Combination of “Combinations of P-values”*

Lan Cheng
Department of Mathematical Sciences, SUNY Fredonia
Lan.Cheng@fredonia.edu

and

Xuguang Sheng †
Department of Economics, American University
sheng@american.edu

Abstract

We investigate the impact of uncertainty over the number of false individual null hypotheses on commonly used p-value combination methods, and develop a combination of “combinations of p-values” (CCP) test that maintains good power properties across such an uncertainty. Our test is motivated by the conflicting results of empirical studies of meta-analysis that differ by methodology. We base our test on a simple union of rejections decision rule that exploits the correlation between two p-value combination methods. Monte Carlo studies show that the CCP test controls size and its power is always close to that of the better individual methods.

JEL Classification: C12; C33

Key words: Combination methods; Hypothesis testing; p value; Union of rejections.

*This paper was presented at the Seaway Section of the Mathematical Association of America in Plattsburgh and North American Summer Meeting of the Econometric Society in St. Louis. We thank Bruce Hansen, Nancy Boynton, Eric Renault, Jonathan Wright, Ulrich Müller and conference participants for helpful comments and suggestions. The usual disclaimer applies.

†Corresponding author. Contact information: 4400 Massachusetts Ave., N.W., Washington, DC 20016; Tel: (202) 885-3782; Fax: (202) 885-3790.

1 Introduction

Combining independent p-values from a set of hypothesis tests has long been used in meta analysis (Hedges and Olkin 1985). Given n individual hypotheses H_{0i}, $i = 1, 2, ..., n$, consider a test for the joint null hypothesis, $H_0 = \bigcap_{i=1}^{n} H_{0i}$. H_0 is true if all the individual null hypotheses are true, and false if at least one of its components is false. Many attempts have been made to compare the overall size (probability of rejecting H_0 when it is true) and the power (probability of rejecting H_0 when there is at least one false individual null hypothesis) of different combination methods. See, for example, Berk and Cohen (1979), Westberg (1985), Zaykin, et al. (2002), Neuhäuser (2003), and Loughin (2004), among others. A key result of this literature is that no uniformly most powerful test exists. Some tests, for example, Tippett's and Simes' methods, are powerful when there is only one or very few false individual null hypotheses. Others, such as Fisher's and Stouffer's methods, are powerful when there are many false individual null hypotheses. Since in practice it is *a priori* unknown whether few or almost all individual null hypotheses are false, researchers have difficulty in selecting which method to use if the results from p-value combination methods conflict. Hence, a new procedure is needed with the aim of maintaining good power properties across such an uncertainty.

Recently, Loughin (2004, p.484) suggested that "combinations of combinations could be considered rather than relying on a single function for all purposes". Neuhäuser (2003) introduced a combination of Simes' method and the truncated product method based on the union of rejections decision rule. In an independent study, Harvey, et al. (2009) used the similar strategy to combine two unit root tests and pointed out that many practitioners actually do the union of rejections, albeit informally or unconsciously. However, as pointed out by Neuhäuser (2003) and Harvey, et al. (2009), such a combination is conservative since two dependent methods are combined using the Bonferroni inequality.

In this paper we propose a combination of "combinations of p-values" (CCP) test, based on a simple union of rejections decision rule. If at least one of the two p-value combination methods yields a rejection at the significance level γ, then the joint null hypothesis is rejected at the significance level α. The value of γ is determined such that (i) the CCP test maintains the size of α, (ii) neither of the p-value combination methods dominates, and the probability that both methods reject H_0 when it is true is minimized. Our Monte Carlo simulations show that the power of the CCP test is close to that of the better individual method, irrespective of the number of false individual null hypotheses in the panel. At the same time, the CCP test avoids the size distortion inherent in separately applying p-value combination methods to the same data set, thus circumventing the problem of "conservative combination". As illustrated by one empirical example on testing purchasing power parity, the proposed test avoids the arbitrary decision about which method to use when the results from individual p-value combination methods conflict. If practitioners were to apply the CCP test to empirical studies, a significant step would be made towards objectively integrating and synthesizing the results.

2 The combination of "combinations of p-values" test

Suppose that tests have been conducted for n individual hypotheses H_{0i}, $i = 1, 2, \ldots, n$. For each test the p-value p_i, $i = 1, 2, \ldots, n$, is calculated: if H_{0i} is true, which is the probability of observing a test statistic at least as extreme as the one that was actually observed. In this paper we consider the problem of testing the joint null hypothesis $H_0 = \bigcap_{i=1}^{n} H_{0i}$ at a specific significance level α when the p-values are independent. The combined alternative, denoted by H_A, is true if at least one of H_{0i} is false for $i = 1, 2, \ldots, n$. Various methods have been proposed in the literature to combine p-values. Here we briefly summarize four popular methods.

- Tippett (1931) suggested rejecting H_0 if $p_i \leq 1 - (1-\alpha)^{1/n}$ for at least one $i = 1, 2, \ldots, n$.

- Simes (1986) proposed that H_0 be rejected if $p_{(i)} \leq i\alpha/n$ for at least one $i = 1, 2, \ldots, n$, where $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(n)}$ are the ordered p-values of $p_1, p_2, \ldots, p_n$.

- Fisher (1932) noted that the test statistic

$$t = -2\sum_{i=1}^{n} \ln(p_i) = -2\ln\prod_{i=1}^{n} p_i$$

has a χ^2 distribution with $2n$ degrees of freedom. The p-value for testing H_0 is

$$p = 1 - F_{\chi^2}(t; 2n),$$

where $F_{\chi^2}(x; k)$ is the cumulative distribution function (c.d.f.) for a χ^2 distribution with k degrees of freedom.

- Stouffer's method, attributed to Stouffer et al. (1949), transforms the p values via the standard normal distribution. The test statistic is defined as

$$z = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \Phi^{-1}(p_i),$$

where $\Phi(\cdot)$ is the standard normal c.d.f. When p_i, $i = 1, 2, \ldots, n$, are independent, $z \sim N(0,1)$ for any $n \geq 1$. Therefore the p-value for testing H_0 is

$$p = \Phi(z) = \Phi\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \Phi^{-1}(p_i)\right).$$

The problem of selecting a method is complicated by the fact that there are many different ways in which H_0 can be false. In general, we cannot expect one method to be sensitive to all possible alternatives. For example, Tippett's and Simes' methods are powerful when the total evidence against the joint null hypothesis is concentrated in one or very few of the p-values being combined, but Fisher's and Stouffer's methods have good performance when evidence against the joint null is spread among more than a small fraction of the units. Since these methods are imperfectly correlated, often one method rejects H_0 whereas the other does not, complicating the test decision.

