OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 75, 4 (2013) 0305-9049
doi: 10.1111/j.1468-0084.2012.00705.x

Truncated Product Methods for Panel Unit Root Tests*

XUGUANG SHENG† and JINGYUN YANG‡

†*Department of Economics, American University,Washington, DC 20016, USA*
(e-mail: sheng@american.edu)
‡*The Methodology Center, Pennsylvania State University, State College, PA 16801, USA*
(email: jxy174@psu.edu)

Abstract

This paper proposes two new panel unit root tests based on Zaykin *et al.* (2002)'s truncated product method. The first one assumes constant correlation between *P*-values and the second one uses sieve bootstrap to allow for general forms of cross-section dependence in the panel units. Monte Carlo simulation shows that both tests have reasonably good size and are powerful in cases of some very large *P*-values. The proposed tests are applied to a panel of real GDP and inflation density forecasts, resulting in evidence that professional forecasters may not update their forecast precision in an optimal Bayesian way.

I. Introduction

Recently, advances in panel unit root studies that provide reliable inference despite cross-section dependence have spurred interest in testing for unit roots in macroeconomic data.[1] O'Connell (1998) considered a GLS-based unit root test for homogeneous panels. Chang (2004) showed that this GLS procedure depends on nuisance parameters and proposed a corrective bootstrap approach. Phillips and Sul (2003), Bai and Ng (2004), Moon and Perron (2004) and Pesaran (2007) proposed dynamic factor models that allow the common factors to have different effects on cross-section units. For a review on these so-called "second generation" panel unit root tests, see Breitung and Pesaran (2008) and Gengenbach *et al.* (2010).

We propose new methods for panel unit root tests by combining dependent *P*-values. The *P*-value combination methods were introduced to panel unit root literature independently by Maddala and Wu (1999) and Choi (2001). Recent contributions include Demet-

*This paper was presented at the 16th International Panel Data Conference (2010) in Amsterdam, the Netherlands and 81st Southern Economic Association Annual Meeting (2011) in Washington, D.C. We thank Stefano Fachin, Christoph Hanck, Joachim Hartung, James MacKinnon, Serena Ng, Hashem Pesaran, Dmitri Zaykin and participants in the conference for helpful comments and suggestions. We also thank the editor Anindya Banerjee and two anonymous referees for the comments that have significantly improved the paper. Dr. Yang's research was supported by Award Number P50DA010075-15 from the National Institute on Drug Abuse. The usual disclaimer applies.
JEL Classification numbers: C12, C33.

[1]See, for example, testing for unit root in the long-term interest rate (Hassler and Tarcolea, 2005), real exchange rate (Pesaran, 2007; Hanck, 2011) and output growth (Choi, 2006; Hanck, 2010).

rescu *et al.* (2006) and Hanck (2011). Compared to combining test statistics, combining *P*-values has several advantages in that it allows different specifications for each panel unit, any unit root test derived and unbalanced panels.

Our proposed tests are based on Zaykin *et al.* (2002) 's truncated product method (TPM), which has been widely used in biostatistics, see Schmidt *et al.* (2008) and Seebacher and Glanville (2010). The TPM takes the product of the *P*-values below some pre-specified cut-off value, increasing power in cases of some very large *P*-values. Building on the TPM, we propose two panel unit root tests: a modified TPM that assumes a constant correlation among the *P*-values and a bootstrap TPM that allows for general forms of cross-section dependence in the panel units. Monte Carlo simulation provides evidence of good power with moderate and large T values for both tests, despite the slightly oversized modified TPM.

For an empirical example, we test the null hypothesis that forecast precision, if perceived properly, should contain a unit root, as implied by the Bayesian learning model developed in Lahiri and Sheng (2008). Based on a panel of density forecasts for real GDP and inflation during 1992–2009, we find evidence that some professional forecasters do not optimally update their forecast precision.

II. Truncated product method

Consider the model

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

The specification in equation (1) allows for heterogeneity in both the intercept and the slope. For convenience, equation (1) is rewritten as

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

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

$$H_0 : \phi_i = 0 \quad \text{for all } i \tag{3}$$

against the alternative

$$H_1 : \phi_i < 0, i = 1, \ldots, N_0; \quad \phi_i = 0, i = N_0 + 1, \ldots, N, \tag{4}$$

such that

$$\lim_{N\to\infty} \frac{N_0}{N} = \delta, 0 < \delta \leq 1. \tag{5}$$

Remark 1. Note that the null and alternative hypotheses can also be written as $H_0 : \delta = 0$ vs. $H_1 : \delta > 0$. Thus, rejection of the null can be interpreted as evidence for rejecting the unit root hypothesis for a non-zero fraction of panel units as $N \to \infty$.

Let S_{i,T_i} be a test statistic applied to the ith unit of the panel in equation (2). Then the corresponding *P*-value is defined as $p_i = F(S_{i,T_i})$, where $F(\cdot)$ denotes the cumulative distribution function (c.d.f.) of S_{i,T_i}. We assume

Assumption 1 (Uniformity). Under H_0, S_{i,T_i} has a continuous distribution function, which ensures a uniform distribution of the *P*-values.

Zaykin *et al.* (2002) suggested the use of the product of all those *P*-values that do not exceed some pre-specified value τ such that

$$W = \prod_{i=1}^{N} p_i^{I(p_i \leq \tau)}, \tag{6}$$

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

Remark 2. When some series in the panel are clearly nonstationary such that the resulting *P*-values are close to 1, traditional *P*-value combination methods may lose power because they could be dominated by these large *P*-values. By truncating, these large components are removed, thereby providing more power, much like how a "trimmed mean" gains efficiency in the presence of outliers.

