Chang (2004) and PSU. We found that this method can generate relatively stronger correlations than their method.

Another distinction from DGP 2 is that dynamic dependencies are allowed within idiosyncratic errors w_t and common factors f_t, i.e. A_{11}, B_{11}, A_{22} and B_{22} are not diagonal, but dynamic independence is still assumed between the two parts.

$$\begin{pmatrix} v_t \\ f_t \end{pmatrix} = \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix} \begin{pmatrix} v_{t-1} \\ f_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} + \begin{pmatrix} B_{11} & 0 \\ 0 & B_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{pmatrix},$$

The ARMA parameters are taken the same way as in PSU[6]. The parameters of the common factor are taken in accordance with the setting for the idiosyncratic errors. To be specific, for one factor, a_2 and b_2 are generated from $U[-0.5, 0.5]$. For two factors, A_{22} and B_{22} are generated according to A_{11} and B_{11}.

Gen 2: In this DGP, we make the generation further by allowing for the dynamic dependency between the factor and the idiosyncratic component. Accordingly, A_{12}, A_{21}, B_{12} and B_{21} are not zero, the whole matrix A and B are generated by the following steps:

(1) Generate an $(N+d) \times (N+d)$ matrix $U \sim U[0,1]$,
(2) Generate $\zeta_1, \ldots, \zeta_{N+d}$, $\zeta_i \sim U[-0.5, 0.5]$, $i = 1, \ldots, N+d$,
(3) Let $Z = diag(\zeta_1, \ldots, \zeta_{N+d})$. Then let $A = UZU'$.

With matrix A and B, f_t and v_t in (9) are generated together as

$$\begin{pmatrix} v_t \\ f_t \end{pmatrix} = A \begin{pmatrix} v_{t-1} \\ f_{t-1} \end{pmatrix} + \varepsilon_t + B\varepsilon_{t-1},$$

where $\varepsilon_t \sim N(0, \Sigma)$ as in Gen 1. Gen 2 includes the cross-sectional dependency to the full extent of the PSU model. σ, A and B are all generated unrestrictedly as shown in the rightmost column in Table 1. The same four settings in DGP 2 are considered for both Gen 1 and Gen 2. We consider the PSU test, BN test and Chang test with all combinations of $T = 25, 50, 100$, and $N = 10, 25, 50, 100$ in the simulation study under this DGP.

---

[6]As in PSU, let

$$A_{11} = \begin{pmatrix} \xi_1 & \xi_1\eta_1 & \xi_1\eta_1^2 & \cdots & \xi_1\eta_1^{N-1} \\ \xi_2\eta_2 & \xi_2 & \xi_2\eta_2 & \cdots & \xi_2\eta_2^{N-2} \\ \vdots & & \ddots & & \vdots \\ \xi_N\eta_N^{N-1} & \xi_N\eta_N^{N-2} & \xi_N\eta_N^{N-3} & \cdots & \xi_N\eta_N \end{pmatrix}$$

where $\xi_i, \eta_i \sim U[-0.5, 0.5]$. The moving-average parameter B_1 is generated as: Let $M = HLH'$ where $H = U(U'U)^{-\frac{1}{2}}$, with U an $N \times N$ matrix of $U[0,1]$ variables, and $L = diag(l_1, \ldots, l_N)$ with $l_1 = 0.1, l_N = 1$ and $l_2, \ldots, l_{N-1} \sim U[0.1, 1]$. Then let $B_1 = 2M - I_N$.

Table 2. Empirical rejection rates under DGP of Chang

Chang's DGP size

		$MA(0.4, 0.8)$				$MA(-0.8, -0.4)$				$ARMA(-0.8, 0.8)$			
N	T	F_{OT}	K_{OT}	τ_p	τ_{gm}	F_{OT}	K_{OT}	τ_p	τ_{gm}	F_{OT}	K_{OT}	τ_p	τ_{gm}
10	25	3.9	4.5	0.2	0.4	14.9	22.6	89.8	86.5	10.0	6.6	41.6	6.2
25	25	5.3	3.7	0	0	19.8	30.0	99.8	99.1	8.7	5.5	62.3	1.8
50	25	16.4	3.2	0	0	32.0	31.3	100	100	30.1	6.0	79.3	0.2
100	25	78.5	9.7	0	0	72.6	21.3	100	100	92.4	21.8	92.3	0.0
10	50	2.2	2.9	1.4	0.7	12.8	19.6	93.8	89.1	4.1	4.6	52.6	8.3
25	50	1.5	1.6	0	0.1	15.7	27.4	99.9	99.5	4.1	3.7	77.2	3.1
50	50	2.3	0.5	0	0	17.7	30.2	100	100	6.3	2.3	93.0	0.7
100	50	12.8	0.1	0	0	35.6	27.4	100	100	28.2	0.9	99.3	0.1
10	100	2.2	2.4	1.5	1.2	8.7	31.8	93.2	88.5	3.4	4.2	56.4	9.1
25	100	1.4	1.5	0.2	0.2	9.4	18.2	99.9	99.4	3.4	3.5	82.3	4.9
50	100	1.5	0.5	0.2	0	10.8	21.6	100	100	3.4	2.1	96.0	2.2
100	100	3.1	0.0	0	0	14.9	20.8	100	100	6.3	0.6	99.8	0.5

Chang's DGP power

		$MA(0.4, 0.8)$				$MA(-0.8, -0.4)$				$ARMA(-0.8, 0.8)$			
N	T	F_{OT}	K_{OT}	τ_p	τ_{gm}	F_{OT}	K_{OT}	τ_p	τ_{gm}	F_{OT}	K_{OT}	τ_p	τ_{gm}
10	25	17.7	33.2	21.3	34.6	64.4	80.3	100	100	26.8	42.2	91.1	51.6
25	25	16.5	42.3	32.2	35.3	78.1	94.6	100	100	28.8	55.0	99.7	53.4
50	25	15.4	39.3	49.6	39.6	73.9	96.3	100	100	26.1	53.6	100	55.2
100	25	42.5	37.5	60.7	47.6	59.3	84.4	100	100	50.6	50.8	100	56.4
10	50	58.0	77.3	84.8	80.4	94.8	98.2	100	100	73.0	87.2	98.9	78.7
25	50	74.0	96.0	99.5	94.8	99.7	100	100	100	89.5	98.8	100	87.5
50	50	65.9	98.4	100	100	100	100	100	100	90.8	99.8	100	92.0
100	50	22.1	93.0	100	100	99.9	100	100	100	70.5	99.4	100	95.5
10	100	98.5	99.6	99.8	95.1	100	100	100	100	99.4	99.9	100	92.4
25	100	100	100	100	99.4	100	100	100	100	100	100	100	96.6
50	100	100	100	100	100	100	100	100	100	100	100	100	98.3
100	100	100	100	100	100	100	100	100	100	100	100	100	99.4

F_{OT} =Chang F test; K_{OT} =Chang K test.