2.1 Construction of the CCP test

By exploiting the imperfect correlation between two combination methods, we propose a CCP test with the aim of retaining robust and good power in the absence of any information regarding the number of false individual null hypotheses. Let T_i denote the test statistics of two p-value combination methods whose significance levels, γ_i, and the corresponding critical values to reject H_0, c_i, are given by

$$\gamma_i(c_i) = \Pr(T_i \geq c_i | H_0 \text{ is true}), \quad i = 1, 2.$$

Given the significance level α, we define the test statistic for the CCP test to be

$$T_c = 1_{\{\bigcup_{i=1}^{2} T_i \geq c_i\}},$$

where $1_{\{A\}}$ is the indicator function of event A, and c_1 and c_2 are to be determined such that

$$\Pr(T_c = 1 | H_0 \text{ is true }) = \alpha. \tag{1}$$

Let w denote the probability that both p-value combination methods reject H_0 when it is true, that is

$$w(c_1, c_2) = \Pr\left(\bigcap_{i=1}^{2} T_i \geq c_i | H_0 \text{ is true}\right).$$

Since

$$\Pr(T_c = 1 | H_0 \text{ is true}) = \Pr\left(\bigcup_{i=1}^{2} T_i \geq c_i | H_0 \text{ is true}\right) = \gamma_1(c_1) + \gamma_2(c_2) - w(c_1, c_2),$$

equation (1) can also be written as

$$\gamma_1(c_1) + \gamma_2(c_2) - w(c_1, c_2) = \alpha. \tag{2}$$

Equation (2) defines an entire family of tests with different pairs of c_1 and c_2, by which the CCP test assigns different weights to each p-value combination method and controls the size at the level of α.

In order for the CCP test to have good power properties across the uncertain number of false individual null hypotheses, the correlation between two p-value combination methods should be made as small as possible. For this reason, the probability that both methods reject H_0 when it is true needs to be minimized. Moreover, we want to avoid the case when one method dominates. For example, if $\gamma_1 \approx 0$ and $\gamma_2 \approx \alpha$, then the CCP test is almost reduced to the second method. Hence, w/γ_1 and w/γ_2 need to be minimized as well. To achieve these goals, we choose c_1 and c_2 to minimize both $w(c_1, c_2)$ and

$$\rho(c_1, c_2) = \sqrt{\frac{w(c_1, c_2)}{\gamma_1(c_1)}}\sqrt{\frac{w(c_1, c_2)}{\gamma_2(c_2)}} = \frac{w(c_1, c_2)}{\sqrt{\gamma_1(c_1)\gamma_2(c_2)}},$$

subject to equation (2).

Indeed ρ plays an important role in the CCP test, since it can measure the correlation between two p-value combination methods. If $\rho = 0$, then $w = 0$ while $\gamma_i \neq 0$, $i = 1, 2$, which implies that both methods never reject H_0 together. In this case, we say two methods are uncorrelated. If $\rho = 1$, then $\gamma_1 = \gamma_2 = w = \alpha$, which indicates two methods have the same chance to rejects H_0 individually and together. In this case, we say two methods are perfectly correlated. For all other cases, $0 < \rho < 1$.

Lemma 2.1. *To minimize $w(c_1, c_2)$ subject to equation (2), c_1 and c_2 need to be chosen such that $d\gamma_1 = -d\gamma_2$.*

Differentiating equation (2) with respect to c_1 and applying the Lagrange multiplier method give the result. Lemma 2.1 shows that when $w(c_1, c_2)$ is minimized, if γ_i is increased, then γ_j has to be decreased by the same amount, where $i, j \in \{1, 2\}$ and $i \neq j$.

Theorem 2.2. *When $w(c_1, c_2)$ is minimized, to minimize $\rho(c_1, c_2)$ subject to equation (2), c_1 and c_2 need to be chosen such that either $\gamma_1 = \gamma_2$ or $w = 0$.*

Proof. According to lemma 2.1, when $w(c_1, c_2)$ is minimized, we have

$$\frac{\gamma_1'(c_1)}{\gamma_2'(c_2)} = \frac{\gamma_1'(c_1) - \dfrac{\partial w}{\partial c_1}}{\gamma_2'(c_2) - \dfrac{\partial w}{\partial c_2}}. \tag{3}$$

To minimize $\rho(c_1, c_2)$ subject to equation (2), we need to choose c_1 and c_2 such that

$$\frac{\gamma_1'(c_1) - \dfrac{\partial w}{\partial c_1}}{\gamma_2'(c_2) - \dfrac{\partial w}{\partial c_2}} = \frac{\dfrac{\partial \rho}{\partial c_1}}{\dfrac{\partial \rho}{\partial c_2}} = \frac{\gamma_2 \left[2\gamma_1 \dfrac{\partial w}{\partial c_1} - w\gamma_1'\right]}{\gamma_1 \left[2\gamma_2 \dfrac{\partial w}{\partial c_2} - w\gamma_2'\right]},$$

which implies

$$\begin{aligned}
\frac{\gamma_1'}{\gamma_2'} &= \frac{\gamma_1'(c_1) - \dfrac{\partial w}{\partial c_1}}{\gamma_2'(c_2) - \dfrac{\partial w}{\partial c_2}} = \frac{2\gamma_1\gamma_2\left[\gamma_1' - \frac{\partial w}{\partial c_1}\right] + \gamma_2\left[2\gamma_1\frac{\partial w}{\partial c_1} - w\gamma_1'\right]}{2\gamma_1\gamma_2\left[\gamma_2' - \frac{\partial w}{\partial c_2}\right] + \gamma_1\left[2\gamma_2\frac{\partial w}{\partial c_2} - w\gamma_2'\right]} \\
&= \frac{2\gamma_2\gamma_2\gamma_1' - w\gamma_2\gamma_1'}{2\gamma_1\gamma_2\gamma_2' - w\gamma_1\gamma_2} = \frac{\gamma_1'(2\gamma_1\gamma_2 - w\gamma_2)}{\gamma_2'(2\gamma_1\gamma_2 - w\gamma_1)},
\end{aligned} \tag{4}$$

where the first equality follows from equation (3). Then equation (4) implies that

$$w(\gamma_1 - \gamma_2) = 0.$$

That is, $\gamma_1(c_1) = \gamma_2(c_2)$ or $w(c_1, c_2) = 0$. It is obvious that ρ achieves its minimum value at the above points. □

According to Theorem 2.2, we need a pair of numbers c_1 and c_2 such that either $w = 0$ or $\gamma_1 = \gamma_2$. However, we cannot always find a pair that yields $w = 0$ for any two p-value combination methods. As a general rule, we propose to choose c_1 and c_2 so that (i) equation (2) holds and (ii) $\gamma_1(c_1) = \gamma_2(c_2) = \gamma$.

In the context of unit root tests, Harvey, et al. (2009) proposed a union of rejections test that rejects if either of the individual methods exceed their α-level critical value adjusted by a common scaling constant. Müller (2009) further showed that this union of rejections test comes close to exploiting the available information efficiently in the sense that it has a power that is never much below the power of the weighted average power maximizing test under uncertainty over the trend.

2.2 Size of the CCP test

Given any two p-value combination methods, the size of the CCP test, $\Pr(T_c = 1|H_0 \text{ is true})$, depends on both γ and n. We define it as

$$g(\gamma; n) = \Pr(T_c = 1|H_0 \text{ is true}),$$

and it is easy to see that $g(\gamma; n) = 2\gamma - w$. Given the significance level for the CCP test $\alpha \in (0, 1)$, γ is to be determined such that

$$g(\gamma; n) = \alpha.$$

Since $g(\gamma; n) = 2\gamma - w$ and $0 \leq w \leq \gamma$, it follows that

$$\frac{\alpha}{2} \leq \gamma \leq \alpha, \tag{5}$$

and consequently

$$\rho = \frac{w}{\sqrt{\gamma_1\gamma_2}} = \frac{2\gamma - \alpha}{\gamma} = 2 - \frac{\alpha}{\gamma} \in [0, 1]. \tag{6}$$

Remark 2.1. *According to equation (6), for each fixed n, ρ is uniquely determined by γ, and vise versa. We consider three different cases regarding the correlation between two p-value combination methods.*

Case (i): If $\gamma = \frac{\alpha}{2}$, then $\rho = 0$ and the two p-value combination methods are uncorrelated;

Case (ii): If $\gamma = \alpha$, then $\rho = 1$ and the two p-value combination methods are perfectly correlated;

Case (iii): If $\frac{\alpha}{2} < \gamma < \alpha$, then $0 < \rho < 1$ and the two p-value combination methods are correlated.

