Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2011, Article ID 617652, 17 pages
doi:10.1155/2011/617652

Research Article

Panel Unit Root Tests by Combining Dependent *P* Values: A Comparative Study

Xuguang Sheng[1] and Jingyun Yang[2]

[1] *Department of Economics, American University, Washington, DC 20016, USA*
[2] *Department of Biostatistics and Epidemiology, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104, USA*

Correspondence should be addressed to Xuguang Sheng, sheng@american.edu and Jingyun Yang, jingyuny@gmail.com

Received 27 June 2011; Accepted 25 August 2011

Academic Editor: Mike Tsionas

We conduct a systematic comparison of the performance of four commonly used *P* value combination methods applied to panel unit root tests: the original Fisher test, the modified inverse normal method, Simes test, and the modified truncated product method (TPM). Our simulation results show that under cross-section dependence the original Fisher test is severely oversized, but the other three tests exhibit good size properties. Simes test is powerful when the total evidence against the joint null hypothesis is concentrated in one or very few of the tests being combined, but the modified inverse normal method and the modified TPM have good performance when evidence against the joint null is spread among more than a small fraction of the panel units. These differences are further illustrated through one empirical example on testing purchasing power parity using a panel of OECD quarterly real exchange rates.

1. Introduction

Combining significance tests, or *P* values, has been a source of considerable research in statistics since Tippett [1] and Fisher [2]. (For a systematic comparison of methods for combining *P* values from independent tests, see the studies by Hedges and Olkin [3] and Loughin [4].) Despite the burgeoning statistical literature on combining *P* values, these techniques have not been used much in panel unit root tests until recently. Maddala and Wu [5] and Choi [6] are among the first who attempted to test unit root in panels by combining independent *P* values. More recent contributions include those by Demetrescu et al. [7], Hanck [8], and Sheng and Yang [9]. Combining *P* values has several advantages over combination of test statistics in that (i) it allows different specifications, such as different deterministic terms and lag orders, for each panel unit, (ii) it does not require a panel to be balanced, and (iii) observed P values derived from continuous test statistics have a uniform distribution under the null hypothesis regardless of the test statistic or distribution from which they arise, and thus it can be carried out for any unit root test derived.

While the formulation of the joint null hypothesis (H_0: all of the time series in the panel are nonstationary) is relatively uncontroversial, the specification of the alternative hypothesis critically depends on what assumption one makes about the nature of the heterogeneity of the panel. (Recent contributions include O'Connell [10], Phillips and Sul [11], Bai and Ng [12], Chang [13], Moon and Perron [14] and Pesaran [15].) The problem of selecting a test is complicated by the fact that there are many different ways in which H_0 can be false. In general, we cannot expect one test to be sensitive to all possible alternatives, so that no single P value combination method is uniformly the best. The goal of this paper is to make a detailed comparison, via both simulations and empirical examples, of some commonly used P value combination methods, and to provide specific recommendation regarding their use in panel unit root tests.

The plan of the paper is as follows. Section 2 briefly reviews the methods of combining P values. Small sample performance of these methods is investigated in Section 3 using Monte Carlo simulations. Section 4 provides the empirical applications, and Section 5 concludes the paper.

2. *P* Value Combination Methods

Consider the model

$$y_{it} = (1-\alpha_i)\mu_i + \alpha_i y_{i,t-1} + \epsilon_{it}, \quad i = 1,\ldots,N;\ t = 1,\ldots,T. \tag{2.1}$$

Heterogeneity in both the intercept and the slope is allowed in (2.1). This specification is commonly used in the literature, see the work of Breitung and Pesaran [16] for a recent review. Equation (2.1) can be rewritten as

$$\Delta y_{it} = -\phi_i\mu_i + \phi_i y_{i,t-1} + \epsilon_{it}, \tag{2.2}$$

where $\Delta y_{it} = y_{it} - y_{i,t-1}$ and $\phi_i = \alpha_i - 1$.

The null hypothesis is

$$H_0 : \phi_1 = \phi_2 = \cdots = \phi_N = 0, \tag{2.3}$$

and the alternative hypothesis is

$$H_1 : \phi_1 < 0, \phi_2 < 0, \ldots, \phi_{N_0} < 0, \quad N_0 \le N. \tag{2.4}$$

Let S_{i,T_i} be a test statistic for the ith unit of the panel in (2.2), and let the corresponding P value be defined as $p_i = F(S_{i,T_i})$, where $F(\cdot)$ denotes the cumulative distribution function (c.d.f.) of S_{i,T_i}. We assume that, under H_0, S_{i,T_i} has a continuous distribution function. This assumption is a regularity condition that ensures a uniform distribution of the P

values, regardless of the test statistic or distribution from which they arise. Thus, P value combinations are nonparametric in the sense that they do not depend on the parametric form of the data. The nonparametric nature of combined P values gives them great flexibility in applications.

In the rest of this section, we briefly review the P value combination methods in the context of panel unit root tests. The first test, proposed by Fisher [2], is defined as

$$P = -2\sum_{i=1}^{N} \ln(p_i), \tag{2.5}$$

which has an χ^2 distribution with $2N$ degrees of freedom under the assumption of cross-section independence of the P values. Maddala and Wu [5] introduced this method to the panel unit root tests, and Choi [6] modified it to the case of infinite N.

Inverse normal method, attributed to Stouffer et al. [17], is another often used method defined as

$$Z = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Phi^{-1}(p_i), \tag{2.6}$$

where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution. Under H_0, $Z \sim N(0,1)$. Choi [6] first applied this method to the panel unit root tests assuming cross-section independence among the panel units. To account for cross-section dependence, Hartung [18] developed a modified inverse normal method by assuming a constant correlation across the probits t_i,

$$\operatorname{cov}(t_i, t_j) = \rho, \quad \text{for } i \neq j,\ i, j = 1, \ldots, N, \tag{2.7}$$

where $t_i = \Phi^{-1}(p_i)$. He proposed to estimate ρ in finite samples by

$$\widehat{\rho}^{\star} = \max\left(-\frac{1}{N-1}, \widehat{\rho}\right), \tag{2.8}$$

where $\widehat{\rho} = 1 - (1/N-1)\sum_{i=1}^{N}(t_i - \bar{t})^2$ and $\bar{t} = (1/N)\sum_{i=1}^{N} t_i$. The modified inverse normal test statistic is formed as

$$Z^{*} = \frac{\sum_{i=1}^{N} t_i}{\sqrt{N + N(N-1)\left[\widehat{\rho}^{\star} + \kappa\sqrt{2/(n+1)}\left(1-\widehat{\rho}^{\star}\right)\right]}}, \tag{2.9}$$

where $\kappa = 0.1(1 + 1/(N-1) - \widehat{\rho}^{\star})$ is a parameter designed to improve the small sample performance of the test statistic. Under the null hypothesis, $Z^{*} \sim N(0,1)$. Demetrescu et al. [7] showed that this method was robust to certain deviations from the assumption of constant correlation between probits in the panel unit root tests.