All of the simulation results are based on 2000 simulations and the Warp-Speed bootstrap is used to obtain the bootstrap distribution. Additionally, the PSU tests τ_p and τ_{gm} use block length $b = 1.75T^{1/3}$ as the simulation study in PSU. The Chang test F_{OT} and K_{OT} use the AR order selected using the MAIC information criterion as the simulation study in Chang (2004).

4. Simulation Result

4.1. **Under Chang's DGP.** Table 2 presents results under Chang's DGP. In general, the PSU test has as good size and power properties as the Chang test except when the moving average coefficients are negative.

From Table 2, it can be clearly seen that the small sample performance of the PSU test depends on the ARMA coefficients, especially the moving average coefficients. In terms of size, when the MA parameters are drawn from (-0.8,-0.4), the PSU test has close-to-unity size, both for the group-mean test τ_{gm} and the pooled test τ_p. when the MA parameters are from (0.4,0.8), the PSU test has close-to-zero size. In the case where all the AR and MA parameters are drawn from (-0.8, 0.8), the group-mean test τ_{gm} has reasonable size whereas the pooled test τ_p is severely oversized. Moreover, in this case, the size of τ_p is increasing with the sample size. In contrast, the Chang test is less sensitive to the ARMA coefficients. When the moving average coefficients are negative, the Chang test is also oversized but the sizes are much smaller than

TABLE 3. Empirical rejection rates under DGP of Bai and Ng

Panel A1: Single factor; $I(1)$ common factor(s) and $I(1)$ idiosyncratic errors

		$MA(0.4,0.8)$					$MA(-0.8,-0.4)$					$ARMA(-0.8,0.8)$				
N	T	$ADF_{\hat{F}}$	$P_{\hat{e}}$	P_a	τ_p	τ_{gm}	$ADF_{\hat{F}}$	$P_{\hat{e}}$	P_a	τ_p	τ_{gm}	$ADF_{\hat{F}}$	$P_{\hat{e}}$	P_a	τ_p	τ_{gm}
25	25	7.8	1.5	3.4	0.3	0.3	7.2	5.9	99.2	91.6	82.8	6.7	2.7	4.0	29.4	6.3
50	25	8.0	1.3	3.0	0.2	0.2	8.0	6.0	100.0	92.1	83.2	6.5	2.0	2.5	29.5	4.8
100	25	8.0	1.0	3.7	0.1	0.2	8.1	6.6	100.0	92.7	83.6	6.3	1.5	2.4	31.1	4.3
25	50	5.5	1.2	3.1	0.4	0.6	8.8	14.9	99.4	93.5	85.1	5.5	2.8	2.8	32.3	7.0
50	50	5.2	0.2	1.5	0.0	0.1	9.0	23.3	100.0	94.8	85.7	5.3	2.4	1.6	33.8	5.0
100	50	5.2	0.2	0.9	0.0	0.3	8.9	37.1	100.0	95.4	86.1	5.8	1.9	1.1	34.1	4.0
25	100	4.8	1.0	2.7	1.1	1.2	10.7	27.4	98.1	90.1	77.7	5.2	4.4	3.0	36.2	9.3
50	100	4.6	0.9	1.6	0.5	0.8	10.9	44.2	100.0	92.4	80.4	5.6	3.6	1.6	40.1	6.8
100	100	4.9	0.6	0.7	0.4	0.5	10.9	69.2	100.0	92.6	80.4	5.5	3.7	1.0	40.2	4.9

Panel A2: Multi factors; $I(1)$ common factor(s) and $I(1)$ idiosyncratic errors

		$MA(0.4,0.8)$					$MA(-0.8,-0.4)$					$ARMA(-0.8,0.8)$				
N	T	MQ_c^c	$P_{\hat{e}}$	P_a	τ_p	τ_{gm}	MQ_c^c	$P_{\hat{e}}$	P_a	τ_p	τ_{gm}	MQ_c^c	$P_{\hat{e}}$	P_a	τ_p	τ_{gm}
25	25	0.0	10.1	6.3	0.6	0.9	3.7	6.9	98.4	74.0	67.0	0.1	10.2	7.8	19.1	6.4
50	25	0.0	19.7	8.5	0.5	0.9	3.2	7.9	99.8	75.6	65.2	0.1	16.7	8.5	18.4	4.6
100	25	0.0	34.9	12.2	0.6	0.7	3.7	9.5	99.9	75.4	69.5	0.3	22.8	9.7	18.6	4.7
25	50	0.0	6.0	3.5	0.8	1.3	70.4	15.2	98.7	73.7	66.1	8.2	7.6	4.7	20.3	7.3
50	50	0.0	10.5	2.5	0.6	0.9	55.0	22.0	100.0	76.6	67.7	2.9	9.5	3.9	23.0	7.0
100	50	0.0	17.0	2.8	1.0	1.1	53.2	34.5	100.0	79.5	69.9	2.6	13.2	3.4	22.1	5.1
25	100	0.3	5.3	3.4	1.6	2.0	68.2	22.8	97.1	73.9	64.2	8.5	6.9	3.6	21.4	7.4
50	100	0.4	6.4	1.8	1.3	2.0	68.0	39.5	99.8	74.7	64.2	10.0	7.7	2.0	22.8	7.3
100	100	0.0	12.7	1.3	1.3	2.1	48.3	63.9	100.0	75.8	67.1	3.2	9.0	1.0	24.2	7.5