When all *P*-values are independent, W has a closed form distribution, see Zaykin *et al.*(2002). However, the distribution of W is unknown when the independency assumption is violated. Motivated by the success of the TPM in biostatistics, below we extend the TPM to allow for cross-section dependence among the *P*-values in panel unit root tests.

Modified TPM

First, we modify the TPM, denoted by W_m, by assuming a constant correlation between the *P*-values. Although constant correlations may not hold in some empirical applications, W_m at least does not require the panel to be balanced. The procedure is as follows:

Step 1: Calculate W_m using equation (6).

Step 2: Estimate the correlation matrix, Σ, for *P*-values. Following Hartung (1999) and Demetrescu *et al.* (2006), we assume a constant correlation between the probits t_i and t_j,

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

where $t_i = \Phi^{-1}(p_i)$ and $t_j = \Phi^{-1}(p_j)$. ρ can be estimated by

$$\tilde{\rho} = \max\left(-\frac{1}{N-1}, \hat{\rho}\right),$$

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

Step 3: The distribution of W_m is calculated based on the following Monte Carlo simulations:

a. Draw pseudo-random probits from the normal distribution with mean zero and the estimated correlation matrix, $\hat{\Sigma}$, and transform them back through the standard normal c.d.f., resulting in N *P*-values, denoted by $\tilde{p}_1, \tilde{p}_2, \cdots, \tilde{p}_N$.
b. Calculate $\tilde{W} = \prod_{i=1}^{N} \tilde{p}_i^{I(\tilde{p}_i \leq \tau)}$.
c. Repeat steps 3a and 3b B times and get $\tilde{W}_l$, $l = 1, \ldots, B$.
d. The *P*-value for W_m is given by $p = \frac{1}{B}\sum_{l=1}^{B} I(\tilde{W}_l \leq W_m)$.

Remark 3. Note that the transformation marginally alters the correlation among *P*-values. However, as pointed out by Zaykin *et al.* (2002), the correlation is approximately invariant

under monotone transformations. Therefore, the correlation between the probits t_i and t_j should be roughly equal to the correlation between the P-values p_i and p_j.

Remark 4. We circumvent the problem of a degenerate correlation matrix Σ in the case for $T < N$ by applying Hartung (1999)'s proposal, thus allowing for cases of very large N.

Bootstrap TPM

To preserve the dependence structure among cross-section units, we extend Palm, Smeekes and Urbain (2008)'s bootstrap method by resampling entire cross sections of residuals. We make the following assumptions:

Assumption 2 (Linearity). The error term ϵ_{it} is given by a general linear process

$$\epsilon_{it} = \psi_i(L)e_{it}, \tag{7}$$

where $\psi_i(z) = \sum_{k=0}^{\infty} \psi_{ik} z^k$ and $\sum_{k=0}^{\infty} |\psi_{ik}| < \infty$ for $i = 1, \ldots, N$.

Assumption 3 (Dependency; see also Chang (2004) Assumption 1). Define the $N \times 1$ vector $\mathbf{e}_t \equiv (e_{1t}, \ldots, e_{Nt})'$ for $t = 1, \ldots, T$. Let $\mathbf{e}_t$ be a sequence of i.i.d. random variables such that $E\mathbf{e}_t = 0$, $E\mathbf{e}_t\mathbf{e}_t' = \mathbf{\Sigma}$ and $E\|\mathbf{e}_t\|^4 < \infty$, where $\|\cdot\|$ is the Euclidean norm.

Below we outline the necessary steps for conducting the bootstrap TPM, W^*.

Step 1: Calculate W^* using equation (6).

Step 2: Obtain the residuals from an Augmented Dickey–Fuller (ADF) regression

$$\hat{e}_{it} = y_{it} - \hat{\alpha}_i y_{i,t-1} - \sum_{j=1}^{J_i} \hat{\phi}_{ij} \Delta y_{i,t-j}, \tag{8}$$

where the lag order J_i is selected according to the modified AIC procedure by Ng and Perron (2001). Then form the time series residual vectors $\hat{\mathbf{e}}_t \equiv (\hat{e}_{1t}, \ldots, \hat{e}_{Nt})'$ for $t = 1, \ldots, T$.[2]

Step 3: Generate the $N \times 1$ vector $\mathbf{e}_t^* \equiv (e_{1t}^*, \ldots, e_{Nt}^*)'$ by resampling from the centered residual vectors $(\hat{\mathbf{e}}_t - T^{-1}\sum_{t=1}^{T} \hat{\mathbf{e}}_t)$, $t = 1, \ldots, T$. The bootstrap samples $\mathbf{e}_t^*$ constructed as such will preserve the cross-section dependence structure of the data, as pointed out by Maddala and Wu (1999).

Step 4: Generate ϵ_{it}^* recursively from e_{it}^* as

$$\epsilon_{it}^* = \sum_{j=1}^{J_i} \hat{\phi}_{ij} \epsilon_{i,t-j}^* + e_{it}^*. \tag{9}$$

Step 5: Impose the null of unit root to obtain bootstrap samples y_{it}^* as

$$y_{it}^* = y_{i,t-1}^* + \epsilon_{it}^*. \tag{10}$$

We set $y_{i,-50}^* = 0$ and run the recursion for 50 initial observations before using y_{it}^* to mitigate the effect of initial conditions.

[2]One can also use the approach in Chang and Park (2003) to resample from the first difference of y_{it} to form a difference-based bootstrap TPM.

Step 6: Based on the bootstrap sample y_{it}^{*}, calculate $\tilde{W}$, defined in equation (6).

Step 7: Repeat steps 3–6 B times and get $\tilde{W}_l$, $l = 1, \ldots, B$.

Step 8: The P-value for W^{*} is given by $p = \frac{1}{B}\sum_{l=1}^{B} I(\tilde{W}_l \leq W^{*})$.

