A Simple Panel Unit Root Test by Combining Dependent P-values*

XUGUANG SHENG[a] and JINGYUN YANG[b]

[a]*Department of Economics, SUNY at Fredonia*
Fredonia, NY 14063, USA.
(e-mail:sheng@fredonia.edu)
[b]*Department of Medicine, Harvard Medical School*
Boston, MA 02115, USA.
(e-mail: jingyuny@gmail.com)

Abstract

This paper proposes a simple panel unit root test based on Zaykin et al.'s (2002) truncated product method. The test is powerful in cases where there are only a few large p-values, and is robust to a certain degree of cross-section dependence. Monte Carlo evidence shows good size and power properties relative to existing p-value combination tests. Unlike the previous tests, the new test allows to make stronger claims in the event of rejection of the null hypothesis. The proposed test is applied to a panel of 27 OECD real exchange rate series as well as to a group of inflation density forecasts in the SPF data.

JEL classification: C12; C33

Keywords: Density Forecast, Panel Data, P-value, PPP, Truncated Product Method, Unit Root.

*We have benefited from discussions with Lan Cheng, Joachim Hartung, John Jones, Terrence Kinal, Xiaomei Li, James MacKinnon and Dmitri Zaykin. The usual disclaimer applies.

1 Introduction

Recently, there has been a growing interest in testing for unit roots in macroeconomic panels.[1] This is largely attributed to the advances in the panel unit root studies that provide reliable inference in the presence of cross-section dependence. O'Connell (1998) considered a GLS-based unit root test for homogeneous panels. Phillips and Sul (2003), Bai and Ng (2004), Moon and Perron (2004) and Pesaran (2007) used dynamic factor modeling for the same purpose. Chang (2004) developed bootstrap methods for panel unit root tests. These so-called "second generation" panel unit root tests are reviewed by Breitung and Pesaran (2008).

In this paper we adopt a different approach to testing a unit root in the panel by combining dependent p-values. Being widely used in meta-analysis, the p-value combination tests were introduced to panel unit root literature independently by Maddala and Wu (1999) and Choi (2001). The approaches most closely related to the one proposed in this paper are by Demetrescu et al. (2006) and Hanck (2008). Demetrescu et al. (2006) demonstrated that in the context of panel unit root test Hartung's modified inverse normal method was robust to certain deviations from the constant correlation assumption. Hanck (2008) found that Simes test had good size and power properties compared to other second-generation panel unit root tests. 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; (iii) it can be carried out for any unit root test derived; and (iv) it is a more balanced approach, because unlike test statistics, p values are usually identically distributed.

Our proposed test is based on Zaykin et al.'s (2002) truncated product method (TPM), which has been widely used in biostatistics.[2] It takes the product of the p-values less than some pre-specified cut-off value, and gains power in cases where there are only a few large p-values. We extend the test to allow for a certain degree of cross-section correlation in the panel. In a systematic comparison of the proposed test with other combination

[1]See, for example, testing for unit root in the long-term interest rate (Hassler and Tarcolea, 2005), stock indices (Choi and Chue, 2007), real exchange rate (Lopez, 2008) and output growth (Hanck, 2009).

[2]When this paper was written, about 146 papers cited their work, based on a computer search from *Google Scholar*.

methods, the empirical size of the TPM is reasonably close to the nominal size for moderate and large T. When quite a few series are stationary in the panel, the TPM tends to be most powerful among all tests considered here. The proposed test has an additional advantage in that it allows one to make stronger claims. When applying the TPM, rejection of the null hypothesis leads one to declare that there is at least one false hypothesis among the ones resulting in p-values less than some pre-specified value. Other combination procedures state that, in the event of rejection, there is at least one false hypothesis among *all* N hypotheses tested. As a by-product, setting the cut-off value equal to 1 results in a modified Fisher test that controls for dependence among a set of p-values.

Application of the combination tests to real exchange rate data does not provide strong evidence in favor of purchasing power parity for the floating regime period (1973-1998). In another application, we test the null hypothesis that forecast precision, if perceived properly, should contain a unit root, as implied by the Bayesian learning model (Lahiri and Sheng, 2008, 2009). Based on a panel of density forecasts for inflation, our result from the TPM and modified Fisher test showed that professional forecasters as a group did not update their inflation forecast percision in a Bayesian way. However, the modified inverse normal test reached an opposite conclusion. This is because it uses *all* p-values and loses power when about 1/3 of p-values are close to 1 in this example. In contrast, by truncating, these large p-values are removed, thus providing more power for the TPM and modified Fisher test.

The plan of the paper is as follows. Section 2 briefly reviews three main methods of combining p-values in the literature. In Section 3, the truncated product method is introduced and then extended to the case of dependent p-values in the context of panel unit root test. Small sample performance of the proposed test is investigated in Section 4 using Monte Carlo simulations. Section 5 provides two empirical applications and Section 6 concludes the paper. Some technical details are left in the Appendix.

2 Combining P-values: A Brief Review

In this section we briefly discuss three main methods of combining p-values in the context of panel unit root tests. Consider the model

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

The specification in equation (1) allows for heterogeneity in both the intercept and the slope, and is commonly used in the literature (Breitung and Pesaran, 2008). For convenience, it is often rewritten as

$$\Delta y_{it} = -\phi_i \mu_i + \phi_i y_{i,t-1} + \epsilon_{it}, \quad (2)$$

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

We are interested in testing the null hypothesis

$$H_0 : \phi_1 = \phi_2 = \cdots = \phi_N = 0$$

against the alternative

$$H_1 : \phi_1 < 0, \phi_2 < 0, \cdots, \phi_{N_0} < 0, N_0 \leq N.$$

Note that, while the null hypothesis implies that all the time series are unit-root nonstationary, the alternative states that there are 1 to N stationary units in the panel. Thus, a rejection of the null neither allows us to conclude that the entire panel is stationary nor does it provide information about the number of stationary units in the panel.[3]

Let S_{i,T_i} be a one-sided time-series unit root test applied to the ith unit of the panel in equation (2). Then the corresponding p-value is defined as $p_i = F(S_{i,T_i})$ if we reject the null of a unit root when a realized value of S_{i,T_i} is smaller than a constant, and $p_i = 1 - F(S_{i,T_i})$ instead when a realized value of S_{i,T_i} is greater than the constant. Here $F(\cdot)$ denotes the cumulative distribution function of S_{i,T_i}. We assume

Assumption 1 Under H_0, S_{i,T_i} has a continuous distribution function.

Assumption 2 When $i \neq j$, ϵ_{it} is independent of ϵ_{js} for all t and s.

Remark 1 Assumption 1 is a regularity condition that ensures a uniform distribution of the p-values. That is, under H_0: $p_i \sim U[0,1]$.

