Research Program on Forecasting

Truncated Product Methods for Panel Unit Root Tests

Xuguang Sheng
American University
Sheng@american.edu

Jingyun Yang
Pennsylvania State University
jingyuny@gmail.com

RPF Working Paper No. 2013-004
http://www.gwu.edu/~forcpgm/2013-004.pdf

April 8, 2013

RESEARCH PROGRAM ON FORECASTING
Center of Economic Research
Department of Economics
The George Washington University
Washington, DC 20052
http://www.gwu.edu/~forcpgm

Truncated Product Methods for Panel Unit Root Tests*

XUGUANG SHENG[a] and JINGYUN YANG[b]

[a] *Department of Economics, American University*
Washington, DC 20016, USA.
[b] *The Methodology Center, Pennsylvania State University*
State College, PA 16801, USA.

Abstract

This paper proposes three 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 latter two use sieve bootstrap that allows for general forms of cross-section dependence in the panel units. Monte Carlo simulation shows that these tests have reasonably good size, are robust to varying degrees of cross-section dependence and are powerful in cases where there are some very large p-values. The proposed tests are applied to a panel of real GDP and inflation density forecasts and provide evidence that professional forecasters may not update their forecast precision in an optimal Bayesian way.

*Corresponding authors: Xuguang Sheng (email: sheng@american.edu) and Jingyun Yang (email: jingyuny@gmail.com).

Truncated Product Methods for Panel Unit Root Tests*

Abstract

This paper proposes new panel unit root tests based on Zaykin et al.'s (2002) truncated product method. The tests are powerful in cases where there are some very large *p*-values, and are able to detect departures from the null of unit root in the subset of units whose *p*-values are below a given threshold. Monte Carlo evidence shows that the new tests are robust to varying degrees of cross-section dependence and have good size and power properties relative to other commonly used panel unit root tests. The proposed tests are applied to a panel of real GDP and inflation density forecasts from the Survey of Professional Forecasters.

JEL classification: C12; C33

Keywords: Density Forecast, Panel Unit Root, P-value, Sieve Bootstrap, Truncated Product Method.

*This paper was presented at the 16th International Panel Data Conference (2010) in Amsterdam, the Netherlands. 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. 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. Chang (2004) showed O'Connell (1998)'s GLS procedure to depend on nuisance parameters and proposed a bootstrap approach for a correction. Phillips and Sul (2003), Bai and Ng (2004), Moon and Perron (2004) and Pesaran (2007) proposed dynamic factor models by allowing the common factors to have different effects on cross-section units. These so-called "second generation" panel unit root tests are reviewed by Breitung and Pesaran (2008).

In this paper we propose new methods for panel unit root test by combining dependent p-values. Being widely used in meta-analysis, the p-value combination methods 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 Hartung's modified inverse normal method was robust to certain deviations from the assumption of constant correlation among p-values. 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 can be carried out for any unit root test derived; and (iii) it can deal with unbalanced panels.

Our proposed tests are based on Zaykin et al.'s (2002) truncated product method (TPM), which has been widely used in biostatistics, cf. Schmidt et al. (2008), Moskvina et al. (2009) and Seebacher and Glanville (2010) among others. 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 some very large p-values. We extend the original TPM to allow for cross-section dependence in the panel units and accordingly develop three tests: modi-

[1] See, for example, testing for unit root in the long-term interest rate (Hassler and Tarcolea, 2005), real exchange rate (Pesaran, 2007; Hanck, 2008) and output growth (Choi, 2006; Hanck, 2010).

fied TPM by assuming a constant correlation among p-values, cf. Hartung (1999), difference-based and residual-based bootstrap TPMs, cf. Maddala and Wu (1999), Chang and Park (2003), Chang (2004) and Palm et al. (2008). The modified TPM has the advantage that it does not require the panel to be balanced but exhibit slight size distortions in most of time. The two bootstrap TPMs are robust to general forms of cross-section dependence and yield good empirical size, close to the 5% nominal level, especially under factor structure with positive serial correlation and under spatial autoregressive specification. All of the proposed tests deliver satisfactory power when T is large. Furthermore, they have an additional advantage that by truncating, an individual rejection is known to have occurred among the small p-values, rather than any of the N p-values.

As 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 the evidence that suggests that some professional forecasters do not update their forecast precision in an optimal Bayesian way.

The plan of the paper is as follows. Section 2 briefly reviews three main methods of combining p-values. In Section 3, the TPM is introduced and then extended to the case of dependent p-values. Small sample performance of the proposed tests is investigated in Section 4 using Monte Carlo simulations. Section 5 provides an empirical application and Section 6 concludes the paper.