III. Monte Carlo study

Initially, we consider dynamic panels with a common factor driving the cross-section dependence. The DGP is as in equation (1), where

$$\epsilon_{it} = \gamma_i f_t + \xi_{it}, \tag{11}$$

for $i = 1, \ldots, N$, $t = -50, -49, \ldots, T$ with the initial value $y_{i,-50} = 0$. The factor loading γ_i is drawn from a uniform distribution as $\gamma_i \sim$ i.i.d. $U[0, 3]$. The individual fixed effect μ_i, the common factor f_t and the error term ξ_{it} are independently drawn from normal distributions as $\mu_i \sim$ i.i.d. $\mathrm{N}(0, 1)$, $f_t \sim$ i.i.d. $\mathrm{N}(0, \sigma_f^2)$ and $\xi_{it} \sim$ i.i.d. $\mathrm{N}(0, 1)$. We explore the properties of the tests under cross-section independence with $\sigma_f^2 = 0$ (DGP 1) and under "high" cross-section dependence with $\sigma_f^2 = 10$ (DGP 2).

Remark 5. Under the null hypothesis of common and idiosyncratic unit roots, the DGP setup in Bai and Ng (2004) is equivalent to our DGP in (1) and (11) when $\alpha_i = 1$ for all i. However, these two DGPs are different in the case of a unit root in the common factor and near-unit roots in the idiosyncratic errors. See Banerjee, Marcellino and Osbat (2004) for a detailed description of this case of cross-unit cointegration.

Next we allow for serial correlation in the error terms. We consider a number of experiments where the errors ξ_{it} in equation (11) are generated either as an AR(1) process $\xi_{it} = \rho_i \xi_{i,t-1} + e_{it}$ (DGP 3), or as an MA(1) process $\xi_{it} = e_{i,t} + \lambda_i e_{i,t-1}$ (DGP 4), where $e_{it} \sim$ i.i.d. $\mathrm{N}(0, 1)$. We choose $\rho_i \sim$ i.i.d. $U[0.2, 0.4]$ or $U[-0.4, -0.2]$ and $\lambda_i \sim$ i.i.d. $U[0.2, 0.4]$ or $U[-0.4, -0.2]$. These DGPs are intended to check the behaviour of our tests under different types of serial correlation.

Finally, we consider spatial dependence as an alternative scenario of panel cross-section dependence. Following Baltagi, Bresson and Pirotte (2007), we consider two commonly used spatial error processes: the spatial autoregressive (SAR) and the spatial moving average (SMA). The SAR specification (DGP 5) for the $N \times 1$ error vector ϵ_t in equation (1) can be expressed as $\epsilon_t = (I_N - \theta_1 W_N)^{-1} \upsilon_t$, where W_N is a known $N \times N$ spatial weights matrix. θ_1 is the spatial autoregressive parameter and the error component υ_t is assumed to be distributed independently across a cross-section dimension with constant variance σ_υ^2. In contrast, the SMA specification (DGP 6) for the error vector ϵ_t can be expressed as $\epsilon_t = (I_N + \theta_2 W_N)\upsilon_t$, where θ_2 is the spatial moving average parameter. Without loss of generality, we let $\sigma_\upsilon^2 = 1$. We consider the spatial dependence with $\theta_1 = 0.8$ and $\theta_2 = 0.8$. We specify the spatial weight matrix W_N as a "1 ahead and 1 behind" matrix with the ith row ($1 < i < N$) of this matrix comprising of nonzero elements in positions $i + 1$ and $i - 1$. Each row of this matrix is normalized such that all non-zero elements are equal to 1/2.

For all of DGPs considered here, we choose

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

The value of δ indicates the fraction of stationary series in the panel, varying in the interval 0–1. When $\delta = 0$, we explore the size of the tests. Choosing $\delta = 0.1$, 0.5 and 0.9, we analyse the impact of the proportion of stationary series on the power of the tests. The tests are one-sided with the nominal size set at 5%, and conducted for all combinations of $N \in \{20, 50\}$ and $T \in \{20, 50, 100\}$. The results are obtained with MATLAB using $M = 2,000$ simulations. Within each simulation, additional $B = 1,000$ bootstrap replications are performed.

We calculate the ADF t-statistics. The number of lags in the ADF regressions is selected according to the modified AIC procedure suggested by Ng and Perron (2001). We obtain P-values of unit root tests using response surface regressions as provided by MacKinnon (1996).[3]

As a preliminary check, we compute the pairwise cross-section correlation coefficient, $\hat{\rho}_{ij}$, of the residuals from the ADF regressions. Following Pesaran (2004), we construct the average of these correlation coefficients as $\bar{\hat{\rho}} = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij}$ and the associated cross-section dependence (CD) test statistics as $CD = \sqrt{\frac{2T}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij}$, where $CD \sim \mathrm{N}(0,1)$ under the null of no cross-section dependence. When the null hypothesis is rejected, we use our proposed tests; if not, we use Zaykin *et al.* (2002)'s original W test. The average correlation coefficient is 0 under cross-section independence (DGP 1), between 3% to 22% under spatial dependence (DGP 5 and 6) and about 80% with a factor structure (DGP 2–4). Thus, our considered DGPs represent a wide range of cross-section dependence. The CD statistics reject the null and prompt us to use W_m and W^* in all cases except DGP 1.