Panel B1: Single factor; $I(0)$ common factor(s) and $I(1)$ idiosyncratic errors

		$MA(0.4,0.8)$					$MA(-0.8,-0.4)$					$ARMA(-0.8,0.8)$				
N	T	$ADF_{\hat{F}}$	$P_{\hat{e}}$	P_a	τ_p	τ_{gm}	$ADF_{\hat{F}}$	$P_{\hat{e}}$	P_a	τ_p	τ_{gm}	$ADF_{\hat{F}}$	$P_{\hat{e}}$	P_a	τ_p	τ_{gm}
25	25	8.3	0.6	1.1	0.5	0.6	8.4	6.2	99.5	98.1	95.5	7.2	2.4	2.5	37.3	8.0
50	25	7.8	0.5	0.5	0.3	0.3	8.0	6.7	100.0	98.4	95.8	7.2	1.7	0.9	38.6	6.6
100	25	7.8	0.4	0.3	0.2	0.2	8.0	7.3	100.0	99.0	95.9	6.9	0.8	0.5	39.7	5.7
25	50	4.9	1.1	2.0	0.9	1.6	11.5	16.1	99.5	99.9	99.3	6.9	2.4	2.4	47.1	10.0
50	50	4.9	0.2	0.6	0.5	0.3	11.5	25.4	100.0	99.9	99.3	6.9	2.4	0.8	48.0	7.3
100	50	4.6	0.1	0.3	0.2	0.5	11.2	40.6	100.0	100.0	99.4	6.8	1.8	0.3	48.6	5.3
25	100	7.3	1.0	2.2	4.7	6.4	23.6	28.5	98.0	100.0	100.0	8.8	4.8	2.6	58.3	13.8
50	100	7.2	1.0	1.2	4.8	4.6	23.9	46.7	100.0	100.0	100.0	10.3	4.0	1.3	64.0	10.4
100	100	8.0	0.5	0.5	3.9	2.6	24.1	71.8	100.0	100.0	100.0	10.5	4.2	0.5	66.3	7.6

Panel B2: Multi factors; $I(0)$ common factor(s) and $I(1)$ idiosyncratic errors

		$MA(0.4,0.8)$					$MA(-0.8,-0.4)$					$ARMA(-0.8,0.8)$				
N	T	MQ_c^c	$P_{\hat{e}}$	P_a	τ_p	τ_{gm}	MQ_c^c	$P_{\hat{e}}$	P_a	τ_p	τ_{gm}	MQ_c^c	$P_{\hat{e}}$	P_a	τ_p	τ_{gm}
25	25	0.0	11.2	3.0	1.3	1.8	4.4	6.9	99.0	92.8	88.4	0.2	10.0	6.0	28.0	9.7
50	25	0.0	21.5	4.0	1.1	1.6	3.6	8.2	100.0	93.9	89.7	0.1	15.1	5.9	27.0	7.6
100	25	0.0	39.3	6.3	0.8	1.2	4.0	9.8	100.0	95.3	90.7	0.2	22.9	6.2	29.0	7.6
25	50	0.0	6.5	2.4	3.0	4.0	88.2	17.0	99.3	98.7	98.1	12.7	7.7	3.7	38.5	14.9
50	50	0.0	11.7	1.3	2.3	3.0	77.1	25.6	100.0	99.2	99.0	3.9	10.0	2.6	41.6	13.7
100	50	0.0	22.3	0.8	2.6	4.3	75.0	41.3	100.0	99.6	99.3	3.9	12.3	2.0	41.7	11.3
25	100	1.2	5.2	2.7	8.8	10.0	95.1	26.5	98.1	100.0	100.0	16.8	7.4	3.4	53.1	21.2
50	100	1.2	7.8	1.1	9.1	11.2	95.6	45.4	99.9	100.0	100.0	19.1	8.1	1.8	56.7	20.4
100	100	0.0	15.9	0.7	8.0	12.0	84.9	71.9	100.0	100.0	100.0	6.5	8.6	0.7	59.4	22.1

$ADF_{\hat{F}}$ =BN ADF test; $P_{\hat{e}}$ =BN combined p-value test; P_a =BN MP-type test; MQ_c^c=BN common trend test; τ_p =pooled PSU test; τ_{gm} =group-mean PSU test.

the PSU test. This difference might be caused by the different treatment for autocorrelation in two bootstrap methods. In the Chang test, the serial correlation is filtered away to some extent by the sieve bootstrap, while the PSU test resamples the whole block to preserve the autocorrelation. In terms of power, the extremely high power of τ_p in case ARMA(-0.8,0.8) and both τ_{gm} and τ_p in case MA(-0.8,-0.4) are meaningless because of the severe size distortion. Apart from this, the PSU tests have as good power as the Chang test.

Panel C1: Single factor; $I(1)$ common factor and $I(0)$ idiosyncratic errors

		$MA(0.4, 0.8)$					$MA(-0.8, -0.4)$					$ARMA(-0.8, 0.8)$				
N	T	$ADF_{\hat{F}}$	$P_{\hat{e}}$	P_a	τ_p	τ_{gm}	$ADF_{\hat{F}}$	$P_{\hat{e}}$	P_a	τ_p	τ_{gm}	$ADF_{\hat{F}}$	$P_{\hat{e}}$	P_a	τ_p	τ_{gm}
25	25	7.2	26.4	85.8	6.4	9.8	7.6	33.2	100.0	94.0	87.0	7.2	27.8	62.6	43.3	24.3
50	25	6.8	46.1	98.1	6.2	10.9	7.6	53.5	100.0	95.2	87.1	6.9	44.6	70.8	43.7	23.8
100	25	7.1	72.4	100.0	6.9	12.0	7.6	79.8	100.0	95.6	87.8	7.1	68.2	82.4	44.3	27.1
25	50	4.4	91.8	99.0	17.5	20.4	8.8	95.8	100.0	96.3	88.7	6.8	87.2	86.7	49.6	34.8
50	50	4.2	99.7	100.0	19.6	22.4	9.3	99.4	100.0	97.0	89.0	6.3	97.8	91.6	51.6	35.2
100	50	4.4	100.0	100.0	20.9	22.7	8.7	100.0	100.0	97.5	90.2	6.7	99.9	96.6	53.5	36.3
25	100	4.5	100.0	99.9	33.2	30.1	11.1	100.0	100.0	95.4	84.4	6.5	99.8	96.4	59.5	45.4
50	100	4.2	100.0	100.0	35.1	31.8	11.3	100.0	100.0	96.3	85.3	6.7	100.0	97.9	63.1	47.2
100	100	4.3	100.0	100.0	37.5	33.2	11.0	100.0	100.0	97.3	85.5	6.9	100.0	99.6	62.7	48.5