Remark 2 Assumption 2 is strong in that different units of the panel must be independent of each other. This assumption is likely to be violated

[3] Ng (2008) is an exception who provided a procedure to consistently estimate and test what fraction of the units in the panel have an autoregressive unit root.

in many macroeconomic applications. Later in the paper we relax this assumption to allow for a certain degree of dependence in the error terms.

We now present three p-value combination methods in the context of panel unit root tests.[4] The first test was proposed by Fisher (1932), who transformed the uniformly distributed p-values and combined them such that the resulting statistic had a chi-square distribution. The test statistic is

$$P = -2 \sum_{i=1}^{N} \ln(p_i), \tag{3}$$

which has a χ^2 distribution with $2N$ degrees of freedom under Assumptions 1 and 2. Being the most widely used in meta-analysis, this test procedure was introduced to the panel unit root tests by Maddala and Wu (1999) and modified to the case of infinite N by Choi (2001), relying on the assumption that the individual time series in the panel are cross-sectionally independent.

Two related modifications of Fisher method are given by Kost and McDermott (2002) and Makambi (2003). Kost and McDermott (2002) extended the Fisher test to the case of dependent p-values when the underlying test statistics were jointly distributed as multivariate t with a common denominator. Makambi (2003) proposed a weighted version of the Fisher test by assuming a positive and constant correlation among the underlying test statistics. In the next section, we propose a third modification, which proves to be more powerful than the original Fisher method.

Another often used procedure is inverse normal test, attributed to Stouffer et al. (1949), which transforms the p values via the standard normal distribution. The test statistic is defined as

$$Z = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Phi^{-1}(p_i), \tag{4}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Under Assumptions 1 and 2, $Z \sim N(0,1)$ for both finite and infinite N. Choi (2001) is the first paper that applied this method to panel unit root tests.

[4]For a systematic comparison of methods for combining p-values from independent tests, see Hedges and Olkin (1985) and Loughin (2004).

His simulation studies showed that the inverse normal method performed best among all combination tests considered in his paper.

To account for cross-section correlation, Hartung (1999) developed a modified inverse normal method by assuming a constant correlation across the probits t_i,

$$\mathrm{cov}(t_i, t_j) = \rho, \text{ for } i \neq j, \; i, j = 1, \ldots, N,$$

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

$$\hat{\rho}^{\star} = \max(-\frac{1}{N-1}, \hat{\rho}),$$

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

$$Z_m = \frac{\sum_{i=1}^{N} t_i}{\sqrt{N + N(N-1)[\hat{\rho}^{\star} + \kappa\sqrt{\frac{2}{N+1}}(1-\hat{\rho}^{\star})]}}, \tag{5}$$

where $\kappa = 0.1(1 + 1/(N-1) - \hat{\rho}^{\star})$ is a parameter designed to improve the small sample behavior of the test statistic. Under the null hypothesis, $Z_m \sim N(0, 1)$. In the context of panel unit root tests, Demetrescu et al. (2006) showed that the test was robust to certain deviations from the constant correlation assumption.

A third method is based on the ordered p-values, proposed by Simes (1986) as an improved Bonferroni procedure. Let $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(N)}$ be the ordered p-values for testing the null hypothesis applied to each time series. Then the joint hypothesis H_0 is rejected if

$$p_{(i)} \leq i\alpha/N, \tag{6}$$

for at least one $i = 1, \ldots, N$. [5] When the test statistics are independent, this procedure has a type I error equal to α. Importantly, Hanck (2008, 2009) showed that Simes test was robust to general patterns of cross-sectional dependence and to nonstationarity in the volatility process of the innovations of the time series in the panel. His simulation studies demonstrated that it had good size and power properties compared to

[5] According to Bonferroni, the null hypothesis H_0 is rejected only if $p_{(1)} \leq \alpha/N$. Thus, Simes adjustment offers more opportunities for rejection.

other second-generation panel unit root tests. However, a feature of this test is that the overall p-value cannot be smaller than the minimum p-value, $p_{(1)}$, while methods that combine p-values explicitly can give an overall p that is smaller than $p_{(1)}$.

3 Truncated Product Method

This section starts with the introduction of truncated product method for combing independent p-values. We then extend the method to the case of dependent p-values in the context of panel unit root tests.

In an influential article, Zaykin et al. (2002) suggested the use of the product of all those p-values that do not exceed some fixed value τ such that

$$W = \prod_{i=1}^{N} p_i^{I(p_i \leq \tau)}, \tag{7}$$

where $I(\cdot)$ is the indicator function.

Remark 3 Note that, obviously, the TPM with $\tau = 1$ is Fisher's original combination method. The ordinary Fisher product test, however, loses power in cases where there are a few large p-values. This can happen when there is a predominance of near-null effects. By truncating, these large components are removed, thereby providing more power, much like a "trimmed mean" gaining efficiency in the presence of outliers.

Remark 4 The TPM emphasizes smaller p-values, somewhat like the Simes and Šidák methods. To be precise, setting $\tau = \min p$ results in Šidák correction. However, as mentioned earlier, Simes and Šidák p-values can never be smaller than $p_{(1)}$, the minimum p-value. In contrast, the TPM p-value could be smaller than $p_{(1)}$, when there are several small and reinforcing p-values in the set.

Remark 5 The value of τ cannot be chosen based on observed p-values, and should be specified in advance. When there are only a few stationary series in the panel, a smaller truncation point is preferable. When almost all series are stationary, a larger truncation point will result in a more powerful test. But it is *a priori* unknown whether a few or almost all series are stationary in the panel. As suggested in Zaykin et al. (2002), a natural, although somewhat arbitrary choice of τ is α (commonly 0.05).

In cases when all p-values are independent, Zaykin et al. (2002) derived the distribution of W under the joint null hypothesis by conditioning on the number, k, of the p_i's less than τ:

$$\begin{aligned}\Pr(W \le w) &= \sum_{k=1}^{N} \Pr(W \le w \mid k)\Pr(k) \\ &= \sum_{k=1}^{N} \binom{N}{k} (1-\tau)^{N-k} \\ &\times \left(w \sum_{s=0}^{k-1} \frac{(k \ln \tau - \ln w)^s}{s!} I(w \le \tau^k) + \tau^k I(w > \tau^k) \right) \quad (8)\end{aligned}$$

Note that, when N is large, the probability in equation (8) should be computed through a Monte Carlo algorithm described below.

Next, we modify the TPM to allow for a certain degree of correlation among the p-values. The procedure is as follows:

Step 1: Estimate the correlation matrix for the p-values. Let Σ be a non-degenerate correlation matrix for the vector of p-values, $\mathbf{R}$. If Σ is positive definite, then by Cholesky decomposition there exists a matrix $\mathbf{C}$ such that $\Sigma = \mathbf{C}\mathbf{C}^T$.