Next, we report the size and power of the modified TPM and bootstrap TPM. For comparison, we also include some other commonly used panel unit root tests.[4] In the absence of clear guidance for the choice of τ, we try ten different cut-off values, ranging from 0.05, 0.1, 0.2, …, up to 0.9. We find that both original and proposed TPMs tend to have relatively small size distortions with a smaller τ, and that their power does not show any clear patterns. We only present the results for $\tau = 0.1$ because our simulation results are similar for $0.05 \leq \tau \leq 0.2$.[5]

With *no* cross-section dependence (not reported here), the P, S, W and CIPS tests yield good empirical size. The other tests are slightly undersized. In the presence of strong cross-section dependence (Tables 1 and 2), the P test shows severe size distortions, the W^* test exhibits good size properties, the K_a^* and K_b^* tests generally perform well but are, as expected, severely oversized when $N = 50$ and $T = 20$ because they are designed for cases of small N and large T values. The CIPS and S tests show size distortions for the case of negative serial correlation but perform reasonably well for other cases. The Z and W_m tests are slightly oversized but the P_e^c test is conservative most of the time.[6]

[3] See Cheung and Lai (1995) as an alternative.

[4] More specifically, we consider Maddala and Wu (1999)'s Fisher test (denoted by P), Demetrescu *et al.* (2006)'s modified inverse normal test (denoted by Z), Hanck (2011)'s Simes test (denoted by S), Pesaran (2007)'s CIPS test, Bai and Ng (2004)'s P_e^c test, Chang (2004)'s K_{OT}^* test (denoted by K_a^*) and Palm *et al.* (2008)'s modified K_{OT}^* test (denoted by K_b^*).

[5] The complete simulation results for the TPM with all candidate truncation points are available upon request.

[6] The results from negative AR serial correlation are qualitatively similar to those from negative MA serial correlation and the results from positive MA serial correlation are qualitatively similar to those from positive AR serial correlation in Table 2 and thus they are not reported here.

TABLE 1

Size and power of tests: factor structure and no serial correlation

	N	T	P	Z	S	$CIPS$	P_e^c	K_a^*	K_b^*	W_m	W^*
	20	20	0.239	0.078	0.034	0.036	0.033	0.047	0.027	0.081	0.065
		50	0.222	0.072	0.030	0.050	0.030	0.033	0.031	0.077	0.075
δ=0		100	0.236	0.071	0.034	0.045	0.040	0.029	0.032	0.074	0.080
	50	20	0.271	0.076	0.030	0.037	0.025	0.217	0.196	0.079	0.063
		50	0.292	0.071	0.022	0.042	0.019	0.101	0.080	0.073	0.073
		100	0.287	0.068	0.034	0.050	0.022	0.083	0.075	0.070	0.075
	20	20	0.228	0.084	0.043	0.032	0.024	0.033	0.026	0.093	0.067
		50	0.257	0.087	0.037	0.256	0.020	0.073	0.058	0.092	0.081
δ=0.1		100	0.280	0.094	0.070	0.621	0.024	0.135	0.142	0.100	0.084
	50	20	0.304	0.094	0.035	0.020	0.016	0.231	0.171	0.102	0.075
		50	0.311	0.080	0.037	0.312	0.020	0.147	0.133	0.085	0.083
		100	0.356	0.091	0.082	0.748	0.030	0.261	0.262	0.118	0.098
	20	20	0.254	0.110	0.040	0.008	0.058	0.035	0.030	0.120	0.065
		50	0.375	0.151	0.069	0.473	0.105	0.125	0.089	0.188	0.117
δ=0.5		100	0.634	0.345	0.191	0.980	0.135	0.330	0.346	0.408	0.328
	50	20	0.326	0.124	0.036	0.000	0.091	0.211	0.166	0.142	0.071
		50	0.456	0.152	0.063	0.583	0.158	0.264	0.224	0.184	0.145
		100	0.706	0.367	0.182	0.997	0.217	0.600	0.592	0.429	0.349
	20	20	0.296	0.117	0.051	0.018	0.088	0.060	0.039	0.120	0.092
		50	0.476	0.164	0.081	0.356	0.321	0.158	0.152	0.176	0.200
δ=0.9		100	0.832	0.460	0.275	0.958	0.637	0.463	0.461	0.467	0.532
	50	20	0.358	0.118	0.047	0.007	0.125	0.291	0.231	0.124	0.093
		50	0.556	0.177	0.083	0.466	0.456	0.467	0.449	0.189	0.210
		100	0.877	0.445	0.244	0.990	0.747	0.841	0.852	0.456	0.516

Note: Rejection rates of panel unit root tests at nominal level $\alpha = 0.05$. P is Maddala and Wu (1999)'s Fisher test, Z is Demetrescu *et al.* (2006)'s modified inverse normal test, S is Hanck (2011)'s Simes test, CIPS is Pesaran (2007)'s cross-sectionally augmented IPS test, P_e^c is Bai and Ng (2004)'s pooled test statistic on idiosyncratic components, K_a^* is Chang (2004)'s K_{OT}^* test, K_b^* is Palm *et al.* (2008)'s modified K_{OT}^* test, W_m is the modified TPM and W^* is the bootstrap TPM.

Table 3 reports the results with spatial dependence. Under SAR specification, the CIPS, P_e^c and W_m tests exhibit size distortions, mainly because spatial correlation is typically weak and not captured by a common factor or constant correlation assumption. Additionally, ignoring such a weak correlation leads to over-rejection, as shown by the result from the P test. The K_a^* and K_b^* tests are severely undersized. The W^* test is undersized for small T values, but the size distortion reduces as T becomes large. The S test performs reasonably well. Under SMA specification, while all bootstrap tests suffer from downward size distortions, the P, CIPS and W_m tests are slightly oversized.

In general, the power of all tests increases when T increases and when N increases as long as δ is fixed, which justifies the use of panel data in unit root tests. All tests become more powerful when δ increases, which is consistent with the findings in Karlsson and Löthgren (2000).