In order to determine γ, we need to find out $g(\gamma; n)$, which can be obtained either analytically or numerically.

2.2.1 Combining Tippett's and Simes' methods: closed-form solution

Let T_1=Tippett's method and T_2 =Simes' method. Now we show that $g(\gamma; n)$ has a closed-form formula.

Theorem 2.3. *Let $P_{(1)}, \cdots, P_{(n)}$ be the ordered p-values for testing individual hypotheses H_{0i}, $i = 1, \ldots, n$. Let $0 < \gamma < 0.2$ and $\zeta = 1 - (1-\gamma)^{1/n}$. Then*

$$\begin{aligned} g(\gamma; n) &= Pr\left[\left(\bigcup_{i=1}^{n}\left(P_{(i)} \leq \frac{i\gamma}{n}\right)\right) \bigcup \left(\bigcup_{i=1}^{n}(P_i \leq \zeta)\right)\right] \\ &= \gamma + n(\zeta - \frac{\gamma}{n})(1-\gamma)(1-\frac{\gamma}{n})^{n-2}. \end{aligned} \tag{7}$$

Proof. See the appendix. □

Once $g(\gamma; n)$ is obtained by equation (7), γ can be determined by solving

$$g(\gamma; n) = \gamma + n(\zeta - \frac{\gamma}{n})(1-\gamma)(1-\frac{\gamma}{n})^{n-2} = \alpha. \tag{8}$$

When $n = 1$, equation (8) is reduced to $\gamma = \alpha$. When $n \geq 2$, equation (8) is nonlinear and Newton-Raphson's iteration can be applied. Since $\gamma > 0$ is usually small, another approach is to use a second order Taylor polynomial at $\gamma = 0$ to approximate $g(\gamma; n)$. Then equation (8) becomes

$$\alpha = g(\gamma; n) \approx \gamma + \frac{n-1}{2n}\gamma^2,$$

which has only one positive solution

$$\gamma \approx \frac{-n + \sqrt{n^2 + 2n^2\alpha - 2n\alpha}}{n-1}.$$

Remark 2.2. *In Theorem 2.3, we only consider the case where* $0 < \gamma < 0.2$*, since* $\gamma < \alpha$ *and in practice the significance level* α *is usually chosen as* 0.01*,* 0.05 *or* 0.1*. Then by equation (8), we have*

$$\alpha = g(\gamma; n) = \gamma + n(\zeta - \frac{\gamma}{n})(1-\gamma)(1-\frac{\gamma}{n})^{n-2} = \gamma + o(\gamma^2) \approx \gamma. \tag{9}$$

Hence Tippett's and Simes' method are almost perfectly correlated according to Remark 2.1.

2.2.2 Combining other tests: numerical approach

In most cases, $g(\gamma; n)$ does not have a closed-form formula. Monte Carlo simulation can then be used to generate $g(\gamma; n)$. Assume that $g(\gamma; n)$ is continuous and third differentiable on $[0, 0.2]$. Then its second order Taylor polynomial at $\gamma = 0$ is

$$g(\gamma; n) = a_0^n + a_1^n\gamma + a_2^n\gamma^2 + b_n\gamma^3, \tag{10}$$

where $b_n\gamma^3$ is the remainder. Linear regression is used to estimate a_0^n, a_1^n, a_2^n and b_n. Let K be a positive integer and step size be $h = \frac{0.2}{K}$. Define

$$\gamma_k = hk, \quad \text{for } k = 1, 2, \ldots, K.$$

For each $k = 1, 2, \ldots, K$, the value of $y_k = g(\gamma_k; n)$ is generated by the Monte Carlo simulation as follows.

1. Generate n i.i.d. uniformly random numbers $p_1, p_2, \ldots p_n$ on $[0, 1]$ as a set of p-values.

2. Compute the value of test statistics $T_{c,k}$ when the significance level for the individual p-value combination method is $\gamma_k = hk$.

3. Repeat the steps above M times and get a random sample $T^1_{c,k}, T^2_{c,k}, \ldots, T^M_{c,k}$. Then $y_k = g(\gamma_k; n) \approx \overline{T_{c,k}} = \frac{1}{M}\sum_{m=1}^{M} T^m_{c,k}$.

Due to the randomness, we have

$$y_k = a^n_0 + a^n_1\gamma_k + a^n_2(\gamma_k)^2 + b_n(\gamma_k)^3 + \varepsilon_k, \qquad k = 1, 2, \ldots, K,$$

where ε_k is the noise. This system of equations can be expressed in matrix form as

$$y = BA + \varepsilon,$$

where

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_K \end{pmatrix}, B = \begin{pmatrix} 1 & \gamma_1 & \gamma_1^2 & \gamma_1^3 \\ 1 & \gamma_2 & \gamma_2^2 & \gamma_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \gamma_K & \gamma_K^2 & \gamma_K^3 \end{pmatrix}, A = \begin{pmatrix} a^n_0 \\ a^n_1 \\ a^n_2 \\ b_n \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_K \end{pmatrix}.$$

The least square estimator for A is

$$\hat{A} = (B'B)^{-1}B'y.$$

After the coefficients in equation (10) are estimated, γ is then obtained by solving

$$g(\gamma; n) \approx a^n_0 + a^n_1\gamma + a^n_2\gamma^2 = \alpha.$$

Note that the error arising from approximating $g(\gamma; n)$ by a second order Taylor polynomial is of size $o(\gamma^3)$ and can be negligible.

2.3 Power of the CCP test

When the joint null hypothesis H_0 is false, it is *a priori* unknown whether one, few or almost all individual null hypotheses H_{0i}, $i = 1, 2, \ldots, n$ are false. Let $L \in \{1, 2, \ldots, n\}$, generally unknown, denote the number of false individual hypotheses. The goal of this subsection is to examine how changes in L affect the power of the CCP test.

A model is needed to simulate the p-value of each individual hypothesis H_{0i}. According to Loughin (2004), the p-value is a random variable, denoted by P, with the following properties: (i) under the individual null hypothesis H_{0i}, $P \sim U(0, 1)$; (ii) under the alternative hypothesis, denoted by H_{Ai}, the density of P should be non-increasing; (iii) it should be possible to define one or more parameters that can stochastically order the densities, mimicking the effects of increasing sample sizes under H_{Ai}, with H_{0i} as a limiting case. Although there are many functions that can serve as the density function of P, we use beta density

$$f_\beta(p) = \beta(a, b)p^{a-1}(1 - p)^{b-1},$$

where $\beta(a, b) = \left(\int_0^1 t^{a-1}(1 - t)^{b-1}\, dt\right)^{-1}$. When $a = 1$ and $b \geq 1$, we have

$$f_\beta(p) = b(1 - p)^{b-1},$$

where b is used to measure the strength of evidence against the individual null hypothesis. The c.d.f. of P is

$$F_\beta(p) = 1 - (1 - p)^b.$$

Let S be the probability integral transformation of P, that is

$$S = F_\beta(P) = 1 - (1 - P)^b.$$

It can be shown that $S \sim U(0, 1)$. Here

$$P = F_\beta^{-1}(S) = 1 - (1 - S)^{1/b}$$

is used to simulate P, where $b = 1$ for a true individual hypothesis and $b > 1$ for a false hypothesis. Then a set of p-values can be simulated when there are L false individual hypotheses. Without loss of generality, let $b_1 = \cdots = b_L > 1$ and $b_{L+1} = \cdots = b_n = 1$. The detailed procedure is described as follows.