Step 2: Calculate the overall p-value based on the following Monte Carlo simulations.

a. Calculate $W_0 = \prod_{i=1}^{N} p_i^{I(p_i \le \tau)}$ for a sample of p_i's (N observed p-values). Set $A = 0$.

b. Generate independent p-values, $u_1^*, \ldots, u_N^*$, from a $U[0,1]$ distribution. These N p-values form the elements of the vector $\mathbf{R}^*$. The dependent p-values, $\mathbf{R} = (u_1, \ldots, u_N)$ are obtained by using a correlation invariant transformation (Zaykin et al. 2002, p.174), $\mathbf{R} = 1 - \Phi\{\mathbf{C}\Phi^{-1}(1 - \mathbf{R}^*)\}$.

c. Calculate $W = \prod_{i=1}^{N} u_i^{I(u_i \le \tau)}$.

d. If $W \le W_0$, increment A by one.

e. Repeat steps b-d B times.

f. The combined p-value is A/B.

Remark 6 The proposed method here has the advantage that N can be very large. It can also be easily modified to incorporate weights, w_i into the analysis as $W_0 = \prod_{i=1}^{N} p_i^{w_i I(p_i \leq \tau)}$ and $W = \prod_{i=1}^{N} u_i^{w_i I(u_i \leq \tau)}$, thus allowing tests of more precision to play a larger role.

Remark 7 This method requires that the correlation structure is known. In many cases, Σ is unknown and has to be estimated from the data. In the appendix we give such a method of calculating $\hat{\Sigma}$ from a set of p-values. The results by using the estimated correlation structure have to be reviewed with caution. This problem is similar to the one of the generalized least-squares technique, when the model is being pre-multiplied with the Cholesky factor obtained from $\hat{\Sigma}$.

Remark 8 We should point out that the proposed test allows one to make stronger claims. When applying the TPM, rejection of the null hypothesis leads one to declare that there is at least one false hypothesis among the ones resulting in p-values $\leq \tau$. Other combination procedures state that, in the event of rejection, there is at least one false hypothesis among *all* N hypotheses tested.

4 Monte Carlo Evidence

This section reports the simulation results of the finite sample size and power of the combination methods in panel unit root tests introduced in Sections 2 and 3.

We use the following data generating process

$$y_{it} = (1 - \alpha_i)\mu_i + \alpha_i y_{i,t-1} + \epsilon_{it}, \tag{9}$$

for $i = 1, \ldots, N$, $t = -50, -49, \ldots, T$ with the initial value $y_{i,-50} = 0$. The parameters are generated in the following way:

$$\mu_i \sim \text{i.i.d. } \mathcal{N}(0, 1)$$

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

We use two different regimes to generate cross-section correlation among the error terms ϵ_{it}.

Regime 1: Factor structure

$$\epsilon_{it} = \gamma_i f_t + \xi_{it} \tag{10}$$

$\gamma_i \sim$ i.i.d. $\mathcal{U}[-1, 3]$, $i = 1, \dots, N$

$f_t \sim$ i.i.d. $\mathcal{N}(0, \sigma_f^2)$, $t = 1, \dots, T$

$\xi_{it} \sim$ i.i.d. $\mathcal{N}(0, 1)$

Regime 2: Equicorrelation
Let $\varepsilon_t = (\epsilon_{1t}, \dots, \epsilon_{Nt})'$. Assume that $\varepsilon_t \sim \mathcal{N}(0, \Sigma)$, with

$$\Sigma = \begin{bmatrix} 1 & \omega & \cdots & \omega \\ \omega & 1 & \cdots & \omega \\ \vdots & \vdots & \vdots & \vdots \\ \omega & \omega & \cdots & 1 \end{bmatrix}$$

All of the above parameters, μ_i, α_i, γ_i are generated independently of each other and also of the error ξ_{it} and of the factor f_t. Moreover, f_t is generated independently of ξ_{it}.

Remark 9 The value of δ indicates the fraction of stationary series in the panel, varying in the interval 0 - 1 (0 gives the size of the test, whereas $\delta > 0$ gives the power of the test). As a result, changes in δ allow us to study the impact of the proportion of stationary series on the power of the tests considered here.

Remark 10 The factor structure in *regime 1* has been widely used in the literature with the dependence being driven by factors in the error terms. See, for example, Phillips and Sul (2003), Bai and Ng (2004), Moon and Perron (2004) and Pesaran (2007). With $\sigma_f^2 = 10$, we explore the properties of the tests under "high" cross-section correlation.

Remark 11 The equicorrelation structure in *regime 2* was advocated by O'Connell (1998) and adopted widely in panel unit root studies by Choi (2006), Demetrescu et al. (2006) and Hanck (2008). We investigate the size and power of the tests under "high" cross-section correlation with $\omega = 0.95$.

When $\delta = 0$, we explore the size of the tests. Choosing $\delta = 0.5$, or 0.9, we analyze the power of the tests under heterogeneous alternatives. We calculated N Augmented Dickey-Fuller (ADF) t-statistics. The p-values were then calculated using the response surfaces estimated in MacKinnon (1996).[6] The tests were one-sided with the nominal size set at 5%, and conducted for all combinations of N and $T = 20, 50, 100$, using 5000 replications per experiment.[7]

Size and power of the tests are reported in Tables 1 to 6.[8] The major findings of our experiments can be summarized as follows.

1. The empirical size of the TPM and modified inverse normal test is reasonably close to the nominal size 0.05 for relatively large T, and they are mildly oversized when T is small. Simes test performs best in terms of size, though it is slightly conservative under equicorrelation (Table 2), consistent with the findings in Hanck (2008). The performance of modified Fisher test is much better than the original one, which, being a first-generation test, shows severe size distortions throughout.

2. The power of all tests is lower in Table 4 (and Table 6) than that in Table 3 (and Table 5). This is not surprising, since fitting a constant leads to loss of power for the tests considered.

3. Compared to other three tests for both small and large T, the power of Simes test is disappointing here.[9] The modified inverse normal test delivers high power in the present setup, similar to the TPM and modified Fisher test. However, we have to emphasize that, although the power of the TPM is satisfactory compared with other combination procedures, it is not our intention to recommend its use for this reason. The more important point is that the alternative hypothesis of interest is different, as discussed in *Remark 8*.

---

[6] These are asymptotic p-values, since the finite-sample p-values are valid only for non-augmented Dickey-Fuller tests. See, MacKinnon (1996), p.614. We are grateful to James G. MacKinnon for kindly providing us with his codes.

[7] An additional $B = 1000$ Monte Carlo simulations were performed for the TPM to control for cross-section correlation of the error terms.