When comparing the different tests under varying scenarios, there is no dominant test. For example, the S test outperforms the others when very few series in the panel are stationary, the CIPS test performs the best under the factor structure, the P_e^c test is preferable under spatial dependence, and the K_a^* and K_b^* tests behave similarly and well under

TABLE 2

Size and power of tests: factor structure with serial correlation

	N	T	P	Z	S	$CIPS$	P_e^c	K_a^*	K_b^*	W_m	W^*
(a) Positive AR serial correlation											
	20	20	0.184	0.095	0.030	0.030	0.050	0.044	0.025	0.089	0.051
		50	0.197	0.083	0.030	0.023	0.032	0.056	0.051	0.084	0.063
δ=0		100	0.173	0.073	0.028	0.033	0.032	0.082	0.068	0.072	0.060
	50	20	0.265	0.100	0.038	0.025	0.033	0.226	0.181	0.100	0.054
		50	0.242	0.081	0.026	0.040	0.019	0.173	0.161	0.087	0.055
		100	0.241	0.076	0.025	0.045	0.027	0.200	0.192	0.077	0.068
	20	20	0.198	0.095	0.031	0.026	0.042	0.044	0.023	0.093	0.055
		50	0.222	0.086	0.035	0.089	0.026	0.078	0.063	0.089	0.063
δ=0.1		100	0.239	0.094	0.062	0.301	0.038	0.216	0.196	0.108	0.075
	50	20	0.259	0.096	0.034	0.019	0.027	0.194	0.164	0.097	0.054
		50	0.289	0.102	0.036	0.103	0.021	0.237	0.213	0.106	0.076
		100	0.325	0.102	0.055	0.401	0.040	0.328	0.414	0.112	0.076
	20	20	0.229	0.116	0.044	0.007	0.072	0.062	0.035	0.136	0.059
		50	0.320	0.130	0.050	0.272	0.113	0.196	0.149	0.158	0.111
δ=0.5		100	0.558	0.305	0.134	0.882	0.163	0.491	0.454	0.363	0.260
	50	20	0.274	0.106	0.032	0.002	0.101	0.200	0.146	0.125	0.053
		50	0.404	0.146	0.048	0.257	0.160	0.385	0.364	0.176	0.103
		100	0.654	0.334	0.120	0.963	0.249	0.786	0.783	0.411	0.303
	20	20	0.246	0.103	0.037	0.017	0.120	0.073	0.060	0.108	0.064
		50	0.428	0.158	0.060	0.152	0.386	0.281	0.215	0.165	0.166
δ=0.9		100	0.761	0.377	0.187	0.818	0.738	0.595	0.590	0.385	0.434
	50	20	0.327	0.129	0.034	0.013	0.158	0.244	0.201	0.137	0.089
		50	0.505	0.154	0.052	0.272	0.563	0.588	0.538	0.167	0.175
		100	0.829	0.393	0.169	0.967	0.855	0.935	0.934	0.410	0.455
(b) Negative MA serial correlation											
	20	20	0.280	0.087	0.063	0.134	0.039	0.030	0.018	0.094	0.077
		50	0.278	0.090	0.079	0.200	0.034	0.015	0.012	0.093	0.091
δ=0		100	0.276	0.082	0.096	0.213	0.035	0.015	0.014	0.081	0.088
	50	20	0.315	0.078	0.057	0.151	0.023	0.205	0.211	0.084	0.083
		50	0.344	0.085	0.097	0.256	0.021	0.053	0.042	0.089	0.087
		100	0.341	0.074	0.111	0.312	0.032	0.034	0.044	0.078	0.095
	20	20	0.281	0.092	0.055	0.080	0.027	0.031	0.020	0.098	0.079
		50	0.315	0.100	0.107	0.488	0.025	0.039	0.026	0.108	0.105
δ=0.1		100	0.337	0.110	0.156	0.819	0.026	0.082	0.090	0.135	0.106
	50	20	0.325	0.082	0.062	0.094	0.021	0.201	0.189	0.086	0.075
		50	0.352	0.080	0.118	0.589	0.027	0.092	0.079	0.094	0.082
		100	0.386	0.103	0.198	0.938	0.034	0.168	0.152	0.133	0.112
	20	20	0.305	0.110	0.070	0.010	0.065	0.047	0.025	0.129	0.089
		50	0.428	0.180	0.154	0.607	0.121	0.084	0.065	0.211	0.161
δ=0.5		100	0.682	0.405	0.324	0.995	0.144	0.262	0.266	0.468	0.412
	50	20	0.374	0.136	0.071	0.004	0.105	0.178	0.157	0.155	0.092
		50	0.502	0.184	0.173	0.757	0.170	0.176	0.155	0.225	0.175
		100	0.744	0.408	0.399	1.000	0.191	0.438	0.443	0.481	0.404
	20	20	0.338	0.124	0.075	0.035	0.065	0.046	0.041	0.132	0.103
		50	0.563	0.218	0.176	0.509	0.253	0.102	0.095	0.221	0.256
δ=0.9		100	0.837	0.525	0.438	0.978	0.525	0.344	0.362	0.519	0.554
	50	20	0.391	0.119	0.081	0.031	0.102	0.269	0.233	0.132	0.113
		50	0.605	0.218	0.236	0.715	0.358	0.362	0.316	0.229	0.236
		100	0.885	0.543	0.532	0.996	0.584	0.676	0.682	0.541	0.568