A third method, proposed by Simes [19] as an improved Bonferroni procedure, is based on the ordered P values, denoted by $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(N)}$. The joint hypothesis H_0 is rejected if

$$p_{(i)} \leq \frac{i\alpha}{N}, \tag{2.10}$$

for at least one $i = 1, \ldots, N$. This procedure has a type I error equal to α when the test statistics are independent. Hanck [8] showed that Simes test was robust to general patterns of cross-sectional dependence in the panel.

The fourth method is Zaykin et al.'s [20] truncated product method (TPM), which takes the product of all those P values that do not exceed some prespecified value τ. The TPM is defined as

$$W = \prod_{i=1}^{N} p_i^{I(p_i \leq \tau)}, \tag{2.11}$$

where $I(\cdot)$ is the indicator function. Note that setting $\tau = 1$ leads to Fisher's original combination method, which could lose power in cases when there are some very large P values. This can happen when some series in the panel are clearly nonstationary such that the resulting P-values are close to 1, and some are clearly stationary such that the resulting P values are close to 0. Ordinary combination methods could be dominated by the large P values. The TPM removes these large P values through truncation, thus eliminating the effect that they could have on the resulting test statistic.

When all the P values are independent, there exists a closed form of the distribution for W under H_0. When the P values are dependent, Monte Carlo simulation is needed to obtain the empirical distribution of W. Sheng and Yang [9] modify the TPM to allow for a certain degree of correlation among the P values. Their procedure is as follows.

Step 1. Calculate W^* using (2.11). Set A = 0.

Step 2. Estimate the correlation matrix, Σ, for P values. Following Hartung [18] and Demetrescu et al. [7], they assume a constant correlation between the probits t_i and t_j,

$$\operatorname{cov}(t_i, t_j) = \rho, \quad \text{for } i \neq j,\ i, j = 1, \ldots, N, \tag{2.12}$$

where $t_i = \Phi^{-1}(p_i)$ and $t_j = \Phi^{-1}(p_j)$. ρ can be estimated in finite samples according to (2.8).

Step 3. The distribution of W^* is calculated based on the following Monte Carlo simulations.

(a) Draw pseudorandom probits from the normal distribution with mean zero and the estimated correlation matrix, $\widehat{\Sigma}$, and transform them back through the standard normal *c.d.f.*, resulting in N P-values, denoted by $\widetilde{p}_1, \widetilde{p}_2, \ldots, \widetilde{p}_N$.

(b) Calculate $\widetilde{W} = \prod_{i=1}^{N} \widetilde{p}_i^{I(\widetilde{p}_i \leq \tau)}$.

(c) If $\widetilde{W} \leq W^*$, increment A by one.

(d) Repeat steps (a)–(c) B times.

(e) The P value for W^* is given by A/B.

3. Monte Carlo Study

In this section we compare the finite sample performance of the P value combination methods introduced in Section 2. We consider "strong" cross-section dependence, driven by a common factor, and "weak" cross-section dependence due to spatial correlation.

3.1. The Design of Monte Carlo

First we consider dynamic panels with fixed effects but no linear trends or residual serial correlation. The data-generating process (DGP) in this case is given by

$$y_{it} = (1-\alpha_i)\mu_i + \alpha_i y_{i,t-1} + \epsilon_{it}, \tag{3.1}$$

where

$$\epsilon_{it} = \gamma_i f_t + \xi_{it}, \tag{3.2}$$

for $i = 1,\ldots,N$, $t = -50,-49,\ldots,T$. The initial values $y_{i,-50}$ are set to be 0 for all i. The individual fixed effect μ_i, the common factor f_t, the factor loading γ_i, and the error term ξ_{it} are independent of each other with $\mu_i \sim$ i.i.d $N(0,1)$, $f_t \sim$ i.i.d $N(0,\sigma_f^2)$, $\gamma_i \sim$ i.i.d $U[0,3]$, and $\xi_{it} \sim$ i.i.d $N(0,1)$.

Remark 3.1. Setting $\sigma_f^2 = 0$, we explore the properties of the tests under cross-section independence, and, with $\sigma_f^2 = 10$, we explore the performance of the tests under "high" cross-section dependence. In the latter case, the average pairwise correlation coefficient of ϵ_{it} and ϵ_{jt} is 70%, representing a strong cross-section correlation in practice.

Next we allow for deterministic trends in the DGP and the Dickey-Fuller (DF) regressions. For this case y_{it} is generated as follows:

$$y_{it} = \kappa_i + (1-\alpha_i)\lambda_i t + \alpha_i y_{i,t-1} + \epsilon_{it}, \tag{3.3}$$

with $\kappa_i \sim$ i.i.d $U[0,0.02]$ and $\lambda_i \sim$ i.i.d $U[0,0.02]$. This ensures that y_{it} has the same average trend properties under the null and the alternative hypotheses. The errors ϵ_{it} are generated according to (3.2) with $\sigma_f^2 = 10$, representing the scenario of high cross-section correlation.

To examine the impact of residual serial correlation, we consider a number of experiments, where the errors ξ_{it} in (3.2) are generated as

$$\xi_{it} = \rho_i \xi_{i,t-1} + e_{it}, \tag{3.4}$$

with $e_{it} \sim$ i.i.d $N(0,1)$. Following Pesaran [15], we choose $\rho_i \sim$ i.i.d $U[0.2,0.4]$ for positive serial correlations and $\rho_i \sim$ i.i.d $U[-0.4,-0.2]$ for negative serial correlations. We use this DGP to check the robustness of the tests to alternative residual correlation models and to the heterogeneity of the coefficients, ρ_i.