[8] The p-values were truncated to lie in the range [0.000001,0.999999], in order to avoid very extreme values affecting these test statistics, cf. Pesaran (2007).

[9] Considering its substantial size distortions, we do not report the power results for the original Fisher test.

4. As illustrated in Figures 1 and 2, in general the power of all tests increases with an increased proportion of stationary series in the panel, cf. Karlsson and Löthgren (2000) and Li (2009). Interestingly, Simes method is most powerful when only very few series are stationary ($\delta = 0.1$), but it is clearly dominated by other tests as the percentage of stationary series in the panel increases. Instead, the TPM is most powerful when quite a few series are stationary ($\delta = 0.2$ to 0.5). With increased stationarity of the series ($\delta > 0.6$), all of the three tests, namely Simes, TPM and modified Fisher, exhibit high power. Since it is *a priori* unknown whether a few or almost all series are stationary in the panel, there is no uniformly most powerful method of combining *p*-values.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

[Table 4 about here.]

[Table 5 about here.]

[Table 6 about here.]

[Figure 1 about here.]

[Figure 2 about here.]

5 Empirical Application

We now present the application of the proposed tests to two important questions in macroeconomics. First, we test for the stationarity of real exchange rates in a panel of OECD countries. Second, we investigate the null hypothesis that there is a unit root in the inflation forecast precision in a panel of density forecasts.

5.1 Purchasing Power Parity

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 (2001) and Lopez (2008), found supporting evidence, others (O'Connell, 1998, Choi and Chue, 2007 and Pesaran, 2007) questioned the validity of PPP for this period. In this subsection, we use the methods discussed in previous sections to investigate if the real exchange rates are stationary among a group of OECD countries.

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

$$q_{it} = s_{it} - p_{us,t} + p_{it}, \tag{11}$$

where s_{it} is the nominal exchange rate of the ith country's currency in terms of US dollar, $p_{us,t}$ 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 7.[10] All data are obtained from the IMF's International Financial Statistics.[11]

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

$$\Delta q_{it} = \mu_i + \rho_i q_{i,t-1} + \sum_{j=1}^{k_i} \varphi_{ij} \Delta q_{i,t-j} + \varepsilon_{it}, i = 1, \ldots, N; t = k_i + 2, \ldots, T. \tag{12}$$

The null and alternative hypotheses for testing PPP are

$$H_0 : \rho_i = 0 \text{ for all } i \tag{13}$$

$$H_1 : \rho_i < 0 \text{ for at least one } i. \tag{14}$$

Note that the alternative hypothesis here is less restrictive since it allows for different convergence rates toward PPP across countries. We used the

[10] 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.

[11] 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.

recursive t-test procedure to select the appropriate lag order for individual real exchange rate series.[12] As shown in Ng and Perron (1995), this sequential testing procedure had better size properties than those based on information criteria. The p-value of the corresponding ADF test statistic was calculated according to MacKinnon (1996). The selected lags and the p-values are reported in Table 7. The results from the left panel in Table 7 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 robust check, we investigated the impact of a change in numeraire on the results. The right panel in Table 7 reports the estimation results when the Deutchemark 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.

[Table 7 about here.]

As is well known, the ADF test has low power with a short time span. Reliance on long time series of data in order to increase the power of the single-series unit root tests has also been problematic due to regime changes and structural breaks in exchange rate. One popular solution is to explore the cross-section dimension. However, as originally pointed out by O'Connell (1998), panel unit root tests can also lead to spurious results if a positive cross-section dependence exists and is ignored. As a preliminary check, we computed the pairwise cross-section correlation coefficients of the residuals from the above individual ADF regressions, $\hat{\rho}_{ij}$. Following Pesaran (2007), we then constructed the simple average of these correlation coefficients as

$$\bar{\hat{\rho}} = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij}, \tag{15}$$

and the associated cross-section dependence (CD) test statistics

$$CD = \sqrt{\frac{2}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \sqrt{T_{ij}} \hat{\rho}_{ij}. \tag{16}$$

[12] Starting with an upper bound, k_{max}, on k, if the last included lag is significant, choose $k = k_{max}$. If not, reduce k 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, 1.645, is used to determine the significance of the last lag. We set k_{max} equal to 8 for quarterly data.

As reported in the bottom two lines of Table 7, the average cross-section error correlation coefficients is 0.396 and 0.513 when US dollar and German mark are considered as the numeraire, respectively. The CD statistics are highly significant. This result is in line with the findings of Choi and Chue (2007) and Pesaran (2007).

[Table 8 about here.]

The panel unit root test statistics are summarized in Tables 7 and 8 for Simes test and in Table 9 for other tests. Simes test does not reject the unit root null, regardless of which numeraire, US dollar or German mark, is used or which panel, all 27 countries or a subset of 20, is considered. However, the evidence is mixed, as illustrated by other test statistics in Table 9. Similar to Lopez (2008), our results are numeraire specific. For 27 OECD countries as a whole, we find substantial evidence against the unit root null with German mark but not with US dollar. For a subset of 20 OECD countries, the two tests - modified inverse normal and modified Fisher - reject the null at the 10% level, when US dollar is used as the numeraire. This result is consistent with Wu and Wu (2001, Table 2) and Lopez (2008, Table 2), who used exactly the same data set as ours. The TPM test does not show any rejection of the unit root null for this subset of countries. In summary, we do not find strong evidence in favor of PPP for the floating regime period, echoing the results in Choi and Chue (2007) and Pesaran (2007).

[Table 9 about here.]

5.2 Precision Updating

In their analysis of the term structure of macroeconomic forecasts, Lahiri and Sheng (2008, 2009) proposed a Bayesian learning model. One of their model implications is that forecast precision (i.e. the reciprocal of forecast uncertainty), if perceived properly, should contain a unit root. To the best of our knowledge, this proposition has never been tested, partly due to lack of a direct measure of forecast uncertainty. Using the density forecasts for inflation, here we discuss a direct test for it.

Following the terminology in Lahiri and Sheng (2009, p.10), the precision of individual i's belief is evolved according to the following equation:

$$a_{ith} = a_{it,h+1} + b_{ith}, \tag{17}$$

where a_{ith} is the precision of individual i's *posterior* belief in predicting annual inflation rate for the target year t and h quarters ahead to the end of the target year, and $a_{it,h+1}$ is the precision of his *prior* belief at $h+1$ quarters ahead to the end of the target year t. Here b_{ith} is individual i's perceived quality of public information, which measures the shock to his precision updating process. In Bloom's (2009) terminology, b_{ith} is called "uncertainty shocks".