2 Combining P-values: A Brief Review

Consider the model

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

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

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

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

We are interested in testing the null hypothesis

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

against the alternative

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

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 providing evidence of rejecting the unit root hypothesis for a non-zero fraction of panel units as $N \to \infty$. Ng (2008) proposes an estimator of the fraction of units under the null.*

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.

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

We now present three p-value combination methods in the context of panel unit root tests.[2] The first test was proposed by Fisher (1932) as

$$P = -2 \sum_{i=1}^{N} \ln(p_i), \tag{6}$$

which has a χ^2 distribution with $2N$ degrees of freedom. This 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) under cross-section independence.

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

Another often used procedure, attributed to Stouffer et al. (1949), is the inverse normal method that transforms the p values via the standard normal distribution. Choi (2001) is the first paper that applied this method to panel unit root tests. His simulation studies showed that the inverse normal method performed best among all combination tests considered in his paper.

To account for cross-section dependence, Hartung (1999) developed a modified inverse normal method by assuming a constant correlation across the probits t_i, where $t_i = \Phi^{-1}(p_i)$ and $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution. In the context of panel unit root tests, Demetrescu et al. (2006) showed that this method was robust to certain deviations from the assumption of constant correlation between probits.

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

$$p_{(i)} \leq \frac{i\alpha}{N}, \tag{7}$$

for at least one $i = 1, \ldots, N$. [3] When the test statistics are independent, this procedure has a type I error equal to α. Importantly, Hanck (2008, 2010) showed that Simes test was robust to general patterns of cross-section dependence and to nonstationarity in the volatility process of the innovations of the time series in the panel if the p-values are computed using the correct distribution function of Dickey-Fuller (DF) statistics.

3 Truncated Product Method

This section starts with the introduction of the TPM for combining 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 pre-specified value τ such that

[3] 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.

$$W = \prod_{i=1}^{N} p_i^{I(p_i \le \tau)}, \tag{8}$$

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

Remark 2. *Note that, obviously, the TPM with $\tau = 1$ is Fisher's original combination method, which, however, loses power in cases when there are some very large p-values. This can happen when some series in the panel are clearly nonstationary such that the resulting p-values are close to 1. Traditional p-value combination methods may lose power as they could be dominated by these large p-values. By truncating, these large components are removed, thereby providing more power, much like a "trimmed mean" gaining efficiency in the presence of outliers.*

Remark 3. *The TPM emphasizes smaller p-values, somewhat like the Simes and Šidák methods.*[4] *However, the p-value of Simes and Šidák methods can never be smaller than $p_{(1)}$, the minimum p-value. In contrast, the p-value of the TPM could be smaller than $p_{(1)}$, when there are several small and reinforcing p-values in the set.*

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 k, the number of the p_i's less than τ:

$$\begin{aligned}
\Pr(W \le w) &= \sum_{k=1}^{N} \Pr(W \le w \mid k)\Pr(k) \\
&= \sum_{k=1}^{N} \binom{N}{k} (1-\tau)^{N-k} \\
&\times \left(w \sum_{s=0}^{k-1} \frac{(k \ln \tau - \ln w)^s}{s!} I(w \le \tau^k) + \tau^k I(w > \tau^k) \right). \quad (9)
\end{aligned}$$

Note that the above distribution of the TPM is guaranteed only under the assumption of cross-section independence. The distribution of W is no longer valid and unknown when the independency assumption is violated. Next, we modify the TPM to allow for cross-section dependence among the p-values.

---

[4]Šidák method is derived under cross-section independence, and suggests rejecting the joint null hypothesis H_0 if $p_i \le 1 - (1-\alpha)^{\frac{1}{N}}$ for at least one $i = 1, \ldots, N$.

3.1 The TPM Relying on a Constant Correlation Assumption

In this subsection we extend 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, 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,

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

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

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

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

Step 2: Calculate the empirical critical value 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 $W_m = \prod_{i=1}^{N} \tilde{p}_i^{I(\tilde{p}_i \leq \tau)}$.

c. Repeat this process B times.

d. Obtain the empirical critical value, w_c, from the finite sample distributions generated by B simulations.

Step 3: If $W < w_c$, then reject the null hypothesis.

Step 4: Repeat the whole process M times to get size and power of the TPM.