Finally we explore the performance of the tests under spatial dependence. We consider two commonly used spatial error processes: the spatial autoregressive (SAR) and the spatial moving average (SMA). Let ϵ_t be the $N \times 1$ error vector in (3.1). In SAR, it can be expressed as

$$\epsilon_t = \theta_1 W_N \epsilon_t + \upsilon_t = (I_N - \theta_1 W_N)^{-1} \upsilon_t, \tag{3.5}$$

where θ_1 is the spatial autoregressive parameter, W_N is an $N \times N$ known spatial weights matrix, and υ_t is the error component which is assumed to be distributed independently across cross-section dimension with constant variance σ_υ^2. Then the full $NT \times NT$ covariance matrix is

$$\Omega_{\text{SAR}} = \sigma_\upsilon^2 \left[I_T \otimes \left(B_N' B_N \right)^{-1} \right], \tag{3.6}$$

where $B_N = I_N - \theta_1 W_N$. In SMA, the error vector ϵ_t can be expressed as

$$\epsilon_t = \theta_2 W_N \upsilon_t + \upsilon_t = (I_N + \theta_2 W_N) \upsilon_t, \tag{3.7}$$

with θ_2 being the spatial moving average parameter. Then the full $NT \times NT$ covariance matrix becomes

$$\Omega_{\text{SMA}} = \sigma_\upsilon^2 \left[I_T \otimes \left(I_N + \theta_2 (W_N + W_N') + \theta_2^2 W_N W_N' \right) \right]. \tag{3.8}$$

Without loss of generality, we let $\sigma_\upsilon^2 = 1$. We consider the spatial dependence with $\theta_1 = 0.8$ and $\theta_2 = 0.8$. The average pairwise correlation coefficient of ϵ_{it} and ϵ_{jt} is 4%–22% for SAR and 2%–8% for SMA, representing a wide range of cross-section correlations in practice. The spatial weight matrix W_N is specified as a "1 ahead and 1 behind" matrix with the ith row, $1 < i < N$, of this matrix having nonzero elements in positions $i + 1$ and $i - 1$. Each row of this matrix is normalized such that all its nonzero elements are equal to $1/2$.

For all of DGPs considered here, we use

$$\alpha_i \begin{cases} \sim \text{i.i.d. } U[0.85, 0.95] & \text{for } i = 1, \dots, N_0, \text{ where } N_0 = \delta \cdot N, \\ = 1 & \text{for } i = N_0 + 1, \dots, N, \end{cases} \tag{3.9}$$

where δ indicates the fraction of stationary series in the panel, varying in the interval 0-1. As a result, changes in δ allow us to study the impact of the proportion of stationary series on the power of tests. When $\delta = 0$, we explore the size of tests. We set $\delta = 0.1$, 0.5 and 0.9 to examine the power of the tests under heterogeneous alternatives. The tests are one-sided with the nominal size set at 5% and conducted for all combinations of N and $T = 20$, 50, and 100. (We also conduct the simulations with the nominal size set at 1% and 10%. The results are qualitatively similar to those at the 5% level, and thus are not reported here.) The results are obtained with MATLAB using $M = 2000$ simulations. To calculate the empirical critical value for the modified TPM, we run additional $B = 1000$ replications within each simulation.

We calculate the augmented Dickey-Fuller (ADF) t statistics. The number of lags in the ADF regressions is selected according to the recursive t-test procedure. (Start with an upper bound, $k_{\max} = 8$, on k. If the last included lag is significant, choose $k = k_{\max}$, if not, reduce k

Table 1: Size and power of panel unit root tests: cross-section independence.

	N	T	P	Z^*	S	W^*
$\delta = 0$	20	20	0.059	0.056	0.052	0.063
		50	0.054	0.046	0.053	0.050
		100	0.047	0.047	0.053	0.052
	50	20	0.040	0.044	0.050	0.045
		50	0.048	0.045	0.048	0.045
		100	0.057	0.054	0.047	0.058
	100	20	0.051	0.050	0.050	0.047
		50	0.047	0.051	0.048	0.057
		100	0.052	0.045	0.049	0.049
$\delta = 0.1$	20	20	0.066	0.072	0.052	0.056
		50	0.087	0.080	0.062	0.083
		100	0.172	0.144	0.112	0.0174
	50	20	0.074	0.081	0.054	0.065
		50	0.123	0.128	0.064	0.102
		100	0.303	0.251	0.121	0.276
	100	20	0.080	0.085	0.048	0.066
		50	0.167	0.165	0.060	0.126
		100	0.464	0.366	0.130	0.435
$\delta = 0.5$	20	20	0.120	0.144	0.066	0.083
		50	0.417	0.489	0.106	0.253
		100	0.951	0.931	0.360	0.860
	50	20	0.181	0.261	0.058	0.108
		50	0.749	0.838	0.119	0.454
		100	1.000	1.000	0.417	0.998
	100	20	0.292	0.422	0.059	0.142
		50	0.950	0.979	0.108	0.683
		100	1.000	1.000	0.447	1.000
$\delta = 0.9$	20	20	0.182	0.283	0.058	0.105
		50	0.816	0.933	0.127	0.495
		100	1.000	1.000	0.580	0.994
	50	20	0.358	0.562	0.054	0.154
		50	0.994	1.000	0.151	0.817
		100	1.000	1.000	0.649	1.000
	100	20	0.591	0.834	0.069	0.257
		50	1.000	1.000	0.156	0.969
		100	1.000	1.000	0.676	1.000

Note. Rejection rates of panel unit root tests at nominal level $\alpha = 0.05$, using 2000 simulations. P is Maddala and Wu's [5] original Fisher test, Z^* is Demetrescu et al.'s [7] modified inverse normal method, S is Hanck's [8] Simes test, and W^* is Sheng and Yang's [9] modified TPM.

by one until the last lag becomes significant. If no lag is significant, set $k = 0$. The 10 percent level of the asymptotic normal distribution is used to determine the significance of the last lag.) As shown in the work of Ng and Perron [21], this sequential testing procedure has better size properties than those based on information criteria in panel unit root tests. The P values in this paper are calculated using the response surfaces estimated in the study by Mackinnon [22].

Table 2: Size and power of panel unit root tests: no serial correlation, cross-section dependence driven by a common factor.