The data in this study are taken from Survey of Professional Forecasters (SPF) that is provided by the Federal Reserve Bank of Philadelphia. A unique feature of the SPF data is that forecasters are also asked to provide density forecasts for inflation. Although the SPF began in 1968, for several reasons as stated in Engelberg et al. (2009), we restricted attention to data collected from the first quarter of 1992 to the second quarter of 2009. We studied the density forecasts for the annual inflation rate. Survey respondents make their first forecasts when there are 8 quarters to the end of the target year; that is, they start forecasting at the first quarter of the previous year, and their last forecast is reported at the fourth quarter of the target year. So the actual horizons for these forecasts are approximately from 8 quarters to 1 quarter. This fixed-target scheme enables us to study the evolution of forecast uncertainty over horizons as specified in equation (17). For the purpose of estimation, we eliminated observations for infrequent respondents. We focused on the "regular" respondents who participated in at least 50 percent of the time. This left us with 24 individuals, whose identification numbers are listed in Table 10.[13] The precision a_{ith} was calculated as the reciprocal of the variance of the density forecast reported by individual i.[14]

We first estimated individual Dickey-Fuller (DF) regressions:

[13] See Giordani and Söderlind (2003) for a detailed discussion on the specification and construction of the analytical sample, and hence not repeated here.

[14] In cases when individuals reported zero variance of the density forecasts, we put an upper limit of 120 on the precision of a_{ith}, since the largest precision available in our sample was 101. Though arbitrarily, we felt that it was better to keep these large precision numbers rather than throw them away, because they reflected 100% certainty underlying individuals' forecasts. More importantly, the original orders of forecast uncertainty were kept, since a precision of 120 indicated a high certainty than that with a precision of 101.

$$\Delta a_{ith} = \rho_i a_{it,h+1} + \varepsilon_{ith}, i = 1, \ldots, N; t = 1, \ldots, T; h = 1, \ldots, H, \qquad (18)$$

where $\Delta a_{ith} = a_{ith} - a_{it,h+1}$. The p-value of the corresponding DF test statistic for each individual was calculated according to MacKinnon (1996). The average pairwise cross-section correlation coefficient of the residuals from the above individual DF regression, $\bar{\hat{\rho}}$, is 0.07 and statistically significant at the conventional level (CD test statistic is 9.41). Now turning to panel unit root tests that account for this positive cross-section correlation, the null and alternative hypotheses for testing the stationarity of inflation forecast uncertainty are in the same form as stated in equations (13) and (14). The TPM, Simes and modified Fisher tests strongly reject the joint null hypothesis at the 5% significance level, but the modified inverse normal method fails to reject the null. To understand these differences, recall that modified inverse normal method uses all p-values and tends to lose power when there are a few large p-values, where in this example about 1/3 of p-values are close to 1. In contrast, by truncating, these large p-values are removed, thus providing more power for the TPM and modified Fisher test. Simes test is also powerful in this case, since there are only few stationary series in the panel.

To summarize, the evidence from panel data analysis shows that professional forecasters as a group did not update their inflation forecast precision in a Bayesian way. One possibility could be that survey measure of uncertainty does not represent the "true" or objective uncertainty correctly. Diebold et al. (1999) concluded that survey uncertainty overestimated the true values. However, Giordani and Söderlind (2003) reached an opposite conclusion. Further studies are warranted to explore the updating process of forecasters' subjective uncertainty.

[Table 10 about here.]

6 Conclusion

This paper proposes a new unit root test for panel data, based on Zaykin et al.'s (2002) truncated product method. The TPM takes the product of the p-values less than some pre-specified cut-off value, and gains power in cases where there are only a few large p-values. We extend the test to allow for a certain degree of cross-section correlation in the panel. As a by-product, setting the cut-off value equal to 1 results in modified Fisher test that controls for dependence among a set of p-values.

We conduct a systematic comparison of the proposed test with other combination methods - original Fisher test, modified Fisher test, modified inverse normal method and Simes test. Monte Carlo evidence shows that the empirical size of the TPM is reasonably close to the nominal size for moderate and large T. When quite a few series are stationary in the panel, the TPM tends to be most powerful. We have to point out that, although the power of the proposed test is satisfactory compared to other combination procedures, it is not our intention to recommend its use for this reason. More importantly, the alternative hypothesis of interest is different: in the event of rejection of the null hypothesis, the TPM states that there is at least one false hypothesis among the ones resulting in p-values less than some pre-specified value; while other combination procedures indicate that there is at least one false hypothesis among *all* N hypotheses tested.

Application of the combination tests to real exchange rate data does not provide strong evidence in favor of PPP for the floating regime period. As expected, our results are numeraire specific and also depend on the panels considered. The convergence toward PPP within 27 OECD countries appears relatively strong during 1973-1998, when the German mark is used as numeraire. In another application, we test the null hypothesis that forecast precision, if perceived properly, should contain a unit root, as implied by the Bayesian learning model (Lahiri and Sheng, 2008, 2009). Based on a panel of density forecasts for inflation, our result from the TPM and modified Fisher test shows that professional forecasters as a group did not update their inflation forecast precision in a Bayesian way. However, modified inverse normal test reaches an opposite conclusion. This is because it uses *all* p-values and loses power when about 1/3 of p-values are close to 1 in this example. In contrast, by truncating, these large p-values are removed, thus providing more power for the TPM and modified Fisher test.

Our testing approach can be extended in a number of directions. One obvious generalization is to incorporate weights, thus allowing tests of more precision to play a larger role. Another worthwhile extension would be to develop a bootstrap version of the current test that are robust to general forms of cross-section dependence in panel data, along the lines of

Chang (2004) and Giacomini et al. (2009). This issue is currently under investigation by the authors. Furthermore, the proposed approach can also be extended easily to test for panel cointegration.

7 Appendix

Here, we propose one method to calculate the correlation matrix $\hat{\Sigma}$ from a set of p-values, p_i, for $i = 1, \ldots, N$.

Consider two random variables X and Y. If $X \sim N(0,1)$, $Y \sim N(0,1)$ and the correlation between X and Y is ρ, then the joint probability density function for (X, Y) is $f(X, Y; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{X^2 - 2\rho XY + Y^2}{2(1-\rho^2)}}$, the density functions of X and Y are the same: $g(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$.

Let $X = \Phi^{-1}(p_i)$, $Y = \Phi^{-1}(p_j)$, then $p_i = \Phi(X)$, $p_j = \Phi(Y)$ and

$$E(p_i) = E(\Phi(X)) = \int_{-\infty}^{\infty} \Phi(X) g(X) dX = \frac{1}{2},$$

$$E(p_i^2) = E(\Phi^2(X)) = \int_{-\infty}^{\infty} \Phi^2(X) g(X) dX = \frac{1}{3},$$

$$Var(p_i) = E(p_i^2) - (E(p_i))^2 = \frac{1}{12}.$$

Similarly, $E(p_j) = \frac{1}{2}$, $Var(p_j) = \frac{1}{12}$. Since

$$\hat{\gamma}(p_i, p_j) = \frac{Cov(p_i, p_j)}{Var(p_i) \cdot Var(p_j)} = \frac{E(p_i p_j) - E(p_i)E(p_j)}{Var(p_i) \cdot Var(p_j)},$$

we need to compute $E(p_i p_j)$, which can be obtained through

$$E(p_i p_j) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Phi(X)\Phi(Y) f(X, Y; \rho) dX dY.$$

If the value of ρ is given, one can use any mathematical software to compute $E(p_i p_j)$. Then the value of $\hat{\gamma}(p_i, p_j)$ is followed.