1. Generate n i.i.d. uniformly distributed random numbers, $s_1, s_2, \ldots s_n$ on $[0, 1]$.

2. Convert s_i to p_i by

$$p_i = F_\beta^{-1}(s_i) = 1 - (1 - s_i)^{1/b}$$

   for $i = 1, 2, \ldots, n$.

3. Apply the CCP test to the p-values generated in step 2.

4. Repeat steps $1 - 3$ M times. Count the number of times, denoted by N, that H_0 is rejected. The power of the CCP test is then given by N/M.

3 Monte Carlo simulation

Now we conduct Monte Carlo simulations to explore size and power of the CCP test. We consider four p-value combination methods: Fisher's, Simes', Stouffer's and Tippett's methods.

Our simulations are performed in Matlab.

3.1 Size consideration

When combining Tippett's and Simes' methods, the size of the CCP test, $g(\gamma; n)$, can be determined either analytically by equation (8) or numerically as described in Section 2.2.2. The values of $g(\gamma; n)$ are displayed in Figure 1 for $n = 2, 5, 10, 20, 40, 80, 160, 500$ and $\gamma \in (0, 0.2]$. The estimated $g(\gamma; n)$ from numerical simulations is in accordance with that derived from the analytic formula for each n.

The simulations are also conducted for combinations of Fisher's and Simes' methods. As illustrated in Figure 2, $g(\gamma; n)$ is between γ and 2γ for different n. Furthermore, $g(\gamma; n)$ is not a linear function of γ. To approximate $g(\gamma; n)$, a second order Taylor polynomial is preferred.

The CCP test is fully specified after γ is determined. The Monte Carlo simulation described in Section 2.2.2 is used to obtain γ, for $\alpha = 0.01$, 0.05, 0.10 and $n = 2$, 5, 10, 20, 40, 80, 160, 500 (Tables 1 and 2). In Table 1, we also include the exact values of γ obtained by solving equation (8). The errors of the numerical estimates are within 0.2% of the exact values, implying that the Monte Carlo simulation results are reliable and acceptable.

As γ measures the correlation between two p-value combination methods under the joint null hypothesis, we also observe the following from Tables 1 and 2:

- In general, the correlation between any two p-value combination methods increases as α increases, but decreases as n increases;
- Tippett's and Simes' methods are almost perfectly correlated, echoing earlier results in Remark 2.2;
- Fisher's and Stouffer's methods are highly correlated;

- The correlation between Simes' (Tippett's) and Fisher's (Stouffer's) methods is very low as n becomes large.

Remark 3.1. *The last observation justifies Neuhäuser (2003)'s combination of Simes' method and truncated product method (TPM) at the individual significance level* $\alpha/2$. *Since TPM with a relatively large truncation point* $\tau = 0.7$ *is close to Fisher's method, we conjecture that the correlation between Simes' method and TPM with large* τ *is very low, i.e.,* $\rho \approx 0$. *According to equation (6),* $\gamma \approx \alpha/2$, *that is,* H_0 *is rejected at the significance level* α *if either Simes' or TPM yields a rejection at the level* $\alpha/2$.

Remark 3.2. *The CCP test is easy to carry out. When combining any two of the four p-value combination methods considered in this paper, one can refer to Tables 1 and 2 for the values of* γ. *Otherwise, one can follow the procedures described in Section 2.2.2 to find* γ *for a given significance level and the number of individual hypotheses to be combined.*

3.2 Power consideration

To construct a robust method, a multiplicity adjustment approach such as Simes' method that is powerful when there is only one or very few false individual null hypotheses, should be combined with a method that is powerful when many individual null hypotheses are false, e.g. Fisher's method. As an example, we choose to combine Fisher's and Simes' methods in accordance with the procedures described in Section 2.3. Figure 3 plots the power of Fisher's method, Simes' method and the CCP test, when $L = 1$ and $L = n$, respectively. When there is only one false individual null hypothesis (i.e. $L = 1$) and the evidence against the individual null is very strong (i.e. $b = 800$), the practitioner should employ Simes' method. But one would most certainly want to employ Fisher's method when all the individual null hypotheses are false (i.e. $L = n$) and the overall amount of evidence against the individual null is moderate (i.e. $b = 1.5$). Because in practice we are uncertain whether the number of false hypotheses is small or large, a risk-averse strategy is simply to apply the CCP test. In both cases the CCP test is not the most powerful method, but performs very well, tracking the better method very closely because the CCP test utilizes the weak correlation between these two methods and derives its power from Simes' method when $L = 1$ and Fisher's method when $L = n$. This simple exercise confirms our expectation that the CCP test insures against selecting an inferior individual method without sacrificing much power. Therefore, we strongly recommend its use in practice.

4 Empirical example

Purchasing Power Parity (PPP) is a key assumption in many theoretical models of international economics. A common way to test for evidence of long-run PPP is to test for the real exchange rate stationarity. Empirical evidence of PPP for the floating regime period (1973-1998) is, however, mixed. While several authors, such as Wu and Wu (2001) and Lopez (2008), found supporting evidence, others (Choi and Chue, 2007 and Pesaran, 2007) questioned the validity of PPP for this period. In this section, we use the CCP test to investigate if the real exchange rates are stationary among a group of countries within the Organization for Economic Cooperation and Development (OECD).

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

$$q_{it} = s_{it} + p_{us,t} - p_{it}, i = 1, \dots, n; t = 1, \dots, T, \tag{11}$$

where s_{it} is the logarithm of the nominal exchange rate of the ith country's currency in terms of US dollars, $p_{us,t}$ and p_{it} are the logarithm of consumer price indices in the US and country i, respectively. We use quarterly data from the first quarter of 1973 to the second quarter of 1998 for 23 OECD countries ($n = 23, T = 102$), as listed in Table 3. All data are obtained from the IMF's International Financial Statistics.[1]

We test the joint null hypothesis $H_0 : q_{it}$ is nonstationary for all i against the alternative $H_A : q_{it}$ is stationary for at least one i for $i = 1, \ldots, n$. This heterogeneous alternative is particularly appropriate for testing PPP, since there are no theoretical grounds for the imposition of the homogeneity hypothesis under PPP, cf. Maddala and Wu (1999). As first pointed out by O'Connell (1998), real exchange rates are dependent induced by the numeraire country and other likely sources of correlation, such as world-wide shocks. Panel unit root tests can lead to spurious results if a positive cross-section dependence exists and is ignored. To deal with this problem, many tests have been developed in the literature, see, for example, Breitung and Pesaran (2008) for a recent review. In this paper we use Bai and Ng (2004)'s panel analysis of nonstationarity in idiosyncratic and common components (PANIC) approach to remove the common factors from the data such that the defactored residuals are independent across panel units. The PANIC approach decomposes q_{it} in the following way:

$$q_{it} = c_i + \lambda_i' F_t + e_{it},$$

where F_t is an $r \times 1$ vector of common factors that induce correlation across panel units, λ_i is an $r \times 1$ vector of factor loadings and e_{it} is an idiosyncratic error. To conduct the unit root test, we first replace e_{it} by $\hat{e}_{it}$, the accumulated principal components estimator of Δe_{it}, and then run the following regression

$$\Delta \hat{e}_{it} = \phi_i \hat{e}_{i,t-1} + \sum_{j=1}^{k_i} \varphi_{ij} \Delta \hat{e}_{i,t-j} + u_{it}. \tag{12}$$

The testing of H_0 can be implemented as a test of the restriction that $\phi_i = 0$ for all i. We conduct the augmented Dickey-Fuller (ADF) test, which is the individual t-test for testing

[1] 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.