			Intercept only				Intercept and trend			
	N	T	P	Z^*	S	W^*	P	Z^*	S	W^*
$\delta = 0$	20	20	0.239	0.076	0.035	0.054	0.233	0.072	0.041	0.061
		50	0.234	0.070	0.034	0.049	0.259	0.074	0.032	0.062
		100	0.243	0.070	0.036	0.049	0.243	0.075	0.035	0.061
	50	20	0.280	0.070	0.042	0.049	0.297	0.069	0.033	0.063
		50	0.290	0.069	0.030	0.046	0.291	0.061	0.028	0.055
		100	0.290	0.066	0.031	0.047	0.275	0.063	0.031	0.054
	100	20	0.311	0.076	0.048	0.051	0.326	0.082	0.038	0.074
		50	0.305	0.070	0.029	0.050	0.340	0.070	0.024	0.061
		100	0.305	0.068	0.029	0.048	0.300	0.062	0.028	0.054
$\delta = 0.1$	20	20	0.244	0.078	0.034	0.054	0.238	0.067	0.031	0.057
		50	0.263	0.078	0.043	0.054	0.243	0.063	0.036	0.057
		100	0.303	0.099	0.094	0.073	0.272	0.083	0.057	0.078
	50	20	0.301	0.074	0.044	0.050	0.280	0.068	0.031	0.058
		50	0.315	0.070	0.035	0.048	0.310	0.085	0.037	0.075
		100	0.373	0.100	0.090	0.082	0.333	0.080	0.047	0.077
	100	20	0.318	0.077	0.064	0.054	0.319	0.070	0.032	0.059
		50	0.364	0.084	0.041	0.062	0.319	0.068	0.027	0.057
		100	0.410	0.094	0.088	0.084	0.350	0.078	0.051	0.079
$\delta = 0.5$	20	20	0.281	0.093	0.042	0.074	0.251	0.083	0.040	0.075
		50	0.406	0.150	0.065	0.116	0.314	0.102	0.052	0.096
		100	0.679	0.396	0.229	0.351	0.476	0.221	0.125	0.228
	50	20	0.338	0.101	0.053	0.075	0.288	0.068	0.029	0.061
		50	0.486	0.166	0.063	0.127	0.373	0.097	0.047	0.096
		100	0.759	0.433	0.212	0.368	0.565	0.225	0.113	0.237
	100	20	0.402	0.106	0.075	0.086	0.352	0.082	0.031	0.075
		50	0.501	0.158	0.058	0.116	0.405	0.098	0.039	0.094
		100	0.792	0.437	0.196	0.384	0.598	0.223	0.104	0.240
$\delta = 0.9$	20	20	0.314	0.094	0.046	0.069	0.260	0.070	0.036	0.064
		50	0.529	0.172	0.091	0.115	0.368	0.107	0.047	0.100
		100	0.872	0.510	0.305	0.382	0.660	0.282	0.163	0.268
	50	20	0.377	0.088	0.058	0.064	0.298	0.073	0.033	0.066
		50	0.590	0.171	0.076	0.117	0.442	0.107	0.051	0.096
		100	0.913	0.514	0.305	0.390	0.742	0.282	0.148	0.270
	100	20	0.432	0.098	0.078	0.064	0.372	0.077	0.028	0.068
		50	0.655	0.176	0.090	0.122	0.491	0.118	0.054	0.111
		100	0.935	0.508	0.276	0.373	0.769	0.291	0.139	0.270

Note. See Table 1.

3.2. Monte Carlo Results

We compare the finite sample size and power of the following tests: Maddala and Wu's [5] original Fisher test (denoted by P), Demetrescu et al.'s [7] modified inverse normal method (denoted by Z^*), Hanck's [8] Simes test (denoted by S), and Sheng and Yang [9]'s

Table 3: Size and power of panel unit root tests: serial correlation, intercept only, cross-section dependence driven by a common factor.

	N	T	Positive serial correlation				Negative serial correlation			
			P	Z^*	S	W^*	P	Z^*	S	W^*
$\delta = 0$	20	20	0.250	0.116	0.085	0.081	0.255	0.097	0.077	0.075
		50	0.240	0.105	0.063	0.071	0.246	0.076	0.044	0.050
		100	0.224	0.090	0.048	0.054	0.237	0.071	0.033	0.048
	50	20	0.309	0.148	0.126	0.112	0.306	0.096	0.090	0.076
		50	0.289	0.091	0.063	0.068	0.288	0.080	0.050	0.063
		100	0.283	0.087	0.043	0.062	0.285	0.076	0.034	0.051
	100	20	0.335	0.141	0.149	0.114	0.308	0.103	0.100	0.078
		50	0.317	0.100	0.057	0.071	0.301	0.078	0.052	0.056
		100	0.308	0.094	0.042	0.066	0.331	0.074	0.037	0.049
$\delta = 0.1$	20	20	0.256	0.139	0.111	0.116	0.260	0.108	0.079	0.082
		50	0.241	0.104	0.070	0.076	0.263	0.091	0.063	0.064
		100	0.282	0.109	0.093	0.082	0.278	0.091	0.096	0.073
	50	20	0.302	0.141	0.117	0.105	0.303	0.114	0.101	0.087
		50	0.308	0.113	0.072	0.083	0.327	0.087	0.064	0.066
		100	0.354	0.125	0.096	0.099	0.368	0.104	0.098	0.081
	100	20	0.330	0.134	0.139	0.106	0.340	0.117	0.120	0.090
		50	0.363	0.118	0.073	0.088	0.331	0.086	0.063	0.064
		100	0.399	0.133	0.110	0.111	0.394	0.100	0.098	0.085
$\delta = 0.5$	20	20	0.285	0.152	0.117	0.117	0.294	0.136	0.099	0.111
		50	0.383	0.174	0.104	0.132	0.393	0.169	0.093	0.136
		100	0.629	0.382	0.221	0.342	0.636	0.348	0.200	0.315
	50	20	0.351	0.164	0.129	0.135	0.338	0.146	0.112	0.124
		50	0.483	0.190	0.110	0.152	0.463	0.184	0.106	0.157
		100	0.757	0.435	0.231	0.367	0.731	0.388	0.210	0.344
	100	20	0.398	0.175	0.169	0.144	0.387	0.153	0.137	0.130
		50	0.529	0.195	0.108	0.157	0.486	0.162	0.100	0.133
		100	0.781	0.439	0.219	0.382	0.740	0.368	0.204	0.336
$\delta = 0.9$	20	20	0.323	0.151	0.128	0.113	0.327	0.132	0.110	0.104
		50	0.511	0.199	0.124	0.146	0.505	0.182	0.116	0.133
		100	0.858	0.505	0.324	0.393	0.833	0.464	0.276	0.355
	50	20	0.376	0.152	0.139	0.116	0.316	0.144	0.111	0.108
		50	0.598	0.208	0.135	0.152	0.572	0.180	0.108	0.130
		100	0.901	0.494	0.300	0.361	0.614	0.185	0.117	0.135
	100	20	0.415	0.157	0.179	0.121	0.413	0.127	0.131	0.093
		50	0.633	0.185	0.128	0.125	0.613	0.185	0.122	0.138
		100	0.918	0.523	0.320	0.392	0.902	0.478	0.291	0.370

Note. See Table 1.

modified TPM (denoted by W^*). The results in Table 1 are obtained for the case of cross-section independence for a benchmark comparison. Tables 2 and 3 consider the cases of cross-section dependence driven by a single common factor with the trend and residual serial correlation. Table 4 reports the results with spatial dependence. Given the size distortions of some methods, we also include the size-adjusted power in Tables 5, 6, and 7. Major findings of our experiments can be summarized as follows.