TABLE 3

Size and power of tests under spatial dependence

	N	T	P	Z	S	$CIPS$	P_e^c	K_a^*	K_b^*	W_m	W^*
(a) Spatial AR dependence											
	20	20	0.126	0.076	0.049	0.115	0.092	0.009	0.006	0.092	0.021
		50	0.135	0.070	0.046	0.118	0.073	0.008	0.007	0.079	0.040
δ=0		100	0.118	0.071	0.044	0.138	0.082	0.004	0.003	0.067	0.054
	50	20	0.117	0.051	0.052	0.096	0.071	0.013	0.013	0.092	0.014
		50	0.129	0.056	0.058	0.109	0.062	0.002	0.002	0.080	0.037
		100	0.125	0.056	0.047	0.097	0.088	0.001	0.001	0.080	0.048
	20	20	0.124	0.078	0.056	0.120	0.094	0.011	0.011	0.114	0.027
		50	0.147	0.092	0.060	0.131	0.091	0.011	0.011	0.130	0.052
δ=0.1		100	0.228	0.134	0.104	0.188	0.127	0.014	0.014	0.196	0.088
	50	20	0.145	0.067	0.053	0.081	0.076	0.011	0.012	0.137	0.021
		50	0.181	0.083	0.051	0.124	0.102	0.006	0.006	0.157	0.064
		100	0.318	0.154	0.106	0.172	0.173	0.007	0.006	0.308	0.166
	20	20	0.174	0.098	0.061	0.132	0.148	0.014	0.011	0.138	0.042
		50	0.361	0.195	0.086	0.240	0.265	0.035	0.034	0.227	0.156
δ=0.5		100	0.781	0.548	0.266	0.520	0.506	0.141	0.116	0.611	0.579
	50	20	0.208	0.095	0.071	0.105	0.155	0.017	0.018	0.165	0.036
		50	0.564	0.275	0.103	0.266	0.456	0.032	0.030	0.317	0.244
		100	0.970	0.822	0.314	0.667	0.849	0.186	0.210	0.847	0.891
	20	20	0.225	0.129	0.064	0.153	0.201	0.025	0.023	0.161	0.051
		50	0.621	0.270	0.133	0.454	0.583	0.088	0.071	0.273	0.284
δ = 0.9		100	0.984	0.721	0.434	0.922	0.957	0.474	0.415	0.721	0.887
	50	20	0.296	0.140	0.070	0.142	0.290	0.021	0.020	0.182	0.058
		50	0.880	0.321	0.122	0.544	0.878	0.109	0.116	0.319	0.487
		100	1.000	0.879	0.478	0.995	1.000	0.748	0.752	0.872	0.996
(b) Spatial MA dependence											
	20	20	0.070	0.049	0.044	0.072	0.060	0.024	0.023	0.070	0.017
		50	0.075	0.050	0.052	0.089	0.046	0.013	0.015	0.074	0.023
δ=0		100	0.077	0.051	0.042	0.088	0.053	0.019	0.020	0.074	0.041
	50	20	0.081	0.039	0.058	0.047	0.044	0.024	0.024	0.089	0.009
		50	0.078	0.036	0.043	0.062	0.035	0.012	0.010	0.071	0.017
		100	0.076	0.031	0.048	0.060	0.046	0.012	0.013	0.074	0.038
	20	20	0.080	0.054	0.050	0.066	0.055	0.025	0.021	0.102	0.024
		50	0.101	0.062	0.053	0.077	0.077	0.029	0.034	0.111	0.034
δ=0.1		100	0.185	0.121	0.097	0.132	0.125	0.056	0.058	0.212	0.107
	50	20	0.076	0.034	0.062	0.068	0.045	0.032	0.025	0.109	0.014
		50	0.133	0.058	0.050	0.085	0.088	0.029	0.025	0.146	0.033
		100	0.281	0.144	0.118	0.127	0.197	0.069	0.071	0.298	0.179
	20	20	0.125	0.078	0.055	0.082	0.101	0.031	0.038	0.121	0.032
		50	0.346	0.221	0.083	0.209	0.283	0.127	0.153	0.243	0.155
δ=0.5		100	0.862	0.668	0.280	0.525	0.667	0.568	0.573	0.700	0.689
	50	20	0.155	0.076	0.050	0.077	0.139	0.024	0.042	0.150	0.016
		50	0.600	0.330	0.098	0.239	0.576	0.213	0.215	0.354	0.232
		100	0.995	0.910	0.350	0.720	0.987	0.890	0.898	0.918	0.955
	20	20	0.187	0.118	0.067	0.109	0.177	0.055	0.054	0.151	0.048
		50	0.652	0.278	0.114	0.399	0.653	0.378	0.367	0.257	0.326
δ=0.9		100	0.999	0.801	0.498	0.953	0.990	0.952	0.968	0.802	0.946
	50	20	0.290	0.150	0.079	0.106	0.316	0.040	0.038	0.198	0.052
		50	0.942	0.344	0.130	0.535	0.956	0.572	0.548	0.320	0.544
		100	1.000	0.930	0.543	1.000	1.000	1.000	1.000	0.930	1.000

factor structure but not under SAR specification. The W_m, Z and W^* tests have comparable power and all perform well under a factor structure with negative serial correlation and spatial dependence.[7] We disregard the P test due to its immense size distortions under cross-section dependence.

IV. Empirical application

In the analysis of the term structure of macroeconomic forecasts, Lahiri and Sheng (2008) proposed a Bayesian learning model, implying that forecast precision, if perceived properly, should contain a unit root. We test this implication directly by using density forecasts for inflation and real GDP.

Following the terminology in Lahiri and Sheng (2008), the precision of individual i's belief is evolved according to the following equation:

$$a_{ith} = a_{it,h+1} + b_{ith}, \tag{12}$$

where a_{ith} is the precision of individual i's *posterior* belief in predicting the variable 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.