$\phi_i = 0$ in equation (12), and get the corresponding p-values.[2] Note that the PANIC residual-based ADF test statistics, or p-values, should be independent. To check this, we compute the pairwise cross-section correlation coefficients of the residuals from the above individual ADF regressions, $\hat{\rho}_{ij}$. The simple average of these correlation coefficients is calculated, cf. Pesaran (2004), as

$$\bar{\hat{\rho}} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \hat{\rho}_{ij}.$$

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

$$CD = \sqrt{\frac{2T}{n(n-1)}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \hat{\rho}_{ij}.$$

Under the null of no cross-section dependence, $CD \sim N(0,1)$. For PANIC residuals, $\bar{\hat{\rho}}$ is estimated as -0.01 and the CD statistic does not reject the null at the 1% level. Thus we can safely proceed to pool the individual p-values under the assumption that they are independent, cf. Westerlund and Larsson (2009) and Bai and Ng (2010).

Table 3 shows the estimation results. The joint null hypothesis H_0 cannot be rejected at $\alpha = 5\%$ according to Simes' and Tippett's methods since $p_{(1)} = 0.01 > \frac{\alpha}{N} = 0.002$ and $p_{(1)} = 0.01 > 1 - (1-\alpha)^{1/N} = 0.002$. However, H_0 is rejected by Fisher's and Stouffer's methods, since they result in very small p-values, as indicated in Table 3. Now turning to the CCP test, we first consider the combination of Simes' and Fisher's methods, that is, H_0 is rejected if either of the methods individually yields a rejection at the significance level γ. According to Table 2, we use $\gamma = 0.0287$. Since the p-value of Fisher's method equals 0.01, which is less than the significance level 0.0287, H_0 is rejected by Fisher's method and thus the CCP test. When combining other pairs of methods that give conflicting results, e.g.

[2] The computation details are given in Bai and Ng (2004) and the MATLAB codes are available in Ng's website.

Simes and Stouffer, Tippett and Fisher, Tippett and Stouffer, we reach the same rejection decision. In summary, the results from the CCP tests unanimously reject the unit root null hypothesis and provide strong evidence in favor of PPP during the floating regime period.

5 Concluding remarks

In this paper we have investigated the impact of uncertainty over the number of false null hypotheses on commonly used p-value combination methods and proposed a new procedure that attempts to retain good power properties despite such an uncertainty. Based on a simple union of rejections decision rule, the proposed test controls for the correlation between the two individual methods and simultaneously exploits the superior power of the two in different situations under the alternative. Our Monte Carlo studies have shown that the power of the CCP test is reasonably close to the best achievable power, while at the same time controls the size at the level of α.

As illustrated by our empirical example, the CCP test avoids the arbitrary decision about which method to use when the results from individual p-value combination methods contradict. Hence the test developed here is rather general and can be applied to any testing problems for which several tests have been proposed. Examples include testing for panel cointegration, testing for structural breaks - to name just a few.

The CCP test can also be generalized to combine more than two p-value combination methods or test statistics in a similar fashion to Harvey, et al. (2011), where they combined four standard unit root tests to deal with uncertainty over both the trend and initial condition.

The power of the CCP test can be further improved once additional information about the proportion of false individual hypotheses is obtained, for example, in the form of pre-testing as suggested in Breitung (2009). We leave it for our future research.

6 Appendix: Proof of theorem 2.3

For each $n \geq 1$ note that $\zeta = 1-(1-\gamma)^{1/n} > \frac{\gamma}{n}$. Now we claim that $\zeta < \frac{2\gamma}{n}$ when $0 < \gamma < 0.2$ for each $n \geq 1$.

For $n = 1$, it is clear that $\zeta = \gamma < 2\gamma$. For each $n \geq 2$, we define

$$f_n(\gamma) := \zeta - \frac{2\gamma}{n} = 1 - (1-\gamma)^{1/n} - \frac{2\gamma}{n}.$$

When $\gamma < 1 - \frac{1}{2}(\frac{1}{2})^{\frac{1}{n-1}}$, $f_n'(\gamma) = \frac{1}{n}(1-\gamma)^{\frac{1}{n}-1} - \frac{2}{n} < 0$. Hence $f_n(\gamma)$ strictly decreases on $(0, 0.2)$ since $0 < \gamma < 0.2 < 1 - \frac{1}{2}(\frac{1}{2})^{\frac{1}{n-1}}$. Then $\zeta - \frac{2\gamma}{n} = f_n(\gamma) < f_n(0) = 0$. Therefore $\zeta < \frac{2\gamma}{n}$. Thus

$$\frac{\gamma}{n} < \zeta < \frac{2\gamma}{n}, \quad \forall\, n \geq 1 \text{ and } 0 < \gamma < 0.2.$$

Now

$$\begin{aligned}
g(\gamma; n) &= Pr\left[\left(\bigcup_{i=1}^{n}\left(P_{(i)} \leq \frac{i\gamma}{n}\right)\right) \bigcup \left(\bigcup_{i=1}^{n}(P_i \leq \zeta)\right)\right] \\
&= Pr\left[\left(\bigcup_{i=1}^{n}\left(P_{(i)} \leq \frac{i\gamma}{n}\right)\right) \bigcup \left(P_{(1)} \leq \zeta\right)\right] \\
&= Pr\left[\left(P_{(1)} \leq \zeta\right) \bigcup \left(\bigcup_{i=2}^{n}\left(P_{(i)} \leq \frac{i\gamma}{n}\right)\right)\right] \\
&= 1 - Pr\left[\left(P_{(1)} > \zeta\right) \bigcap \left(\bigcap_{i=2}^{n}\left(P_i > \frac{i\gamma}{n}\right)\right)\right].
\end{aligned}$$