Table 4: Size and power of panel unit root tests: intercept only, spatial dependence.

			Spatial autoregressive				Spatial moving average			
	N	T	P	Z^*	S	W^*	P	Z^*	S	W^*
$\delta = 0$	20	20	0.121	0.059	0.040	0.046	0.079	0.050	0.051	0.034
		50	0.126	0.063	0.044	0.048	0.086	0.054	0.050	0.039
		100	0.133	0.066	0.044	0.049	0.091	0.060	0.047	0.043
	50	20	0.140	0.054	0.040	0.038	0.081	0.030	0.042	0.020
		50	0.142	0.063	0.058	0.042	0.089	0.038	0.057	0.024
		100	0.123	0.052	0.051	0.038	0.089	0.039	0.042	0.025
	100	20	0.143	0.046	0.053	0.021	0.095	0.026	0.055	0.010
		50	0.152	0.046	0.051	0.022	0.092	0.027	0.060	0.013
		100	0.136	0.047	0.049	0.023	0.089	0.023	0.052	0.012
$\delta = 0.1$	20	20	0.144	0.068	0.049	0.048	0.089	0.060	0.050	0.040
		50	0.167	0.089	0.057	0.058	0.128	0.073	0.063	0.055
		100	0.244	0.135	0.104	0.119	0.196	0.135	0.113	0.105
	50	20	0.160	0.071	0.053	0.039	0.095	0.043	0.052	0.023
		50	0.198	0.092	0.057	0.056	0.159	0.075	0.069	0.038
		100	0.353	0.189	0.124	0.158	0.320	0.166	0.133	0.129
	100	20	0.161	0.048	0.052	0.024	0.113	0.032	0.050	0.011
		50	0.264	0.097	0.058	0.038	0.203	0.064	0.065	0.018
		100	0.479	0.231	0.121	0.171	0.499	0.217	0.141	0.148
$\delta = 0.5$	20	20	0.199	0.095	0.056	0.064	0.155	0.082	0.053	0.048
		50	0.414	0.217	0.098	0.141	0.425	0.239	0.099	0.132
		100	0.857	0.626	0.318	0.505	0.903	0.745	0.355	0.585
	50	20	0.272	0.098	0.052	0.054	0.224	0.085	0.059	0.033
		50	0.669	0.308	0.118	0.166	0.710	0.369	0.107	0.163
		100	0.989	0.855	0.380	0.732	0.999	0.938	0.402	0.833
	100	20	0.356	0.101	0.068	0.037	0.289	0.076	0.055	0.019
		50	0.848	0.383	0.109	0.166	0.929	0.446	0.117	0.159
		100	1.000	0.967	0.418	0.888	1.000	0.987	0.460	0.955
$\delta = 0.9$	20	20	0.281	0.117	0.072	0.067	0.222	0.090	0.055	0.049
		50	0.687	0.264	0.141	0.163	0.752	0.297	0.152	0.161
		100	0.992	0.754	0.482	0.578	1.000	0.850	0.558	0.652
	50	20	0.407	0.116	0.067	0.057	0.379	0.102	0.067	0.037
		50	0.934	0.276	0.145	0.147	0.980	0.279	0.140	0.131
		100	1.000	0.904	0.579	0.737	1.000	0.958	0.618	0.792
	100	20	0.538	0.108	0.065	0.031	0.541	0.092	0.067	0.027
		50	0.996	0.295	0.145	0.146	1.000	0.247	0.144	0.100
		100	1.000	0.984	0.661	0.846	1.000	0.997	0.675	0.901

Note. See Table 1.

(1) In the absence of clear guidance regarding the choice of τ, we try 10 different values, ranging from 0.05, 0.1, 0.2, . . ., up to 0.9. Our simulation results show that W^* tends to be slightly oversized with a small τ but moderately undersized with a large τ and that its power does not show any clear patterns. We also note that W^* yields similar results as τ varies between 0.05 and 0.2. In our paper we select $\tau = 0.1$. (To save space, the complete simulation results are not reported here, but are available upon request.)

Table 5: Size-adjusted power of panel unit root tests: no serial correlation, cross-section dependence driven by a common factor.

			Intercept only			Intercept and trend		
	N	T	P	Z^*	W^*	P	Z^*	W^*
$\delta = 0.1$	20	20	0.043	0.046	0.044	0.038	0.052	0.037
		50	0.048	0.056	0.049	0.049	0.050	0.050
		100	0.062	0.084	0.063	0.066	0.062	0.064
	50	20	0.061	0.063	0.061	0.047	0.046	0.047
		50	0.053	0.054	0.054	0.042	0.046	0.040
		100	0.066	0.080	0.058	0.056	0.062	0.053
	100	20	0.099	0.121	0.094	0.052	0.058	0.052
		50	0.043	0.051	0.042	0.046	0.045	0.046
		100	0.065	0.069	0.065	0.058	0.067	0.055
$\delta = 0.5$	20	20	0.044	0.071	0.042	0.045	0.056	0.046
		50	0.070	0.116	0.068	0.062	0.075	0.058
		100	0.200	0.349	0.207	0.123	0.163	0.110
	50	20	0.057	0.080	0.057	0.052	0.056	0.047
		50	0.085	0.133	0.076	0.051	0.062	0.041
		100	0.161	0.348	0.160	0.116	0.182	0.099
	100	20	0.114	0.154	0.119	0.048	0.053	0.047
		50	0.069	0.123	0.064	0.057	0.068	0.054
		100	0.189	0.355	0.189	0.099	0.169	0.086
$\delta = 0.9$	20	20	0.059	0.050	0.058	0.061	0.061	0.060
		50	0.140	0.114	0.137	0.088	0.077	0.084
		100	0.456	0.433	0.431	0.250	0.201	0.231
	50	20	0.084	0.073	0.082	0.054	0.053	0.053
		50	0.150	0.114	0.144	0.087	0.073	0.081
		100	0.453	0.424	0.422	0.243	0.225	0.222
	100	20	0.144	0.151	0.131	0.054	0.051	0.054
		50	0.143	0.129	0.136	0.088	0.070	0.086
		100	0.446	0.395	0.414	0.208	0.198	0.191