The data in this study are taken from Survey of Professional Forecasters (SPF), which uniquely includes density forecasts for inflation and real GDP. For several reasons as stated in Engelberg, Manski and Williams (2009), we restrict attention to data collected from the first quarter of 1992 to the second quarter of 2009. Due to the design of the survey, the actual horizons for these forecasts are approximately from 8 quarters to 1 quarter for a target year. This fixed-target scheme enables us to study the evolution of forecast precision over horizons. For the purpose of estimation, we eliminate observations by infrequent respondents, and focus on the "regular" respondents who participated in at least 50% of the forecast periods. This leaves us with 24 individuals, whose identification numbers are listed in Table 4.[8] The precision a_{ith} is calculated as the reciprocal of the variance of the density forecast reported by individual i.[9]

We first estimate individual DF regressions and then compute the pairwise cross-section correlation coefficient of the residuals. In our sample, the average of these correlation coefficients is estimated to be 0.07 for inflation and 0.09 for real GDP. The CD statistics, 9.41 for inflation and 11.70 for real GDP, strongly reject the null of no cross-section dependence for both variables.

[7]Also notable is that our residual-based bootstrap TPM is slightly more powerful than the difference-based bootstrap TPM, consistent with the findings in Palm *et al.* (2008).

[8]See Giordani and Söderlind (2003) for a detailed discussion on the specification and construction of the analytical sample.

[9]In cases when the variance of the density forecast for an individual is zero, we put an upper bound of 120 on the precision a_{ith}, since the largest precision in our sample is 101. Though arbitrarily, it is better to keep these large precision numbers rather than throw them away, because they reflect 100% certainty underlying individuals' forecasts. More importantly, the original order of forecast uncertainty is preserved, since a precision of 120 indicates a higher certainty than a precision of 101.

TABLE 4

Panel unit root tests of forecast precision

	Inflation			*Real GDP*		
ID	*DF statistics*	*P-value*	*Simes criterion*	*DF statistics*	*P-value*	*Simes criterion*
20	−0.94	0.298	0.019	−2.12	0.989	0.027
65	−2.72	0.009	0.008	6.97	1.000	0.038
84	−0.16	0.616	0.025	−1.15	0.219	0.008
99	3.62	1.000	0.044	2.81	0.998	0.033
407	−3.91	0.001	0.002	2.82	0.998	0.035
411	−2.60	0.012	0.010	3.30	1.000	0.040
420	−1.84	0.063	0.015	−0.56	0.460	0.015
421	0.55	0.827	0.029	−3.90	0.001	0.002
426	−1.21	0.201	0.017	−0.99	0.278	0.010
428	1.72	0.975	0.033	−2.25	0.027	0.004
431	2.59	0.996	0.042	−0.07	0.647	0.019
433	2.32	0.993	0.038	−0.35	0.547	0.017
439	4.01	1.000	0.046	1.12	0.926	0.025
446	−0.42	0.517	0.021	4.87	1.000	0.042
456	−2.43	0.018	0.013	5.21	1.000	0.044
463	3.55	1.000	0.048	4.20	1.000	0.046
472	2.18	0.990	0.035	0.95	0.903	0.023
483	2.46	0.995	0.040	2.25	0.992	0.029
484	3.31	1.000	0.050	5.28	1.000	0.048
504	1.55	0.965	0.031	0.93	0.900	0.021
507	0.00	0.670	0.027	−1.80	0.069	0.006
508	−3.19	0.003	0.004	−0.83	0.343	0.013
510	−2.73	0.009	0.006	4.64	1.000	0.050
512	−0.17	0.614	0.023	2.37	0.993	0.031
P	0.004				0.796	
Z	1.000				1.000	
W_m	0.000				0.023	

Note: The DF statistics are based on univariate AR(1) specification in the level of the variable without an intercept. The corresponding *P*-values are obtained using the response surfaces estimated in MacKinnon (1996). Simes criterion is calculated as $i\alpha/N$ based on ordered *P*-values for $i = 1, \ldots, N$. P is Maddala and Wu (1999)'s Fisher test, Z is Demetrescu *et al.* (2006)'s modified inverse normal test and W_m is the modified TPM. All statistics are calculated based on the same sample period, namely 1992:Q1-2009:Q2, using the density forecasts from Survey of Professional Forecasters. The significance level α is set at 0.05.

Now we turn to panel unit root tests that account for this positive cross-section dependence.[10] The joint null and alternative hypotheses are specified in equations (3) and (4). For inflation forecasts, the S, P and W_m tests reject the joint null hypothesis of non-stationarity in forecasters' precision updating process at the 5% significance level, but the Z test fails to reject the null. As for real GDP forecasts, the S and W_m tests show strong evidence of rejection, but the P and Z tests do not reject. To understand the mixed evidence against the null, recall that the Z test uses all *P*-values and tends to lose power when some *P*-values

[10]Note that Bai and Ng (2004)'s P_e^c test, Pesaran (2007)'s CIPS test and the bootstrap tests require balanced panels and are not calculated for this empirical example of unbalanced panel.

are very large. In this example, about 40% of the P-values are close to 1 for inflation and 60% for real GDP. In contrast, by truncating these large P-values are removed, increasing the power of W_m. The S test is also powerful in this case, since there are some very small and reinforcing P-values in the panel. Thus, the evidence from panel data analysis shows that in predicting real GDP and inflation, some professional forecasters do not update their forecast precision in an optimal Bayesian way. Using the approaches of Hanck (2009) or Moon and Perron (2012) may give insight to which forecasters do not behave optimally.

V. Conclusion

In this paper, we propose two panel unit root tests: the modified TPM, W_m, and the residual-based bootstrap TPM, W^*. The W_m test allows for an unbalanced panel and the W^* test is robust to general forms of cross-section dependence in the panel. We conduct a systematic comparison of the proposed tests with other commonly used panel unit root tests. Evidence from Monte Carlo simulations shows that both tests deliver good power with moderate and large T values. Although the W_m test is slightly oversized, the W^* test yields good empirical size. To illustrate the usefulness of the proposed tests, we apply them to a panel of real GDP and inflation density forecasts. The resulting evidence indicates that some professional forecasters do not update their forecast precision in an optimal Beyesian way.