Following Hartung (1999) and Demetrescu et al. (2006), we assume a constant correlation between the probits t_i and t_j

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

where $t_i = x = \Phi^{-1}(p_i)$ and $t_j = y = \Phi^{-1}(p_j)$. ρ is estimated in finite samples by

$$\hat{\rho}^{\star} = \max(-\frac{1}{N-1}, \hat{\rho}),$$

where $\hat{\rho} = 1 - \frac{1}{N-1} \sum_{i=1}^{N} (t_i - \bar{t})^2$ and $\bar{t} = \frac{1}{N} \sum_{i=1}^{n} t_i$.

References

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

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica* 77, 623-685.

Breitung, J. and Pesaran, M.H. (2008). 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*, 279-322, Kluwer Academic Publishers.

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

Choi, I. (2001). Unit Root Tests for Panel Data. *Journal of International Money and Finance* 20, 249-272.

Choi, I. (2006). Combination Unit Root Tests for Cross-sectionally Correlated Panels. In Corbae, D., Durlauf, S.N. and Hansen, B. (eds.), *Econometric Theory and Practice: Frontiers of Analysis and Applied Research*. Cambridge University Press, Cambridge, UK, 311-333.

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

Demetrescu, M., Hassler, U. and Tarcolea, A-I. (2006). Combining Significance of Correlated Statistics with Application to Panel Data. *Oxford Bulletin of Economics and Statistics* 68, 647-663.

Diebold, F.X., Tay, A.S. and Wallis, K.F. (1999). Evaluating Density Forecasts of Inflation: The Survey of Professional Forecasters. In Engle, R.F. and White, H. (eds.), *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger*. Oxford University Press, Oxford.

Engelberg, J., Manski, C. and Williams, J. (2009). Comparing the Point Predictions and Subjective Probability Distributions of Professional Forecasters. *Journal of Business and Economics Statistics* 27, 30-41.

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

Giordani, P. and Söderlind, P. (2003). Inflation Forecast Uncertainty. *European Economic Review* 47, 1037-1059.

Giacomini, R., Politis, D. and White, H. (2009). A Warp-Speed Method for Conducting Monte Carlo Experiments Involving Bootstrap Estimators. *Working paper*, Department of Economics, UCSD.

Hanck, C. (2008). Intersection Test for Panel Unit Roots. *Working paper*, Department of Economics and Econometrics, University of Groningen.

Hanck, C. (2009). Nonstationary-Volatility Robust Panel Unit Root Tests and the Great Moderation. *Working paper*, Department of Economics and Econometrics, University of Groningen.

Hartung, J. (1999). A Note on Combining Dependent Tests of Significance. *Biometrical Journal* 41, 849-855.

Hassler, U. and Tarcolea, A-I. (2005). Combining Multi-country Evidence on Unit Roots: The Case of Long-term Interest Rates. *Applied Economics Quarterly* 51, 181-189.

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

Karlsson, S. and Löthgren, M. (2000). On the Power and Interpretation of Panel Unit Root Tests. *Economics Letters* 6, 249-255.

Kost, J.T. and McDermott, M.P. (2002). Combining Dependent P-values. *Statistics and Probability Letters* 60, 183-190.

Lahiri, K. and Sheng, X. (2008). Evolution of Forecast Disagreement in a Bayesian Learning Model. *Journal of Econometrics* 144, 325-340.

Lahiri, K. and Sheng, X. (2009). Learning and Heterogeneity in GDP and Inflation Forecasts. Forthcoming in *International Journal of Forecasting*.

Li, X. (2009). On the Power of Panel Unit Root Tests: A Simulation Study. *Working paper*, Department of Economics, SUNY Albany.

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

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

MacKinnon, J.G. (1996). Numerical Distribution Functions for Unit Roots and Cointegration Tests. *Journal of Applied Econometrics* 11, 601-618.

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

Makambi, K.H. (2003). Weighted Inverse Chi-square Method for Correlated Significance Tests. *Journal of Applied Statistics* 30, 225-234.

Moon, H.R. and Perron, B. (2004). Testing for a Unit Root in Panels with Dynamic Factors. *Journal of Econometrics* 122, 81-126.

Ng, S. (2008). A Simple Test for Nonstationarity in Mixed Panels. *Journal of Business and Economic Statistics* 26, 113-127.

Ng, S. and Perron, P. (1995). Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag. *Journal of American Statistical Association* 90, 268-281.

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

Pesaran, M.H. (2007). A Simple Panel Unit Root Test in the Presence of Cross-section Dependence. *Journal of Applied Econometrics* 22, 265-312.

Phillips, P.C.B. and Sul, D. (2003). Dynamic Panel Estimation and Homogeneity Testing under Cross Section Dependence. *Econometrics Journal* 6, 217-259.

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

Stouffer, S.A., Suchman, E.A., DeVinney, L.C., Star, S.A. and Williams, R.M.Jr. (1949). *The American Soldier*, Vol. 1 - Adjustment during Army Life. Princeton, Princeton University Press.

Wu, J-L. and Wu, S. (2001). Is Purchasing Power Parity Overvalued? *Journal of Money, Credit and Banking* 33, 804-812.

Zaykin, D.V., Zhivotovsky, L.A., Westfall, P.H., and Weir, B.S. (2002). Truncated Product Method for Combining P-Values. *Genetic Epidemiology* 22, 170-185.