Note. The power is calculated at the exact 5% level. The 5% critical values for these tests are obtained from their finite sample distributions generated by 2000 simulations for sample sizes $T = 20$, 50, and 100. P is Maddala and Wu's [5] original Fisher test, Z^* is Demetrescu et al.'s [7] modified inverse normal method, and W^* is Sheng and Yang's [9] modified TPM.

(2) With *no* cross-section dependence, all the tests yield good empirical size, close to the 5% nominal level (Table 1). As expected, P test shows severe size distortions under cross-section dependence driven by a common factor or by spatial correlations. For a common factor with *no* residual serial correlation, while Z^* test is mildly oversized and S test is slightly undersized, W^* test shows satisfactory size properties (Table 2). The presence of serial correlation leads to size distortions for all statistics when T is small, which even persist when $T = 100$ for P and Z^* tests. On the contrary, S and W^* tests exhibit good size properties with $T = 50$ and 100 (Table 3). Under spatial dependence, S test performs the best in terms of size, while Z^* and W^* tests are conservative for large N (Table 4).

(3) All the tests become more powerful as N increases, which justifies the use of panel data in unit root tests. When a linear time trend is included, the power of all the

Table 6: Size-adjusted power of panel unit root tests: serial correlation, intercept only, cross-section dependence driven by a common factor.

			Positive correlation			Negative correlation		
	N	T	P	Z^*	W^*	P	Z^*	W^*
$\delta = 0.1$	20	20	0.041	0.051	0.043	0.042	0.048	0.041
		50	0.053	0.054	0.050	0.043	0.054	0.042
		100	0.062	0.065	0.057	0.067	0.077	0.066
	50	20	0.042	0.058	0.043	0.050	0.053	0.048
		50	0.059	0.069	0.063	0.053	0.058	0.053
		100	0.058	0.075	0.052	0.056	0.066	0.054
	100	20	0.045	0.058	0.041	0.044	0.051	0.044
		50	0.056	0.061	0.052	0.063	0.061	0.065
		100	0.054	0.067	0.053	0.069	0.083	0.065
$\delta = 0.5$	20	20	0.040	0.056	0.040	0.045	0.064	0.039
		50	0.072	0.100	0.069	0.069	0.117	0.059
		100	0.188	0.255	0.199	0.158	0.298	0.150
	50	20	0.049	0.059	0.048	0.051	0.093	0.045
		50	0.081	0.120	0.083	0.068	0.127	0.058
		100	0.207	0.302	0.210	0.168	0.309	0.150
	100	20	0.051	0.054	0.046	0.037	0.084	0.033
		50	0.100	0.124	0.095	0.071	0.121	0.063
		100	0.209	0.330	0.221	0.148	0.345	0.127
$\delta = 0.9$	20	20	0.058	0.060	0.056	0.066	0.066	0.065
		50	0.153	0.083	0.148	0.119	0.110	0.114
		100	0.424	0.242	0.391	0.390	0.384	0.376
	50	20	0.068	0.054	0.066	0.065	0.069	0.061
		50	0.162	0.078	0.156	0.136	0.130	0.134
		100	0.454	0.262	0.415	0.376	0.376	0.352
	100	20	0.062	0.052	0.061	0.058	0.063	0.057
		50	0.169	0.088	0.157	0.135	0.114	0.135
		100	0.431	0.268	0.411	0.358	0.371	0.326

Note. See Table 5.

tests decreases substantially. Also notable is the fact that the power of tests increases when the proportion of stationary series increases in the panel.

(4) Compared to the other three tests, the size-unadjusted power of S test is somewhat disappointing here. An exception is that, when only very few series are stationary, S test becomes most powerful. When the proportion of stationary series in the panel increases, however, S test is outperformed by other tests. For example, in the case of *no* cross-section dependence in Table 1 with $\delta = 0.9$, $N = 100$, and $T = 50$, the power of S test is 0.156, and, in contrast, all other tests have power close to 1.

(5) Because P test has severe size distortions, we only compare Z^* and W^* tests in terms of size-adjusted power. (The power is calculated at the exact 5% level. The 5% critical values for these tests are obtained from their finite sample distributions generated by 2000 simulations for sample size $T = 20, 50$, and 100. Since Hanck's [8] test does not have an explicit form of finite sample distribution, we do not calculate its size-adjusted power.) With the cross-section dependence driven by a common

Table 7: Size-adjusted power of panel unit root tests: intercept only, spatial dependence.

			Autoregressive			Moving average		
	N	T	P	Z^*	W^*	P	Z^*	W^*
$\delta = 0.1$	20	20	0.046	0.048	0.046	0.059	0.056	0.049
		50	0.085	0.083	0.078	0.072	0.069	0.067
		100	0.108	0.096	0.102	0.144	0.142	0.148
	50	20	0.055	0.055	0.060	0.058	0.061	0.047
		50	0.098	0.101	0.096	0.098	0.095	0.089
		100	0.165	0.151	0.190	0.234	0.207	0.235
	100	20	0.069	0.071	0.066	0.065	0.068	0.062
		50	0.103	0.091	0.088	0.125	0.119	0.105
		100	0.283	0.247	0.311	0.360	0.334	0.369
$\delta = 0.5$	20	20	0.088	0.067	0.081	0.117	0.094	0.076
		50	0.238	0.178	0.190	0.298	0.211	0.211
		100	0.669	0.557	0.640	0.852	0.760	0.785
	50	20	0.114	0.093	0.093	0.141	0.104	0.085
		50	0.442	0.298	0.321	0.593	0.381	0.379
		100	0.957	0.850	0.922	0.995	0.961	0.984
	100	20	0.168	0.114	0.116	0.211	0.146	0.121
		50	0.654	0.393	0.465	0.859	0.578	0.615
		100	0.999	0.974	0.996	1.000	1.000	1.000
$\delta = 0.9$	20	20	0.100	0.066	0.087	0.158	0.105	0.100
		50	0.519	0.223	0.359	0.644	0.272	0.438
		100	0.968	0.636	0.918	0.999	0.845	0.977
	50	20	0.218	0.113	0.143	0.284	0.142	0.135
		50	0.812	0.269	0.578	0.958	0.371	0.709
		100	1.000	0.864	0.998	1.000	0.987	1.000
	100	20	0.329	0.124	0.174	0.472	0.143	0.227
		50	0.977	0.285	0.795	0.999	0.569	0.930
		100	1.000	0.988	1.000	1.000	1.000	1.000