Panel C2: Two factors; $I(1)$ common factors and $I(0)$ idiosyncratic errors

		$MA(0.4, 0.8)$					$MA(-0.8, -0.4)$					$ARMA(-0.8, 0.8)$				
N	T	MQ_c^c	$P_{\hat{e}}$	P_a	τ_p	τ_{gm}	MQ_c^c	$P_{\hat{e}}$	P_a	τ_p	τ_{gm}	MQ_c^c	$P_{\hat{e}}$	P_a	τ_p	τ_{gm}
25	25	0.0	15.6	71.0	2.9	5.0	3.6	29.8	99.7	82.6	74.5	0.1	22.7	58.2	29.3	18.2
50	25	0.0	26.1	90.7	3.0	5.5	3.5	47.9	100.0	81.7	72.5	0.1	35.7	68.2	31.5	18.9
100	25	0.0	44.7	99.2	4.5	6.4	3.7	71.7	100.0	80.7	72.6	0.2	55.3	79.5	32.0	19.1
25	50	0.1	80.0	97.5	9.6	10.6	71.4	90.0	100.0	81.1	73.4	11.5	77.5	84.6	36.3	25.8
50	50	0.0	96.7	99.8	9.9	11.1	58.6	97.8	100.0	82.8	73.5	4.6	93.2	91.4	37.4	27.5
100	50	0.0	100.0	100.0	10.7	10.2	54.0	99.4	100.0	84.1	73.9	4.8	99.0	96.5	37.9	26.0
25	100	0.6	99.8	99.9	18.0	16.9	68.4	99.4	100.0	81.1	68.4	13.0	97.6	95.3	40.4	31.9
50	100	0.4	100.0	100.0	19.0	17.6	67.8	99.9	100.0	82.0	69.0	13.3	99.8	98.3	43.5	34.1
100	100	0.1	100.0	100.0	21.5	19.4	51.8	100.0	100.0	81.5	68.9	5.9	99.9	99.1	43.1	36.2

Panel D1: Single factor; $I(0)$ common factor and $I(0)$ idiosyncratic errors

		$MA(0.4, 0.8)$					$MA(-0.8, -0.4)$					$ARMA(-0.8, 0.8)$				
N	T	$ADF_{\hat{F}}$	$P_{\hat{e}}$	P_a	τ_p	τ_{gm}	$ADF_{\hat{F}}$	$P_{\hat{e}}$	P_a	τ_p	τ_{gm}	$ADF_{\hat{F}}$	$P_{\hat{e}}$	P_a	τ_p	τ_{gm}
25	25	7.5	28.8	86.0	11.2	18.8	7.3	35.3	100.0	99.5	97.8	7.7	30.1	62.1	56.5	34.4
50	25	7.1	48.4	97.7	11.5	19.1	7.8	57.7	100.0	99.8	97.7	8.1	49.0	69.3	56.9	34.2
100	25	7.1	75.8	100.0	12.8	23.2	7.7	85.3	100.0	99.9	98.1	7.9	75.3	81.1	58.4	36.6
25	50	4.6	95.9	99.1	42.1	52.6	12.9	97.9	100.0	100.0	99.8	7.3	92.3	86.8	72.2	54.8
50	50	5.0	100.0	100.0	46.1	54.7	12.9	99.9	100.0	100.0	99.8	7.9	99.4	91.7	74.6	56.5
100	50	4.9	100.0	100.0	51.0	56.0	12.6	100.0	100.0	100.0	99.8	8.1	100.0	96.4	77.0	59.2
25	100	7.4	100.0	100.0	92.3	90.7	26.8	100.0	100.0	100.0	100.0	11.7	99.9	96.5	88.9	80.8
50	100	7.8	100.0	100.0	94.0	92.5	26.7	100.0	100.0	100.0	100.0	11.8	100.0	97.8	92.0	85.8
100	100	7.7	100.0	100.0	95.4	92.9	26.4	100.0	100.0	100.0	100.0	11.5	100.0	99.5	91.9	86.3

Panel D2: Two factors; $I(0)$ common factors and $I(0)$ idiosyncratic errors

		$MA(0.4, 0.8)$					$MA(-0.8, -0.4)$					$ARMA(-0.8, 0.8)$				
N	T	MQ_c^c	$P_{\hat{e}}$	P_a	τ_p	τ_{gm}	MQ_c^c	$P_{\hat{e}}$	P_a	τ_p	τ_{gm}	MQ_c^c	$P_{\hat{e}}$	P_a	τ_p	τ_{gm}
10	25	0.0	12.2	50.5	8.5	9.9	8.7	20.0	99.1	95.6	93.3	0.5	16.5	50.8	39.7	28.5
25	25	0.0	17.2	73.1	7.0	9.5	3.6	34.8	100.0	97.5	94.9	0.2	26.3	59.2	42.7	28.0
50	25	0.0	31.3	92.0	7.1	10.2	3.9	54.7	100.0	97.2	93.8	0.1	40.7	68.0	45.5	30.9
100	25	0.0	50.0	99.4	9.5	12.0	3.8	81.1	100.0	97.1	93.6	0.2	62.9	78.4	46.5	29.2
10	50	0.2	51.0	87.7	25.3	29.3	89.7	66.8	100.0	99.6	99.6	16.1	54.2	80.6	59.9	46.1
25	50	0.1	90.5	98.4	32.9	33.0	89.9	97.3	100.0	99.8	99.6	17.7	87.9	87.2	66.3	50.8
50	50	0.0	99.5	100.0	32.7	35.0	80.9	100.0	100.0	99.9	99.5	6.5	98.0	92.0	67.7	56.2
100	50	0.0	100.0	100.0	33.7	34.1	77.7	100.0	100.0	99.9	99.6	7.2	100.0	96.4	67.8	55.5
10	100	1.4	95.4	97.0	71.5	70.3	96.7	98.2	100.0	100.0	100.0	21.9	93.3	93.3	82.3	73.7
25	100	1.9	100.0	100.0	78.1	77.8	96.7	100.0	100.0	100.0	100.0	24.6	99.8	96.3	85.5	76.9
50	100	2.4	100.0	100.0	80.4	77.8	96.5	100.0	100.0	100.0	100.0	25.7	100.0	98.5	88.0	79.9
100	100	0.2	100.0	100.0	83.6	82.0	88.4	100.0	100.0	100.0	100.0	13.8	100.0	99.3	88.8	82.0

$ADF_{\hat{F}}$ =BN ADF test; $P_{\hat{e}}$ =BN combined p-value test; P_a =BN MP-type test; MQ_c^c=BN common trend test; τ_p =pooled PSU test; τ_{gm} =group-mean PSU test.