Table 1: Size of panel unit root tests with factor structure

N	Test	*No intercept*			*Intercept only*		
		T			T		
		20	50	100	20	50	100
20	S	0.075	0.037	0.036	0.103	0.049	0.039
	Z_m	0.106	0.079	0.070	0.121	0.078	0.073
	W	0.095	0.072	0.066	0.116	0.071	0.066
	P_w	0.109	0.084	0.076	0.128	0.082	0.076
	P	0.234	0.230	0.239	0.265	0.229	0.233
50	S	0.075	0.041	0.037	0.098	0.051	0.040
	Z_m	0.101	0.076	0.076	0.110	0.087	0.075
	W	0.099	0.074	0.072	0.104	0.075	0.062
	P_w	0.106	0.080	0.082	0.113	0.090	0.077
	P	0.278	0.277	0.299	0.294	0.286	0.291
100	S	0.082	0.054	0.036	0.118	0.056	0.037
	Z_m	0.097	0.090	0.075	0.122	0.076	0.064
	W	0.090	0.089	0.075	0.115	0.068	0.057
	P_w	0.103	0.095	0.078	0.126	0.081	0.067
	P	0.314	0.325	0.332	0.325	0.307	0.306

Note: Rejection rates of the panel unit root tests at nominal level $\alpha = 0.05$, using 5000 replications. S is Simes test as in Hanck (2008), Z_m is modified inverse normal test as in Demetrescu et al. (2006), W is the TPM method that accounts for the cross-section correlation, P_w is the modified Fisher test when $\tau = 1$ in the TPM, and P is the original Fisher test as in Maddala and Wu (1999).

Table 2: Size of panel unit root tests with equicorrelation

		No intercept			*Intercept only*		
N	Test	T			T		
		20	50	100	20	50	100
20	S	0.047	0.040	0.039	0.073	0.044	0.030
	Z_m	0.073	0.071	0.063	0.098	0.066	0.057
	W	0.080	0.078	0.070	0.100	0.066	0.058
	P_w	0.079	0.076	0.067	0.100	0.067	0.058
	P	0.233	0.262	0.278	0.270	0.246	0.251
50	S	0.056	0.032	0.027	0.074	0.051	0.027
	Z_m	0.082	0.063	0.064	0.106	0.082	0.062
	W	0.095	0.075	0.073	0.107	0.083	0.066
	P_w	0.091	0.071	0.068	0.106	0.084	0.065
	P	0.304	0.313	0.300	0.334	0.315	0.299
100	S	0.044	0.038	0.032	0.061	0.039	0.032
	Z_m	0.072	0.072	0.069	0.096	0.071	0.070
	W	0.080	0.079	0.074	0.099	0.074	0.072
	P_w	0.077	0.074	0.072	0.100	0.075	0.071
	P	0.322	0.340	0.331	0.338	0.338	0.348

Note: see Table 1.

Table 3: Power of panel unit root tests with factor structure and no intercept

			$\delta = 0.5$			$\delta = 0.9$		
N	Test	T	20	50	100	20	50	100
20	S		0.124	0.215	0.606	0.160	0.327	0.744
	Z_m		0.213	0.442	0.815	0.220	0.540	0.938
	W		0.209	0.430	0.860	0.211	0.487	0.911
	P_w		0.224	0.468	0.873	0.229	0.542	0.940
50	S		0.141	0.203	0.578	0.160	0.316	0.738
	Z_m		0.239	0.476	0.827	0.223	0.518	0.932
	W		0.234	0.466	0.892	0.207	0.480	0.913
	P_w		0.251	0.513	0.903	0.230	0.536	0.943
100	S		0.140	0.223	0.631	0.185	0.304	0.766
	Z_m		0.243	0.480	0.855	0.231	0.512	0.950
	W		0.247	0.474	0.909	0.217	0.459	0.929
	P_w		0.262	0.533	0.920	0.237	0.521	0.952

Note: see Table 1.

Table 4: Power of panel unit root tests with factor structure and intercept

			$\delta = 0.5$			$\delta = 0.9$		
N	Test	T	20	50	100	20	50	100
20	S		0.114	0.106	0.225	0.110	0.131	0.297
	Z_m		0.143	0.176	0.381	0.142	0.188	0.485
	W		0.143	0.174	0.400	0.136	0.171	0.447
	P_w		0.152	0.189	0.418	0.143	0.194	0.495
50	S		0.134	0.103	0.228	0.124	0.132	0.315
	Z_m		0.166	0.179	0.404	0.137	0.193	0.492
	W		0.163	0.183	0.428	0.132	0.178	0.465
	P_w		0.178	0.202	0.453	0.143	0.202	0.503
100	S		0.146	0.107	0.213	0.144	0.138	0.307
	Z_m		0.148	0.195	0.400	0.145	0.197	0.475
	W		0.152	0.199	0.427	0.136	0.186	0.448
	P_w		0.161	0.212	0.451	0.150	0.206	0.488

Note: see Table 1.

Table 5: Power of panel unit root tests with equicorrelation and no intercept

			$\delta = 0.5$			$\delta = 0.9$		
N	Test	T	20	50	100	20	50	100
20	S		0.099	0.187	0.508	0.121	0.233	0.639
	Z_m		0.198	0.392	0.738	0.200	0.461	0.895
	W		0.202	0.401	0.804	0.192	0.431	0.863
	P_w		0.209	0.410	0.807	0.205	0.468	0.899
50	S		0.118	0.166	0.484	0.114	0.234	0.611
	Z_m		0.234	0.404	0.758	0.208	0.462	0.890
	W		0.246	0.427	0.834	0.204	0.436	0.869
	P_w		0.250	0.444	0.834	0.220	0.475	0.897
100	S		0.104	0.156	0.472	0.114	0.223	0.600
	Z_m		0.218	0.418	0.772	0.203	0.458	0.908
	W		0.233	0.435	0.848	0.196	0.436	0.889
	P_w		0.231	0.458	0.850	0.206	0.467	0.911

Note: see Table 1.

Table 6: Power of panel unit root tests with equicorrelation and intercept

			$\delta = 0.5$			$\delta = 0.9$		
N	Test	T	20	50	100	20	50	100
20	S		0.101	0.083	0.206	0.085	0.103	0.240
	Z_m		0.146	0.157	0.364	0.126	0.191	0.435
	W		0.150	0.170	0.390	0.124	0.178	0.414
	P_w		0.156	0.170	0.398	0.128	0.196	0.440
50	S		0.093	0.069	0.158	0.082	0.094	0.223
	Z_m		0.145	0.166	0.338	0.119	0.191	0.432
	W		0.159	0.177	0.372	0.121	0.182	0.420
	P_w		0.158	0.183	0.377	0.124	0.196	0.439
100	S		0.088	0.070	0.156	0.086	0.081	0.216
	Z_m		0.138	0.159	0.358	0.135	0.173	0.463
	W		0.149	0.171	0.404	0.132	0.161	0.442
	P_w		0.147	0.178	0.398	0.136	0.175	0.470

Note: see Table 1.