Note. See Table 5.

factor, Z^* test tends to deliver higher power for $\delta = 0.5$ but lower power for $\delta = 0.9$ than W^* test (Tables 5 and 6). Under spatial dependence, however, the former is clearly dominated by the latter in most of the time. This is especially true for SAR process, where W^* test exhibits substantially higher size-adjusted power than Z^* test (Table 7).

4. Empirical Application

Purchasing Power Parity (PPP) is a key assumption in many theoretical models of international economics. Empirical evidence of PPP for the floating regime period (1973–1998) is, however, mixed. While several authors, such as Wu and Wu [23] and Lopez [24], found supporting evidence, others [10, 15, 25] questioned the validity of PPP for this period. In this section, we use the methods discussed in previous sections to investigate if the real exchange rates are stationary among a group of OECD countries.

Table 8: Unit root tests for 27 OECD real exchange rates.

US dollar real exchange rate				Deutchemark real exchange rate			
Country	k	P value	Simes criterion	Country	k	P value	Simes criterion
New Zealand	8	0.008	0.002	Mexico	3	0.006	0.002
Sweden	8	0.053	0.004	Iceland	0	0.010	0.004
United Kingdom	7	0.055	0.006	Australia	3	0.012	0.006
Finland	7	0.058	0.007	Korea	0	0.014	0.007
Spain	8	0.061	0.009	Canada	7	0.040	0.009
Mexico	3	0.066	0.011	Sweden	0	0.074	0.011
Iceland	8	0.069	0.013	United States	4	0.148	0.013
Switzerland	4	0.071	0.015	New Zealand	0	0.171	0.015
France	4	0.080	0.017	Finland	6	0.232	0.017
Netherlands	4	0.099	0.019	Turkey	8	0.241	0.019
Austria	4	0.102	0.020	Netherlands	1	0.415	0.020
Italy	4	0.103	0.022	Norway	7	0.417	0.022
Belgium	4	0.135	0.024	Spain	0	0.459	0.024
Korea	0	0.138	0.026	France	0	0.564	0.026
Germany	4	0.148	0.028	Italy	0	0.565	0.028
Greece	4	0.150	0.030	Poland	5	0.579	0.030
Norway	7	0.167	0.031	Hungary	4	0.612	0.031
Denmark	3	0.206	0.033	Belgium	0	0.618	0.033
Ireland	7	0.235	0.035	Luxembourg	0	0.655	0.035
Japan	4	0.246	0.037	Japan	5	0.656	0.037
Luxembourg	3	0.276	0.039	United Kingdom	0	0.697	0.039
Portugal	8	0.332	0.041	Denmark	0	0.698	0.041
Australia	3	0.386	0.043	Ireland	0	0.708	0.043
Poland	0	0.414	0.044	Austria	0	0.720	0.044
Turkey	8	0.418	0.046	Switzerland	8	0.733	0.046
Canada	6	0.580	0.048	Portugal	0	0.786	0.048
Hungary	0	0.816	0.050	Greece	5	0.880	0.050
P		0.097				0.015	
Z^*		0.095				0.016	
W^*		0.257				0.002	

Note. Simes criterion is calculated using the 5% significance level.

The log real exchange rate between country i and the US is given by

$$q_{it} = s_{it} - p_{\mathrm{us},t} + p_{it}, \tag{4.1}$$

where s_{it} is the nominal exchange rate of the ith country's currency in terms of US dollar and $p_{\mathrm{us},t}$ and p_{it} are consumer price indices in the US and country i, respectively. All these variables are measured in natural logarithms. We use quarterly data from 1973 : 1 to 1998 : 2 for 27 OECD countries, as listed in Table 8. (Two countries, Czech Republic and Slovak Republic, are excluded from our analysis, since their data span a very limited period of time, starting at 1993 : 1.) All data are obtained from the IMF's International Financial Statistics. (Note that, for Iceland, the consumer price indices are missing during 1982:Q1–1982:Q4 in the original data. We filled out this gap by calculating the level of CPI from its percentage changes in the IMF database.)

As the first stage in our analysis we estimated individual ADF regressions:

$$\Delta q_{it} = \mu_i + \phi_i q_{i,t-1} + \sum_{j=1}^{k_i} \varphi_{ij} \Delta q_{i,t-j} + \varepsilon_{it}, \quad i = 1, \ldots, N; \; t = k_i + 2, \ldots, T. \tag{4.2}$$

The null and alternative hypotheses for testing PPP are specified in (2.3) and (2.4), respectively. The selected lags and the P values are reported in Table 8. The results in the left panel show that the ADF test does not reject the unit root null of real exchange rate at the 5% level except for New Zealand. As a robustness check, we investigated the impact of a change in numeraire on the results. The right panel reports the estimation results when the Deutsche mark is used as the numeraire. Out of 27 countries, only 5—Mexico, Iceland, Australia, Korea, and Canada—reject the null of unit root at the 5% level.

As is well known, the ADF test has low power with a short time span. Exploring the cross-section dimension is an alternative. However, if a positive cross-section dependence is ignored, panel unit root tests can also lead to spurious results, as pointed out by O'Connell [10]. As a preliminary check, we compute the pairwise cross-section correlation coefficients of the residuals from the above individual ADF regressions, $\widehat{\rho}_{ij}$. The simple average of these correlation coefficients is calculated, according to Pesaran [26], as

$$\overline{\widehat{\rho}} = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \widehat{\rho}_{ij}. \tag{4.3}$$

The associated cross-section dependence (CD) test statistic is calculated using

$$\text{CD} = \sqrt{\frac{2T}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \widehat{\rho}_{ij}. \tag{4.4}$$

In our sample $\overline{\widehat{\rho}}$ is estimated as 0.396 and 0.513 when US dollar and Deutchemark are considered as the numeraire, respectively. The CD statistics, 71.137 for the former and 93.368 for the latter, strongly reject the null of no cross-section dependence at the conventional significance level.