Since the joint density function of the ordered p-values $P_{(1)}, \ldots, P_{(n)}$ is $n!$,

$$
\begin{aligned}
g(\gamma; n) &= 1 - \int_{\gamma}^{1} \int_{\frac{(n-1)\gamma}{n}}^{p_{(n)}} \cdots \int_{\frac{2\gamma}{n}}^{p_{(3)}} \int_{\zeta}^{p_{(2)}} n! dp_{(1)} dp_{(2)} \cdots dp_{(n-1)} dp_{(n)} \\
&= 1 - \left(\int_{\gamma}^{1} \int_{\frac{(n-1)\gamma}{n}}^{p_{(n)}} \cdots \int_{\frac{2\gamma}{n}}^{p_{(3)}} \int_{\frac{\gamma}{n}}^{p_{(2)}} n! dp_{(1)} dp_{(2)} \cdots dp_{(n-1)} dp_{(n)} \right. \\
&\quad \left. - \int_{\gamma}^{1} \int_{\frac{(n-1)\gamma}{n}}^{p_{(n)}} \cdots \int_{\frac{2\gamma}{n}}^{p_{(3)}} \int_{\frac{\gamma}{n}}^{\zeta} n! dp_{(1)} dp_{(2)} \cdots dp_{(n-1)} dp_{(n)} \right) \\
&= 1 - Pr\left[\bigcap_{i=1}^{n} P_{(i)} > \frac{i\gamma}{n} \right] + \int_{\gamma}^{1} \int_{\frac{(n-1)\gamma}{n}}^{p_{(n)}} \cdots \int_{\frac{2\gamma}{n}}^{p_{(3)}} \int_{\frac{\gamma}{n}}^{\zeta} n! dp_{(1)} dp_{(2)} \cdots dp_{(n-1)} dp_{(n)} \\
&= Pr\left[\bigcup_{i=1}^{n} P_{(i)} \leq \frac{i\gamma}{n} \right] \\
&\quad + \int_{\gamma}^{1} n dp_1 \int_{\frac{(n-1)\gamma}{n}}^{p_{(n)}} \cdots \int_{\frac{2\gamma}{n}}^{p_{(3)}} \int_{\frac{\gamma}{n}}^{\zeta} (n-1)! dp_{(1)} dp_{(2)} \cdots dp_{(n-1)} dp_{(n)} \\
&= \gamma + n\left(\zeta - \frac{\gamma}{n}\right) \int_{\gamma}^{1} \int_{\frac{(n-1)\gamma}{n}}^{p_{(n)}} \cdots \int_{\frac{2\gamma}{n}}^{p_{(3)}} (n-1)! dp_{(2)} \cdots dp_{(n-1)} d_p(n) \\
&= \gamma + n\left(\zeta - \frac{\gamma}{n}\right) I(\gamma; n),
\end{aligned}
$$

where

$$
I(\gamma; n) = \begin{cases} 1 & n = 1, \\ \displaystyle\int_{\gamma}^{1} \int_{\frac{(n-1)\gamma}{n}}^{p_{(n)}} \cdots \int_{\frac{2\gamma}{n}}^{p_{(3)}} (n-1)! dp_{(2)} \cdots dp_{(n-1)} d_p(n) & n \geq 2. \end{cases}
$$

Now we claim that

$$
I(\gamma; n) = (1-\gamma)(1 - \frac{\gamma}{n})^{n-2}. \tag{13}
$$

It is easy to see that equation (13) holds when $n = 1$ and $n = 2$. By induction, we assume that equation (13) holds for $n = k$, that is $I(\gamma; k) = (1-\gamma)(1-\frac{\gamma}{k})^{k-2}$. When $n = k+1$,

$$I(\gamma; k+1) = \int_{\gamma}^{1} \int_{\frac{k\gamma}{k+1}}^{p_{(k+1)}} \cdots \int_{\frac{2\gamma}{k+1}}^{p_{(3)}} k! dp_{(2)} \cdots dp_{(k)} dp_{(k+1)}.$$

By using the change of variables, let $\widetilde{p_{(i)}} = \dfrac{p_{(i)}}{p_{(k+1)}}$, $i = 2, \ldots, k$. Then

$$\begin{aligned}
I(\gamma; k+1) &= \int_{\gamma}^{1} \int_{\frac{k\gamma}{k+1}}^{p_{(k+1)}} \cdots \int_{\frac{2\gamma}{k+1}}^{p_{(3)}} k! dp_{(2)} \cdots dp_{(k)} dp_{(k+1)} \\
&= \int_{\gamma}^{1} \left(\int_{\frac{\frac{k\gamma}{k+1}}{p_{(k+1)}}}^{1} \int_{\frac{\frac{(k-1)\gamma}{k+1}}{p_{(k+1)}}}^{\widetilde{p_{(k)}}} \cdots \int_{\frac{\frac{2\gamma}{k+1}}{p_{(k+1)}}}^{\widetilde{p_{(3)}}} k! p_{(k+1)}^{k-1} d\widetilde{p_{(2)}} \cdots d\widetilde{p_{(k-1)}} d\widetilde{p_{(k)}} \right) dp_{(k+1)} \\
&= \int_{\gamma}^{1} k p_{(k+1)}^{k-1} \left(\int_{\frac{\frac{k\gamma}{k+1}}{p_{(k+1)}}}^{1} \int_{\frac{\frac{(k-1)\gamma}{k+1}}{p_{(k+1)}}}^{\widetilde{p_{(k)}}} \cdots \int_{\frac{\frac{2\gamma}{k+1}}{p_{(k+1)}}}^{\widetilde{p_{(3)}}} (k-1)! d\widetilde{p_{(2)}} \cdots d\widetilde{p_{(k-1)}} d\widetilde{p_{(k)}} \right) dp_{(k+1)} \\
&= \int_{\gamma}^{1} k p_{(k+1)}^{k-1} \, I\left(\frac{k\gamma}{(k+1)p_{(k+1)}}; k \right) dp_{(k+1)} \\
&= \int_{\gamma}^{1} k p_{(k+1)}^{k-1} \left(1 - \frac{k\gamma}{(k+1)p_{(k+1)}} \right) \left(1 - \frac{\gamma}{(k+1)p_{(k+1)}} \right)^{k-2} dp_{(k+1)} \\
&= (1-\gamma) \left(1 - \frac{\gamma}{k+1} \right)^{k-1}.
\end{aligned}$$

By induction, equation (13) holds for all $n \geq 1$. Hence

$$g(\gamma; n) = \gamma + n\left(\zeta - \frac{\gamma}{n}\right) I(\gamma; n) = \gamma + n\left(\zeta - \frac{\gamma}{n}\right)(1-\gamma)\left(1 - \frac{\gamma}{n}\right)^{n-2}.$$

References

Bai, Jushan and Serena Ng, "A PANIC Attack on Unit Roots and Cointegration," *Econometrica*, 72 (2004), 1127-1177.

Bai, Jushan and Serena Ng, "Panel Unit Root Tests with Cross-section Dependence: A Further Investigation," *Econometric Theory*, 26 (2010), 1088-1114.

Breitung, Jörg, "Comment on Unit Root Testing in Practice: Dealing with Uncertainty over the Trend and Initial Condition," *Econometric Theory*, 25 (2009), 649-653.

Breitung, Jörg and M. Hashem Pesaran, "Unit Root and Cointegration in Panels," in Matyas, L. and Sevestre, P. (eds.) *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice* (2008), 279-322, Kluwer Academic Publishers.

Berk, Robert H. and Arthur Cohen, "Asymptotically Optimal Methods of Combining Tests," *Journal of the American Statistical Association*, 74 (1979), 812-814.

Choi, In and Timothy K. Chue, "Subsampling Hypothesis Tests for Nonstationary Panels with Applications to Exchange Rates and Stock Prices," *Journal of Applied Econometrics*, 22 (2007), 233-264.

Fisher, Ronald A., *Statistical Methods for Research Workers*, 4th edition (1932), Oliver and Boyd, London.