Table 7: Augmented Dickey-Fuller tests for 27 OECD countries

US dollar real exchange rate				German mark 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
$\bar{\hat{\rho}}$		0.396				0.513	
CD		71.137				93.368	

Table 8: Augmented Dickey-Fuller tests for 20 OECD countries

US dollar real exchange rate				German mark real exchange rate			
Country	k	p-value	Simes criterion	Country	k	p-value	Simes criterion
New Zealand	8	0.008	0.003	Australia	3	0.012	0.003
Sweden	8	0.053	0.005	Canada	7	0.040	0.005
United Kingdom	7	0.055	0.008	Sweden	0	0.074	0.008
Finland	7	0.058	0.010	United States	4	0.148	0.010
Spain	8	0.061	0.013	New Zealand	0	0.171	0.013
Switzerland	4	0.071	0.015	Finland	6	0.232	0.015
France	4	0.080	0.018	Netherlands	1	0.415	0.018
Netherlands	4	0.099	0.020	Norway	7	0.417	0.020
Austria	4	0.102	0.023	Spain	0	0.459	0.023
Italy	4	0.103	0.025	France	0	0.564	0.025
Belgium	4	0.135	0.028	Italy	0	0.565	0.028
Germany	4	0.148	0.030	Belgium	0	0.618	0.030
Greece	4	0.150	0.033	Japan	5	0.656	0.033
Norway	7	0.167	0.035	United Kingdom	0	0.697	0.035
Denmark	3	0.206	0.038	Denmark	0	0.698	0.038
Ireland	7	0.235	0.040	Ireland	0	0.708	0.040
Japan	4	0.246	0.043	Austria	0	0.720	0.043
Portugal	8	0.332	0.045	Switzerland	8	0.733	0.045
Australia	3	0.386	0.048	Portugal	0	0.786	0.048
Canada	6	0.580	0.050	Greece	5	0.880	0.050
$\bar{\hat{\rho}}$		0.563				0.573	
CD		74.435				77.747	

Table 9: P-values of panel unit root tests for the real exchange rate

Tests	US dollar real exchange rate		German mark real exchange rate	
	N=27	N=20	N=27	N=20
Z_m	0.095	0.090	0.016	0.374
W	0.257	0.168	0.002	0.167
P_w	0.097	0.090	0.015	0.330

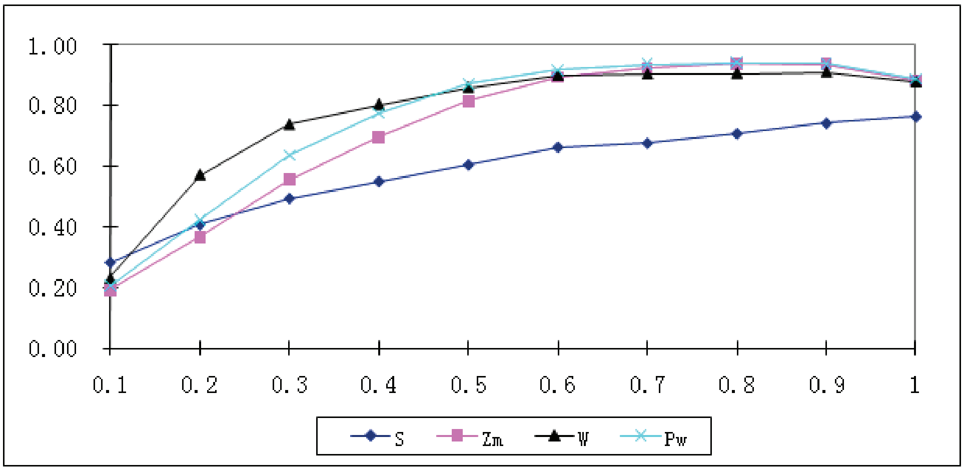
Note: Z_m is the modified inverse normal test as in Demetrescu et al. (2006), W is the TPM method that accounts for the cross-section correlation, and P_w is the modified Fisher test by setting $\tau = 1$ in calculating the TPM.

Table 10: Panel unit root test of inflation forecast uncertainty

Identification number	p-value	Simes criterion
407	0.0000	0.0021
508	0.0014	0.0042
510	0.0061	0.0063
65	0.0064	0.0083
411	0.0090	0.0104
456	0.0146	0.0125
420	0.0623	0.0146
426	0.2091	0.0167
20	0.3105	0.0188
446	0.5303	0.0208
512	0.6267	0.0229
84	0.6283	0.0250
507	0.6823	0.0271
421	0.8358	0.0292
504	0.9707	0.0313
428	0.9797	0.0333
472	0.9935	0.0354
433	0.9955	0.0375
483	0.9970	0.0396
431	0.9980	0.0417
484	0.9998	0.0438
463	0.9999	0.0458
99	0.9999	0.0479
439	1.0000	0.0500
Z_m	0.9990	
W	0.0000	
P_w	0.0010	

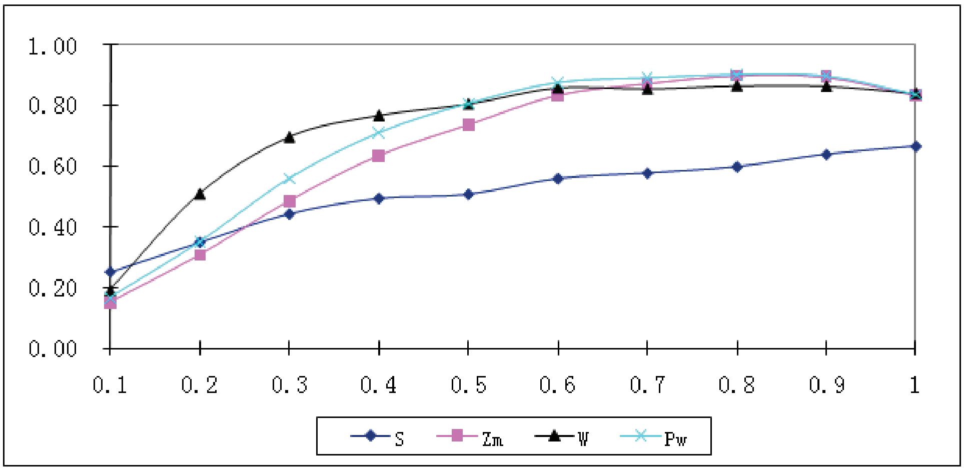
Note: Z_m is the modified inverse normal test as in Demetrescu et al. (2006), W is the TPM method that accounts for the cross-section correlation, and P_w is the modified Fisher test by setting $\tau = 1$ in calculating the TPM.

Figure 1: Power comparison with factor structure and no intercept ($N = 20$, $T = 100$)



Note: The horizontal axis shows the percentage of stationary series in the panel, and the vertical axis shows the power of the tests. S is Simes test as in Hanck (2008), Z_m is the modified inverse normal test as in Demetrescu et al. (2006), W is the TPM method that accounts for the cross-section correlation, and P_w is the modified Fisher test by setting $\tau = 1$ in calculating the TPM.

Figure 2: Power comparison with equicorrelation and no intercept ($N = 20$, $T = 100$)



Note: See Figure 1.