Another aspect regarding the Chang test is worth mentioning. The F test has large size distortions in all cases when $N > T$, and the size distortion becomes larger as the ratio N/T grows. For instance, in column MA(0.4,0.8), the size of F test is 0.164 when $T = 25, N = 50$, 0.785 when $T = 25, N = 100$. Similar patterns can also be detected from column ARMA(-0.8,0.8) in the same panel.

4.2. **Under BN's DGP.** Table 3 gives results under BN's DGP. Apart from the large size distortion for large negative MA parameters, the PSU test has generally good small sample

Table 4. Empirical rejection rates under DGP of PSU, one factor and Gen 1.

		$F: I(1),\, e: I(1)$							$F: I(0),\, e: I(1)$						
N	T	$ADF_{\hat{F}}$	P_e	P_a	τ_{gm}	τ_p	F_{OT}	K_{OT}	$ADF_{\hat{F}}$	P_e	P_a	τ_{gm}	τ_p	F_{OT}	K_{OT}
10	25	7.3	4.3	13.7	5.2	3.5	9.7	6.7	7.8	4.3	13.0	11.0	8.0	8.4	8.9
25	25	7.3	5.3	15.7	6.4	2.4	26.3	10.5	7.6	5.7	15.8	9.2	6.4	21.4	12.0
50	25	7.2	7.0	21.7	5.8	2.2	63.0	17.2	7.9	7.3	22.8	10.9	9.0	55.2	15.8
100	25	6.9	10.6	29.7	5.8	1.4	95.6	12.6	7.4	11.2	30.1	11.7	8.2	95.1	12.8
10	50	5.7	5.8	14.6	8.2	5.5	5.4	5.0	6.3	5.4	14.4	17.4	13.6	5.1	7.0
25	50	5.7	6.8	16.4	7.6	4.6	10.2	4.6	6.8	7.1	16.9	16.5	11.6	7.8	7.2
50	50	5.6	9.9	21.2	8.1	3.0	23.4	7.0	6.8	10.5	22.7	18.0	14.6	16.0	7.5
100	50	5.6	14.4	29.1	8.5	2.6	55.7	7.7	7.0	15.6	29.8	17.6	15.2	42.9	8.9
10	100	5.2	5.2	13.9	6.4	4.4	5.0	4.2	8.4	4.9	13,2	23.2	16.5	8.7	10.4
25	100	5.9	8.6	16.8	6.7	5.0	7.7	4.2	8.6	9.0	17.6	24.6	17.6	9.1	11.5
50	100	5.6	13.1	24.3	8.3	3.5	13.9	6.2	9.0	13.0	24.4	28.7	20.0	12.1	10.6
100	100	5.5	18.4	31.3	8.6	3.0	28.6	8.2	8.9	18.8	32.8	26.2	19.8	20.8	10.5

		$F: I(1),\, e: I(0)$							$F: I(0),\, e: I(0)$						
N	T	$ADF_{\hat{F}}$	P_e	P_a	τ_{gm}	τ_p	F_{OT}	K_{OT}	$ADF_{\hat{F}}$	P_e	P_a	τ_{gm}	τ_p	F_{OT}	K_{OT}
10	25	7.1	18.0	57.0	17.3	14.5	10.7	13.5	7.1	21.0	65.0	29.6	29.2	8.3	14.0
25	25	6.6	30.2	68.0	17.8	14.2	26.2	18.9	7.4	34.9	76.2	34.9	36.0	20.8	24.5
50	25	6.8	39.7	74.6	20.2	16.6	50.1	15.9	7.5	47.3	81.8	37.3	38.0	42.6	14.0
100	25	7.1	51.3	80.2	18.7	15.5	85.8	9.5	7.7	60.4	86.8	40.2	38.4	82.2	6.8
10	50	6.9	41.5	70.6	26.2	22.0	10.8	16.1	7.6	59.6	84.2	58.6	60.3	16.3	27.8
25	50	6.9	59.0	78.1	26.8	22.6	13.7	20.0	7.5	79.3	88.9	67.0	68.2	17.8	36.0
50	50	6.9	69.0	82.5	29.6	24.6	20.5	21.5	7.4	88.4	92.5	68.3	71.8	20.7	35.0
100	50	6.9	75.6	85.7	28.6	24.8	41.8	9.2	7.9	92.9	94.6	70.6	71.8	33.2	9.9
10	100	9.4	63.8	76.1	35.6	28.2	18.8	27.9	13.9	89.9	92.3	86.3	85.2	36.8	57.2
25	100	9.0	77.5	82.5	37.0	29.0	19.3	34.6	14.1	97.4	95.7	92.2	91.6	41.8	68.2
50	100	9.3	83.9	86.1	41.0	31.6	23.1	34.4	13.9	99.3	97.4	93.9	92.5	41.8	75.2
100	100	9.0	87.1	88.6	40.6	32.4	27.4	34.4	14.1	99.9	98.7	94.5	93.0	54.6	73.8

Table 5. Empirical rejection rates under DGP of PSU, two factors and Gen 1.