Harvey, David I., Stephen J. Leybourne and A.M. Robert Taylor, "Unit Root Testing in Practice: Dealing with Uncertainty over the Trend and Initial Condition," *Econometric Theory*, 25 (2009), 587-636.

Harvey, David I., Stephen J. Leybourne and A.M. Robert Taylor, "Testing for Unit Roots in the Presence of Uncertainty over Both the Trend and Initial Condition," *Journal of Econometrics*, (2011), forthcoming.

Hedges, Larry V. and Ingram Olkin, *Statistical Methods for Meta-Analysis*. (1985), Academic Press, San Diego.

Lopez, Claude, "Evidence of Purchasing Power Parity for the Floating Regime Period," *Journal of International Money and Finance*, 27 (2008), 156-164.

Loughin, Thomas M., "A Systematic Comparison of Methods for Combining p-values from Independent Tests," *Computational Statistics and Data Analysis*, 47 (2004), 467-485.

Maddala, G.S. and Shaowen Wu, "A Comparative Study of Unit Root Tests with Panel Data and A New Simple Test," *Oxford Bulletin of Economics and Statistics*, 61 (1999), 631-652.

Müller, Ulrich K., "Comment on Unit Root Testing in Practice: Dealing with Uncertainty over the Trend and Initial Condition," *Econometric Theory*, 25 (2009), 643-648.

Neuhäuser, Markus, "Tests for Genetic Differentiation," *Biometrical Journal*, 8 (2003), 974-984.

O'Connell, Paul G.J., "The Overvaluation of Purchasing Power Parity," *Journal of International Economics*, 44 (1998), 1-19.

Pesaran, M. Hashem, "General Diagnostic Tests for Cross Section Dependence in Panels," Cambridge Working Papers in Economics No. 435 (2004), University of Cambridge.

Simes, R.J., "An Improved Bonferroni Procedure for Multiple Tests of Significance," *Biometrika*, 73 (1986), 751-754.

Stouffer, Samuel A., Edward A. Suchman, Leland C. DeVinney, Shirley A Star. and Robin M. Williams Jr., *Studies in Social Psychology in World War II: The American Soldier. Vol. 1, Adjustment during Army Life*. (1949), Princeton, Princeton University Press.

Tippett, Leonard H.C., *The Method of Statistics*. (1931), Williams & Norgate, London.

Westberg, Margareta, "Combining Independent Statistical Tests," *The Statistician*, 34 (1985), 287-296.

Westerlund, Joakim and Rolf Larsson, "A Note on the Pooling of Individual PANIC Unit Root Tests," *Econometric Theory*, 25 (2009), 1851-1868.

Wu, Jyh-Lin and Shaowen Wu, "Is Purchasing Power Parity Overvalued?" *Journal of Money, Credit and Banking*, 33 (2001), 804-812.

Zaykin, Dmitri V., Lev A. Zhivotovsky, Peter H. Westfall, and Bruce S. Weir, "Truncated Product Method for Combining p-Values," *Genetic Epidemiology* 22 (2002), 170-185.

Table 1: Selection of γ: combination of Tippett's and Simes' methods

	α	$n=2$	5	10	20	40	80	160	500
	0.01	0.0100	0.0100	0.0100	0.0100	0.0099	0.0099	0.0100	0.0100
I	0.05	0.0494	0.0490	0.0489	0.0489	0.0488	0.0488	0.0489	0.0489
	0.10	0.0977	0.0964	0.0960	0.0956	0.0956	0.0954	0.0955	0.0956
	0.01	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100
II	0.05	0.0494	0.0491	0.0490	0.0489	0.0489	0.0489	0.0488	0.0489
	0.10	0.0977	0.0966	0.0963	0.0961	0.0960	0.0960	0.0960	0.0960

I: Simulated values; II: Exact values

Table 2: Selection of γ: combination of other methods

	α	$n=2$	5	10	20	40	80	160	500
	0.01	0.0071	0.0060	0.0055	0.0053	0.0048	0.0046	0.0046	0.0047
A	0.05	0.0385	0.0330	0.0304	0.0287	0.0270	0.0261	0.0257	0.0254
	0.10	0.0817	0.0708	0.0648	0.0606	0.0573	0.0551	0.0538	0.0527
	0.01	0.0071	0.0059	0.0055	0.0053	0.0048	0.0046	0.0046	0.0047
B	0.05	0.0383	0.0328	0.0303	0.0286	0.0270	0.0260	0.0256	0.0254
	0.10	0.0808	0.0697	0.0641	0.0601	0.0570	0.0549	0.0537	0.0526
	0.01	0.0058	0.0052	0.0051	0.0051	0.0047	0.0047	0.0047	0.0048
C	0.05	0.0325	0.0286	0.0273	0.0265	0.0257	0.0253	0.0252	0.0252
	0.10	0.0696	0.0605	0.0570	0.0550	0.0534	0.0525	0.0520	0.0517
	0.01	0.0058	0.0052	0.0051	0.0051	0.0047	0.0047	0.0047	0.0048
D	0.05	0.0324	0.0285	0.0272	0.0265	0.0256	0.0252	0.0252	0.0252
	0.10	0.0690	0.0601	0.0568	0.0549	0.0533	0.0524	0.0520	0.0517
	0.01	0.0079	0.0072	0.0070	0.0069	0.0062	0.0060	0.0061	0.0063
E	0.05	0.0410	0.0382	0.0374	0.0370	0.0361	0.0357	0.0357	0.0359
	0.10	0.0846	0.0793	0.0778	0.0771	0.0771	0.0768	0.0765	0.0763

A: Combination of Fisher's and Simes' methods
B: Combination of Fisher's and Tippett's methods
C: Combination of Stouffer's and Simes' methods
D: Combination of Stouffer's and Tippett's methods
E: Combination of Fisher's and Stouffer's methods

Table 3: PANIC residual-based unit root tests for 23 OECD real exchange rates

Panel (a): augmented Dickey-Fuller test		
Country	Lag	p-value
Australia	0	0.585
Austria	6	0.760
Belgium	0	0.365
Canada	0	0.905
Denmark	0	0.080
Finland	0	0.265
France	0	0.405
Germany	0	0.760
Greece	2	0.100
Iceland	5	0.250
Ireland	1	0.185
Italy	0	0.115
Japan	0	0.525
Korea	0	0.035
Luxembourg	0	0.650
Mexico	0	0.035
Netherlands	2	0.075
Norway	0	0.010
Portugal	0	0.205
Sweden	0	0.430
Switzerland	0	0.520
Turkey	0	0.435
United Kingdom	1	0.120
Panel (b): p-values for panel unit root tests		
	Test	p-value
	Fisher's method	0.010
	Stouffer's method	0.003