Now turning to panel unit root tests, Simes test does not reject the unit root null, regardless of which numeraire, US dollar or Deutchemark, is used. However, the evidence is mixed, as illustrated by other test statistics. For 27 OECD countries as a whole, we find substantial evidence against the unit root null with Deutchemark but not with US dollar. In summary, our results from panel unit root tests are numeraire specific, consistent with Lopez [24], and provide mixed evidence in support of PPP for the floating regime period.

5. Conclusion

We conduct a systematic comparison of the performance of four commonly used P-value combination methods applied to panel unit root tests: the original Fisher test, the modified inverse normal method, Simes test, and the modified TPM. Monte Carlo evidence shows that, in the presence of both "strong" and "weak" cross-section dependence, the original Fisher test is severely oversized but the other three tests exhibit good size properties with moderate and large T. In terms of power, Simes test is useful when the total evidence against the joint null hypothesis is concentrated in one or very few of the tests being combined, and the modified inverse normal method and the modified TPM perform well when evidence against the joint null is spread among more than a small fraction of the panel units. Furthermore, under spatial dependence, the modified TPM yields the highest size-adjusted power. We investigate the PPP hypothesis for a panel of OECD countries and find mixed evidence.

The results of this work provide practitioners with guidelines to follow for selecting an appropriate combination method in panel unit root tests. A worthwhile extension would be to develop bootstrap P value combination methods that are robust to general forms of cross-section dependence in panel data. This issue is currently under investigation by the authors.

Acknowledgment

The authors have benefited greatly from discussions with Kajal Lahiri and Dmitri Zaykin. They also thank the guest editor Mike Tsionas and an anonymous referee for helpful comments. The usual disclaimer applies.

References

[1] L. H. C. Tippett, *The Method of Statistics*, Williams and Norgate, London, UK, 1931.

[2] R. A. Fisher, *Statistical Methods for Research Workers*, Oliver and Boyd, London, UK, 4th edition, 1932.

[3] L. V. Hedges and I. Olkin, *Statistical Methods for Meta-Analysis*, Academic Press, Orlando, Fla, USA, 1985.

[4] T. M. Loughin, "A systematic comparison of methods for combining -values from independent tests," *Computational Statistics & Data Analysis*, vol. 47, no. 3, pp. 467–485, 2004.

[5] G. S. Maddala and S. Wu, "A comparative study of unit root tests with panel data and a new simple test," *Oxford Bulletin of Economics and Statistics*, vol. 61, pp. 631–652, 1999.

[6] I. Choi, "Unit root tests for panel data," *Journal of International Money and Finance*, vol. 20, no. 2, pp. 249–272, 2001.

[7] M. Demetrescu, U. Hassler, and A. I. Tarcolea, "Combining significance of correlated statistics with application to panel data," *Oxford Bulletin of Economics and Statistics*, vol. 68, no. 5, pp. 647–663, 2006.

[8] C. Hanck, "Intersection test for panel unit roots," *Econometric Reviews*. In press.

[9] X. Sheng and J. Yang, "A simple panel unit root test by combining dependent p-values," SSRN working paper, no. 1526047, 2009.

[10] P. G. J. O'Connell, "The overvaluation of purchasing power parity," *Journal of International Economics*, vol. 44, no. 1, pp. 1–19, 1998.

[11] P. C. B. Phillips and D. Sul, "Dynamic panel estimation and homogeneity testing under cross section dependence," *Econometrics Journal*, vol. 6, no. 1, pp. 217–259, 2003.

[12] J. Bai and S. Ng, "A PANIC attack on unit roots and cointegration," *Econometrica*, vol. 72, no. 4, pp. 1127–1177, 2004.

[13] Y. Chang, "Bootstrap unit root tests in panels with cross-sectional dependency," *Journal of Econometrics*, vol. 120, no. 2, pp. 263–293, 2004.

[14] H. R. Moon and B. Perron, "Testing for a unit root in panels with dynamic factors," *Journal of Econometrics*, vol. 122, no. 1, pp. 81–126, 2004.

[15] M. H. Pesaran, "A simple panel unit root test in the presence of cross-section dependence," *Journal of Applied Econometrics*, vol. 22, no. 2, pp. 265–312, 2007.

[16] J. Breitung and M. H. Pesaran, "Unit root and cointegration in panels," in *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, L. Matyas and P. Sevestre, Eds., pp. 279–322, Kluwer Academic Publishers, Boston, Mass, USA, 2008.

[17] S. A. Stouffer, E. A. Suchman, L. C. DeVinney, S. A. Star, and R. M. Williams Jr., *The American Soldier*, vol. 1 of *Adjustment during Army Life*, Princeton University Press, Princeton, NJ, USA, 1949.

[18] J. Hartung, "A note on combining dependent tests of significance," *Biometrical Journal*, vol. 41, no. 7, pp. 849–855, 1999.

[19] R. J. Simes, "An improved Bonferroni procedure for multiple tests of significance," *Biometrika*, vol. 73, no. 3, pp. 751–754, 1986.

[20] D. V. Zaykin, L. A. Zhivotovsky, P. H. Westfall, and B. S. Weir, "Truncated product method for combining P-values," *Genetic Epidemiology*, vol. 22, no. 2, pp. 170–185, 2002.

[21] S. Ng and P. Perron, "Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag," *Journal of the American Statistical Association*, vol. 90, no. 429, pp. 268–281, 1995.

[22] J. G. Mackinnon, "Numerical distribution functions for unit root and cointegration tests," *Journal of Applied Econometrics*, vol. 11, no. 6, pp. 601–618, 1996.

[23] J.-L. Wu and S. Wu, "Is purchasing power parity overvalued?" *Journal of Money, Credit and Banking*, vol. 33, pp. 804–812, 2001.

[24] C. Lopez, "Evidence of purchasing power parity for the floating regime period," *Journal of International Money and Finance*, vol. 27, no. 1, pp. 156–164, 2008.

[25] I. Choi and T. K. Chue, "Subsampling hypothesis tests for nonstationary panels with applications to exchange rates and stock prices," *Journal of Applied Econometrics*, vol. 22, no. 2, pp. 233–264, 2007.

[26] M. H. Pesaran, "General diagnostic tests for cross section dependence in panels," Cambridge working papers in economics, no. 435, University of Cambridge, 2004.