		$F: I(1),\, e: I(1)$							$F: I(0),\, e: I(1)$						
N	T	$ADF_{\hat{F}}$	P_e	P_a	τ_{gm}	τ_p	F_{OT}	K_{OT}	$ADF_{\hat{F}}$	P_e	P_a	τ_{gm}	τ_p	F_{OT}	K_{OT}
10	25	0.6	3.8	12.1	11.9	8.8	12.4	8.5	0.3	3.6	10.5	18.8	13.6	12.2	12.5
25	25	0.0	4.1	10.9	13.6	9.8	29.8	14.3	0.0	3.2	9.3	21.4	17.0	25.8	17.2
50	25	0.0	5.5	13.9	12.1	9.8	64.0	19.5	0.1	5.3	11.6	20.2	16.4	56.8	17.4
100	25	0.0	7.0	19.9	13.6	10.4	92.7	15.3	0.0	6.8	15.7	21.0	16.7	89.9	13.6
10	25	5.9	4.5	12.5	16.4	10.2	8.3	8.0	10.3	3.7	11.1	29.9	21.1	9.8	12.8
25	50	5.6	4.3	11.3	17.0	10.4	15.8	10.5	9.2	3.9	9.0	31.6	23.4	15.0	16.4
50	50	0.6	6.6	13.8	18.8	10.8	32.2	13.6	1.5	5.5	10.9	34.0	23.5	24.2	18.4
100	50	0.7	9.6	18.7	17.8	11.4	62.3	13.5	1.4	7.7	15.6	33.4	24.0	48.4	11.6
10	25	6.2	5.5	12.9	19.0	9.6	7.4	7.4	14.1	4.1	9.5	38.3	30.4	13.5	17.6
25	100	6.0	5.6	11.9	20.0	11.0	14.4	10.9	13.5	4.3	9.9	45.1	35.5	17.2	24.1
50	100	6.0	7.5	15.9	20.7	11.5	24.0	12.6	14.6	6.7	12.3	47.9	38.0	24.6	26.3
100	100	2.1	11.5	20.9	22.4	11.7	36.9	13.0	5.8	8.9	18.4	48.8	37.0	32.5	20.2

		$F: I(1),\, e: I(0)$							$F: I(0),\, e: I(0)$						
N	T	$ADF_{\hat{F}}$	P_e	P_a	τ_{gm}	τ_p	F_{OT}	K_{OT}	$ADF_{\hat{F}}$	P_e	P_a	τ_{gm}	τ_p	F_{OT}	K_{OT}
10	25	0.4	12.7	44.6	22	16	10.8	12.2	0.3	17.0	54.8	35.2	24.8	10.4	14.4
25	25	0.0	21.3	56.9	20.8	15	25.4	17	0.1	27.5	68.4	33.6	25.8	20.8	19.2
50	25	0.0	31.9	63.4	23	16	57.9	18.6	0.0	41.9	75.1	35.9	24.7	48.6	20.1
100	25	0.0	41.7	70.8	21.2	16.4	87.5	10.4	0.0	55.6	82.8	33.6	25	83.4	7.6
10	50	9.5	28.5	54.6	31.8	20.4	10	12.7	13.6	47.4	78.3	59.2	41.9	12.6	22.6
25	50	8.7	44.7	64.0	32.6	20.7	14.4	16	12.7	75.0	87.5	61.2	43.1	16.8	29.4
50	50	1.2	54.8	69.6	31.4	19.7	24.6	16.8	1.6	87.4	90.4	60.3	40	20.4	27.6
100	50	1.3	62.1	73.7	32.2	20.8	48.2	11.8	1.7	93.8	94.0	62.7	43	36.2	7.2
10	25	9.2	46.5	61.6	39.4	24.4	12.8	19.8	22.7	85.9	90.4	80.8	66.4	32.4	50
25	100	9.3	60.2	66.4	42	24.4	13.7	20.4	21.9	97.6	94.9	85.2	68.8	29.3	56.4
50	100	9.2	68.9	72.2	42	24.8	17.3	22.8	22.1	99.3	97.2	87	68.4	33.2	58
100	100	2.9	73.5	75.4	43.1	25	26.6	21.2	8.6	100.0	98.1	88.2	70.2	37.6	44.4

properties in cases F1, e1, F0, e1 and F0, e0. However, in case F1, e0, the PSU test has size distortions and the size increases with the sample size.

TABLE 6. Empirical rejection rates under DGP of PSU, single factor and Gen 2.

		$F: I(1), e: I(1)$							$F: I(0), e: I(1)$						
N	T	$ADF_{\hat{F}}$	P_e	P_a	τ_{gm}	τ_p	F_{OT}	K_{OT}	$ADF_{\hat{F}}$	P_e	P_a	τ_{gm}	τ_p	F_{OT}	K_{OT}
10	25	6.6	12.3	33.7	7.4	5.8	20.0	11.4	6.6	10.7	27.6	11.2	9.4	31.2	20.2
25	25	6.4	16.9	41.5	7.2	5.2	47.5	11.7	7.1	15.2	32.9	9.4	8.2	53.4	17.3
50	25	6.6	21.3	46.0	7.7	6.0	73.5	7.2	7.0	19.2	37.4	10.1	8.6	80.0	14.0
100	25	6.7	25.0	49.7	7.8	6.4	93.5	6.3	7.1	22.5	41.3	9.9	8.7	97.4	14.3
10	25	6.5	15.4	32.5	10.2	7.4	16.3	11.1	8.0	12.1	22.3	12.3	9.9	28.2	21.0
25	50	6.0	20.1	40.6	10.2	7.0	29.2	8.2	7.0	15.7	29.2	13.7	11.1	43.6	24.8
50	50	6.0	23.5	45.6	10.1	7.3	42.5	6.7	6.9	19.3	33.5	13.4	10.4	57.2	18.4
100	50	6.0	27.7	49.9	8.9	7.0	67.1	6.2	6.9	22.5	37.3	12.1	9.8	76.6	17.8
10	25	6.2	15.2	33.7	11.7	6.5	12.5	10.2	10.2	11.9	21.9	17.2	14.0	31.2	25.8
25	100	6.5	21.3	41.8	11.3	6.5	23.5	9.7	9.7	17.3	27.7	18.0	15.6	44.6	24.6
50	100	6.4	26.3	46.6	10.8	6.4	33.6	6.3	10.4	19.9	32.4	19.2	16.2	53.2	22.5
100	100	6.3	30.4	51.9	11.6	7.0	45.3	5.7	10.5	23.6	35.3	18.2	15.6	62.9	20.2
		$F: I(1), e: I(0)$							$F: I(0), e: I(0)$						
N	T	$ADF_{\hat{F}}$	P_e	P_a	τ_{gm}	τ_p	F_{OT}	K_{OT}	$ADF_{\hat{F}}$	P_e	P_a	τ_{gm}	τ_p	F_{OT}	K_{OT}
10	25	8.4	18.9	39.2	12.8	10.7	33.8	22.8	8.1	23.6	49.8	17.4	14.0	31.2	27.0
25	25	8.1	27.7	46.2	14.8	11.4	58.8	19.6	8.2	33.5	55.9	18.4	13.4	54.4	20.2
50	25	8.1	33.5	50.6	12.6	10.4	83.1	19.6	8.4	39.8	60.6	19.2	15.0	78.4	15.1
100	25	8.3	37.7	54.4	13.9	11.2	98.3	18.1	8.4	45.4	65.1	19.8	15.4	97.1	14.2
10	25	7.6	26.3	41.3	20.0	14.8	35.9	31.4	9.6	43.5	61.1	33.1	26.0	31.2	28.6
25	50	7.5	36.0	48.0	19.2	14.4	55.8	30.0	9.6	55.5	67.5	38.3	27.3	48.3	28.8
50	50	7.9	41.4	52.4	18.8	14.1	71.2	24.3	9.6	63.0	72.4	37.4	27.0	63.6	18.4
100	50	7.6	45.3	55.4	18.4	12.6	86.0	19.4	9.4	67.9	75.7	33.2	23.5	80.6	1.34
10	25	8.9	34.7	41.9	26.2	18.0	45.0	43.1	15.2	71.1	76.4	61.9	47.1	47.8	55.2
25	100	8.8	42.9	49.3	24.4	16.9	64.2	39.1	15.4	81.8	80.2	62.4	42.0	64.5	40.0
50	100	8.9	47.7	53.0	23.4	17.0	76.2	29.3	15.2	86.8	84.2	70.7	48.0	73.5	17.2
100	100	8.8	52.2	57.0	25.0	18.1	84.2	24.6	15.1	89.9	85.7	68.9	45.1	81.6	14.7