Our 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 an adaptive TPM that optimizes the selection of the truncation point among a set of candidates.

Final Manuscript Received: April 2012

References

Bai, J. and Ng, S. (2004). 'A PANIC attack on unit roots and cointegration', *Econometrica*, Vol. 72, pp. 1127–1177.

Baltagi, B. H., Bresson, G. and Pirotte, A. (2007). 'Panel unit root tests and spatial dependence', *Journal of Applied Econometrics*, Vol. 22, pp. 339–360.

Banerjee, A., Marcellino, M. and Osbat, C. (2004). 'Some cautions on the use of panel methods for integrated series of macro-economic data', *Econometrics Journal*, Vol. 7, pp. 322–340.

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*, New York: Springer-Verlag Publisher, pp. 279–322.

Chang, Y. and Park, J. Y. (2003). 'A Sieve bootstrap for the test of a unit root', *Journal of Time Series Analysis*, Vol. 24, pp. 379–400.

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

Cheung, Y.-W. and Lai, K.S. (1995). 'Lag order and critical values of the augmented Dickey-Fuller Test', *Journal of Business and Economic Statistics*, Vol. 13, pp. 277–280.

Choi, I. (2001). 'Unit root tests for panel data', *Journal of International Money and Finance*, Vol. 20, pp. 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, UK. Cambridge University Press, pp. 311–333.

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*, Vol. 68, pp. 647–663.
Engelberg, J., Manski, C. and Williams, J. (2009). 'Comparing the point predictions and subjective probability distributions of professional forecasters', *Journal of Business and Economic Statistics*, Vol. 27, pp. 30–41.
Gengenbach, C., Palm, F. C. and Urbain, J. P. (2010). 'Panel unit root tests in the presence of cross-sectional dependencies: comparison and implications for modelling', *Econometric Reviews*, Vol. 29, pp. 111–145.
Giordani, P. and Söderlind, P. (2003). 'Inflation forecast uncertainty', *European Economic Review*, Vol. 47, pp. 1037–1059.
Hanck, C. (2011). 'Intersection test for panel unit roots', *Econometric Reviews*, DOI: 10.1080/07474938.2011.608058.
Hanck, C. (2009). 'For which countries did PPP hold? A multiple testing approach', *Empirical Economics*, Vol. 37, pp. 93–103.
Hanck, C. (2010). *Nonstationary-Volatility Robust Panel Unit Root Tests and the Great Moderation*. MPRA working paper No. 11988.
Hartung, J. (1999). 'A note on combining dependent tests of significance', *Biometrical Journal*, Vol. 41, pp. 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*, Vol. 51, pp. 181–189.
Karlsson, S. and Löthgren, M. (2000). 'On the power and interpretation of panel unit root tests', *Economics Letters*, Vol. 66, pp. 249–255.
Lahiri, K. and Sheng, X. (2008). 'Evolution of forecast disagreement in a Bayesian learning model', *Journal of Econometrics*, Vol. 144, pp. 325–340.
MacKinnon, J.G. (1996). 'Numerical distribution functions for unit roots and cointegration tests', *Journal of Applied Econometrics*, Vol. 11, pp. 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*, Vol. 61, pp. 631–652.
Moon, H.R. and Perron, B. (2004). 'Testing for a unit root in panels with dynamic factors', *Journal of Econometrics*, Vol. 122, pp. 81–126.
Moon, H. R. and Perron, B. (2012). Beyond panel unit root tests: using multiple testing to determine nonstationarity properties of individual Series', *Journal of Econometrics*, DOI: 10.1016/j.jeconom.2012.01.008.
Ng, S. and Perron, P. (2001). 'Lag length selection and the construction of unit root tests with good size and power', *Econometrica*, Vol. 69, pp. 1519–1554.
O'Connell, P. G. J. (1998). 'The overvaluation of purchasing power parity', *Journal of International Economics*, Vol. 44, pp. 1–19.
Palm, F. C., Smeekes, S. and Urbain, J.-P. (2008). 'Bootstrap unit-root tests: comparison and extensions', *Journal of Time Series Analysis*, Vol. 29, pp. 371–401.
Pesaran, M. H. (2004). *General Diagnostic Tests for Cross Section Dependence in Panels*. Cambridge Working Papers in Economics No. 435, University of Cambridge.
Pesaran, M. H. (2007). 'A simple panel unit root test in the presence of cross-section dependence', *Journal of Applied Econometrics*, Vol. 22, pp. 265–312.
Phillips, P. C. B. and Sul, D. (2003). 'Dynamic panel estimation and homogeneity testing under cross section dependence', *Econometrics Journal*, Vol. 6, pp. 217–259.
Schmidt, C., Gonzaludo, N. P., Strunk, S., Dahm, S., Schuchhardt, J., Kleinjung, F., Wuschke, S., Joost, H. G. and Al-Hasani, H. (2008). 'A meta-analysis of QTL for diabetes-related traits in rodents', *Physiological Genomics*, Vol. 34, pp. 42–53.
Seebacher, F. and Glanville, E. J. (2010). 'Low levels of physical activity increase metabolic responsiveness to cold in a rat (Rattus fuscipes)', *PLoS ONE*, Vol. 5, p. e13022.
Zaykin, D. V., Zhivotovsky, L. A., Westfall, P. H., and Weir, B. S. (2002). 'Truncated product method for combining P-values', *Genetic Epidemiology*, Vol. 22, pp. 170–185.