Figure 1: Size of the CCP test: combining Tippett's and Simes' methods

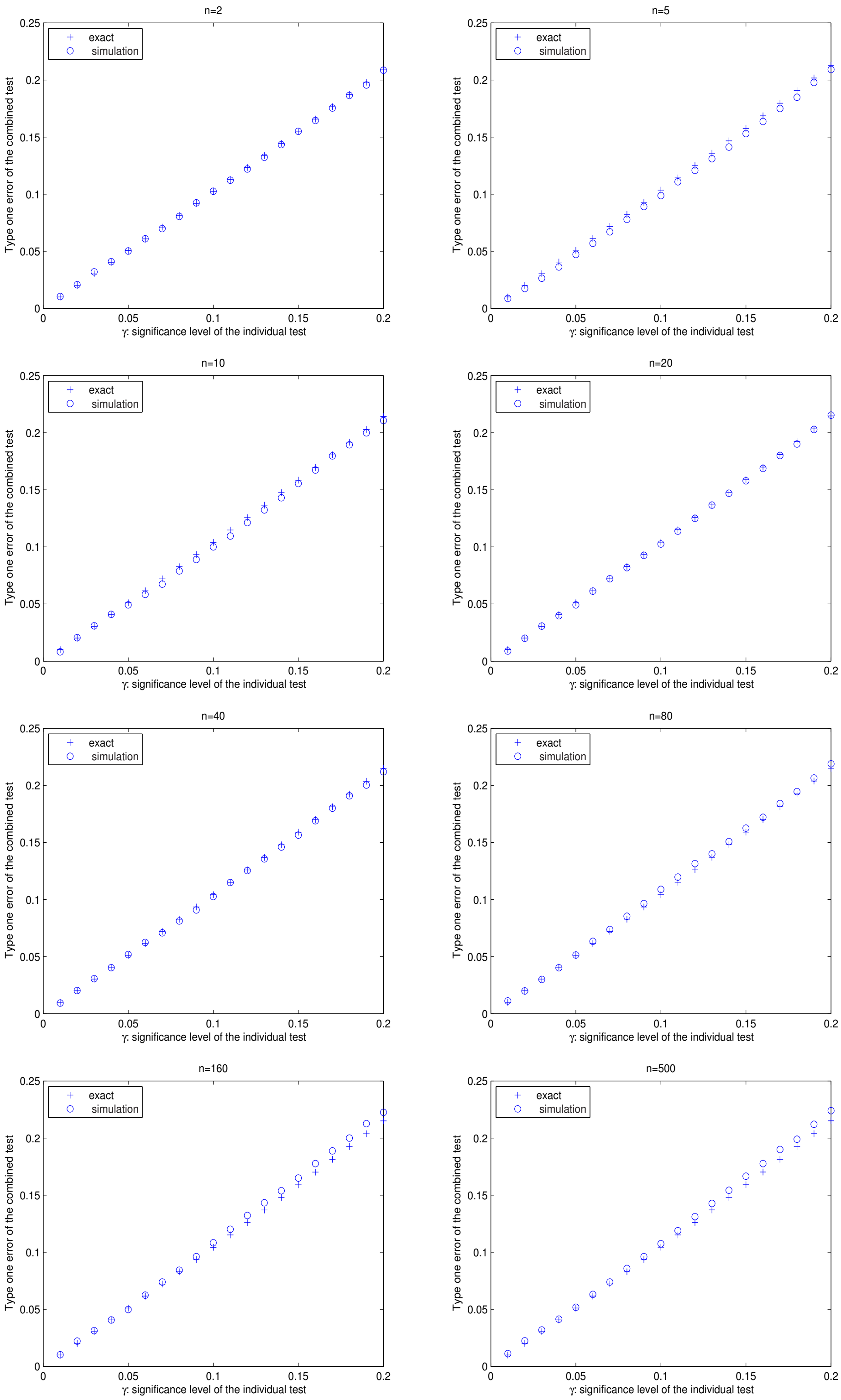

Figure 2: Size of the CCP test: combining Fisher's and Simes' methods

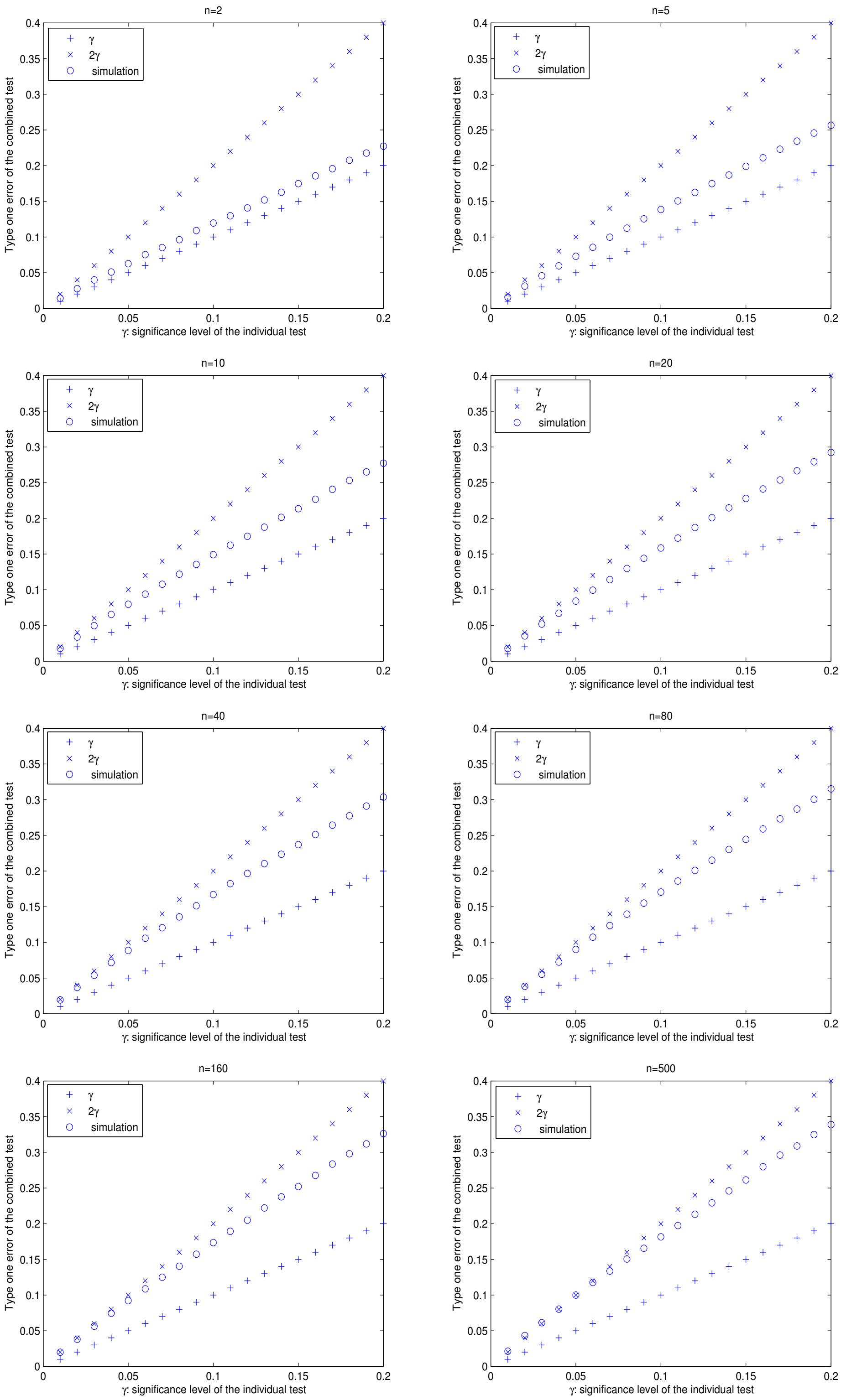

Figure 3: Power of the CCP test: combining Fisher's and Simes' methods