TABLE 7. Empirical rejection rates under DGP of PSU, two factors and Gen 2.

		$F: I(1), e: I(1)$							$F: I(0), e: I(1)$						
N	T	$ADF_{\hat{F}}$	P_e	P_a	τ_{gm}	τ_p	F_{OT}	K_{OT}	$ADF_{\hat{F}}$	P_e	P_a	τ_{gm}	τ_p	F_{OT}	K_{OT}
10	25	0.8	9.8	29.8	9.2	7.8	20.6	13.6	0.8	9.6	19.2	13	12.2	31	22.3
25	25	0.0	14.9	38.4	7.8	7.2	43.6	7.4	0.0	14.0	24.1	11.8	11.4	55.6	13.7
50	25	0.0	18.8	44.5	8.1	7.2	77	7.8	0.1	16.6	28.9	11.6	10.5	77.3	12.7
100	25	0.1	23.2	48.7	8.3	7.2	94	7.4	0.1	19.6	31.6	12.3	10.5	95.3	11.5
10	25	6.7	12.6	29.0	9.8	7.2	16.6	11	10.0	8.4	14.6	17.6	16.2	30.9	23.1
25	50	6.0	18.4	37.4	11.3	8.6	29.9	6.6	9.6	12.6	19.2	18.8	18	44.7	15.1
50	50	0.7	23.3	42.7	10.2	8.2	44	6.4	1.3	16.2	23.7	17.4	16.6	57.8	14.5
100	50	0.8	28.2	47.3	10.2	6.7	65	6.5	1.2	20.2	27.9	16.4	14.6	74.4	14.9
10	25	6.1	15.7	30.8	12.6	8.6	15.2	11.7	14.8	10.1	13.3	23.2	21.5	35.4	30.7
25	100	5.5	22.8	37.7	11.8	8	26.3	9.6	14.6	14.4	18.2	23.8	22.7	48.7	25.5
50	100	5.5	27.1	42.8	12.2	8.4	35	6.3	14.5	17.4	22.7	25.1	23.4	57.8	18.9
100	100	0.9	31.9	47.5	13	8	3.7	7.2	2.9	20.2	26.4	25.2	23	66.9	18.3
		$F: I(1), e: I(0)$							$F: I(0), e: I(0)$						
N	T	$ADF_{\hat{F}}$	P_e	P_a	τ_{gm}	τ_p	F_{OT}	K_{OT}	$ADF_{\hat{F}}$	P_e	P_a	τ_{gm}	τ_p	F_{OT}	K_{OT}
10	25	0.6	13.8	24.9	10.2	9	38.1	22.8	1.2	17.4	35.3	16.7	13.2	29.7	20
25	25	0.1	20.3	30.5	10.9	8.8	62.7	17.6	0.0	25.4	42.2	17.8	14.4	54.8	13
50	25	0.0	25.5	34.1	10.6	8	84.8	17.1	0.0	32.8	45.9	16.5	13.8	78.5	12.8
100	25	0.0	29.7	37.2	9.6	7.8	97.4	14.2	0.0	38.2	49.8	14.2	11.6	95.9	10.2
10	25	7.7	17.0	23.9	13.1	10.2	34	25.4	12.4	34.7	49.1	29.8	22.6	24.8	24.4
25	50	7.6	25.4	30.4	13.3	9.9	56.6	23.8	12.7	48.4	55.1	31.4	23	43.4	13.6
50	50	0.9	30.1	35.6	12.9	9.8	70.6	19.6	1.1	54.8	59.2	28	21.1	57.4	11.2
100	50	0.8	34.1	38.7	14	10.3	86.9	15.8	1.0	60.5	62.2	31.1	22.8	77.2	10.8
10	25	9.7	20.9	25.8	18	12.3	44.1	32.1	21.8	61.2	64.2	55.7	45.6	41.9	45.4
25	100	9.3	28.8	31.8	17.6	12.4	60.2	34.2	21.1	74.1	68.1	55.9	43.3	57.2	19.6
50	100	9.2	32.8	36.4	17.8	12.3	71.2	24.7	21.7	80.7	71.7	60.6	47.8	66	11.5
100	100	1.6	36.9	40.2	16.5	11.4	78.8	21.1	6.5	83.9	74.3	59.2	46.3	74	9.6

Some more specific findings are as follows. First, comparing the single factor with two factors, PSU has higher size with two factors in cases F1, e1 and F0 e1, and lower power in case F0, e0. In case F1, e0, however, the serious size distortion is improved with an increased number of factors, as shown in panel C1 and C2. Second, there is no obvious difference between the tests τ_p and τ_{gm}, except in the case ARMA(-0.8,0.8), where τ_{gm} always has smaller rejection probability than τ_p. Especially in case F1, e1, the size of τ_{gm} is close to nominal size. Third, the BN test is also influenced by the negative MA parameters but to a different extent. The MP-type test is mostly influenced while the combined p value test and the test for common factors are less influenced. At last, for idiosyncratic errors the tests for the common factors, as in Gengenbach et al. (2010), have low rejection probabilities regardless of if the factor part is $I(0)$ or $I(1)$, both for one factor and two factors. In this regard, the BN test is not a good choice to investigate nonstationarity of a whole panel.

4.3. **Under PSU's DGP with general cross-sectional dependency.** In general, PSU has significant robustness against the complicated cross-sectional dependence compared to other tests especially under cases F1, e1 and F0, e0, i.e. the common factors and idiosyncratic errors have identical integrated order. However, for cases F1, e0 and F0, e1 the PSU tests become oversized. More details are listed as follows:

1) Table 4 shows results under DGP Gen 1 with single factor. About the PSU tests, τ_p has better size properties than τ_{gm} in general. Specifically, τ_p has a lower size under the case F1, e1 and smaller size distortion under the other two cases for null hypothesis. The power of τ_p and τ_{gm} are similar. For the Chang tests, they have serious size distortion under the case F1, e1, and the size distortion becomes more serious as N increases. Even though they have a lower size than PSU tests under cases F1, e0 and F0, e1, the power is much lower than PSU tests, as shown in case F0, e0. Regarding the BN tests, the $ADF_{\hat{F}}$ test on the common factor part, as in DGP 2, always has small rejection probability, and all tests on the idiosyncratic errors have serious size distortion, which becomes more serious with the increase of N. Another interesting point for BN tests is that if we fix $N = 10$, we find that the P_e test has good size properties and acceptable power when T is large. In other words, the P_e test has consistently good performance in the small N and large T case even though the common factors correlate with the idiosyncratic errors and the idiosyncratic errors are not independent.

2) Results under DGP Gen 1 with two factors are shown in Table 5. The BN tests are not influenced by increasing the number of common factors, and the interesting property of the P_e test is still kept. The performance of all the tests based on the bootstrap approach will be influenced negatively. Comparing Table 4 and 5 we can see that for both PSU and Chang tests, the size increases and power decreases.

3) Results under DGP Gen 2 are given in Table 6 and 7. In this case, the PSU tests are still robust and all the properties are kept. Moreover, the severe size distortion in cases F1, e0 and F0, e1 even become smaller. As a contrast, the Chang tests F_{OT}, K_{OT} and the BN tests $P_{\hat{e}}$ and P_a get more serious size distortion and lower power in all cases. In addition, the interesting property of the BN test P_e disappeared. Finally, the effects of increasing the number of factors are similar to the results under DGP Gen1.

5. Conclusion and Further study

We have investigated the robustness of the PSU test from specific to general DGPs with the comparison with the Chang test and the BN test. The main conclusions are:

1) Under the specific DGPs of Chang (2004) and Bai and Ng (2004), the PSU test, especially τ_{gm} has generally as good size and power properties as the Chang and BN tests, except in the case when negative moving average coefficients are present. In the case with negative moving average coefficients, both τ_{gm} and τ_p have extreme size distortions.

2) Under the DGP of PSU with general cross-sectional dependency structure, the PSU test

exhibits robustness with good size and power properties in the cases F1, e1 and F0, e0, i.e. the idiosyncratic and the factor parts have the same integrated orders, and the performance is better than the other two tests.

3) In the cases F1, e0 and F0, e1, i.e. only one part, either idiosyncratic or common factor, is $I(1)$ and the other part is $I(0)$, the PSU test is oversized. This problem is more severe in the F1, e0 case.

4) The group-mean test τ_{gm} is more robust than the pooled test τ_p. In both the DGP of Chang (2004) and Bai and Ng (2004), when the ARMA coefficients are randomly chosen from $U[-0.8, 0.8]$, τ_p exhibits severe size distortion.

The PSU test is an attractive and encouraging method because of its flexibility in handling the cross-sectional dependency. In practice, however, directly using the PSU test has obvious limitations as the PSU test is restricted to the model without deterministic terms, and dealing with deterministic terms is not trivial in the PSU test. Since many economic variables exhibit non-zero mean or deterministic trend behaviors, the PSU test will be more applicable if deterministic terms can be taken into account. This will be reported in further study by the authors.

References

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

Bai, J., Ng, S. (2004). A Panic Attack on Unit Roots and Cointegration, *Econometrica* **72**: 1127–177.

Bai, J., Ng, S. (2010). Panel unit root tests with cross-section dependence: a further investigation, *Econometric theory* **26**: 1088–1114.

Chang, Y. (2004). Bootstrap unit root tests in panels with cross-sectional dependency, *Journal of Econometrics* **120**: 263–293.

Gengenbach, C., Palm, F.C., Urbain, J.P. (2010). Panel unit root tests in the presence of cross sectional dependencies: comparison and implications for modelling, *Econometric Reviews* **29(2)**: 111–145.

Im, K.S., Pesaran, M.H., Shin, Y. (2003). Testing for unit roots in heterogeneous panels, *Journal of Econometrics* **115**: 53–74.

Levin, A., Lin, C.F., Chu, C.S.J. (2010). Unit root tests in panel data: asymptotic and finite-sample properties, *Journal of Econometrics* **108**: 1–24.

Moon, R., Perron, P. (2004). Testing for Unit Root in Panels with Dynamic Factors, *Journal of Econometrics* **122**: 81–126.

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

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

Palm, F.C., Smeekes, S., Urbain, J.P. (2011). Cross-sectional dependence robust block bootstrap panel unit root tests, *Journal of Econometrics* **163**: 85–104.

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

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

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

Politis, D.N., Romano, J.P., Wolf, M. (1999). Subsampling. Spring-Verlag, New York.

Sargan, J.D., Bhargava A. (1983). Testing for residuals from least squares regression being generated by Gaussian random walk, *Econometrica* **51**: 153–174.

Schwert, G.W. (1989). Testing for unit roots: a Monte Carlo investigation, *Journal of Business and Economic Statistics* **7**: 147–159.

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