Remark 4. *Note that the transformation alters the correlation among p-values, although marginally. As pointed out by Zaykin et al. (2002), however, the correlation is approximately invariant under monotone transformations. As a result, 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. The simulation algorithm for deriving the critical values assumes correlation to be the only form of dependence between the probits, as in the case when the test statistics jointly follow a multivariate normal distribution.*

Remark 5. *Zaykin et al. (2002) require the correlation matrix Σ to be non-degenerate, which is not the case for $T < N$. We circumvent this problem by resorting to Hartung's (1999) proposal. The proposed method here has the advantage that N can be very large. In that case, the probability in equation (9) should be computed through the Monte Carlo algorithm described above to avoid numerical overflow.*

The modified TPM, W_m, has the advantage that it does not require the panel to be balanced. Its disadvantage is also obvious, since it relies on a constant correlation assumption, which may not be true in some empirical applications. To allow for general forms of cross-section dependence, we propose bootstrap TPMs in the next subsection.

3.2 The TPM by Allowing for General Forms of Cross-section Dependence

We make the following assumptions:

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

$$\epsilon_{it} = \psi_i(L)e_{it}, \tag{10}$$

where L is the usual lag operator and $\psi_i(z) = \sum_{k=0}^{\infty} \psi_{ik} z^k$ for $i = 1, \ldots, N$.

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

We use the sieve bootstrap proposed by Bühlmann (1997). The sieve bootstrap approximates ϵ_{it} with a finite-order autoregressive process, where the order increases with sample size. Our proposed test closely follows from the difference-based sieve bootstrap, advocated by Chang and Park (2003) and Chang (2004). Throughout the paper, we use the notation * to denote bootstrap samples or statistics. Below we outline the necessary steps for conducting bootstrap TPM. We also discuss various issues arising in practical implementation of the proposed method.

Difference-based bootstrap TPM, W_a^*:

Step 1: Fit the approximated autoregression to $\hat{\epsilon}_{it}$, where $\hat{\epsilon}_{it} = \Delta y_{it}$ in equation (2) under the unit root null hypothesis,

$$\hat{\epsilon}_{it} = \sum_{j=1}^{J_i} \phi_{ij} \hat{\epsilon}_{i,t-j} + e_{it} \tag{11}$$

by the usual OLS regression, where the selection of lag order J_i is specified in *Remark 7*. Denote by $\hat{\phi}_{ij}$ the OLS estimates and by $\hat{e}_{it}$ the residuals in regression (11). Then form the time series residual vectors $\hat{\boldsymbol{e}}_t \equiv (\hat{e}_{1t}, \ldots, \hat{e}_{Nt})'$ for $t = 1, \ldots, T$.

Step 2: Generate the $N \times 1$ vector $\boldsymbol{e}_t^* \equiv (e_{1t}^*, \ldots, e_{Nt}^*)'$ by resampling from the centered residual vectors $\hat{\boldsymbol{e}}_t$. That is, obtain $\boldsymbol{e}_t^*$ from the empirical distribution of $(\hat{\boldsymbol{e}}_t - T^{-1} \sum_{t=1}^{T} \hat{\boldsymbol{e}}_t)$, $t = 1, \ldots, T$. The bootstrap samples $\boldsymbol{e}_t^*$ constructed as such will preserve the cross-section dependence structure of the data.

Step 3: Generate ϵ_{it}^* recursively from e_{it}^* as

$$\epsilon_{it}^* = \sum_{j=1}^{J_i} \hat{\phi}_{ij} \epsilon_{i,t-j}^* + e_{it}^*, \tag{12}$$

where $\hat{\phi}_{ij}$ are the estimated coefficients from the fitted regression (11).

Step 4: Impose the null of unit root to obtain bootstrap samples y_{it}^* as

$$y_{it}^* = y_{i,t-1}^* + \epsilon_{it}^*. \tag{13}$$

Step 5: Based on the bootstrap sample y_{it}^*, calculate the bootstrap TPM, W_a^*, defined in equation (8).

Step 6: Repeat steps 2-5 B times, where B denotes the number of bootstrap replications.

Step 7: Obtain $c^*(\lambda)$ such that $P\{W_a^* \leq c^*(\lambda)\} = \lambda$ for any given significance level λ. The bootstrap test W_a^* rejects the unit root null hypothesis if $W_a \leq c^*(\lambda)$, where W_a is the TPM using the original sample.

Residual-based bootstrap TPM, W_b^*:

Based on the test by Chang and Park (2003), Palm et al. (2008) propose a residual-based unit root test. We extend Palm et al. (2008)'s method to panel unit root test by resampling entire cross sections of residuals to preserve the dependence structure among cross-section units. All steps remain unchanged for the residual-based bootstrap TPM, W_b^*, except *Step 1*, in which the residuals are obtained from an Augmented Dickey-Fuller (ADF) regression as in the following equation

$$\hat{e}_{it} = y_{it} - \hat{\alpha}_i y_{i,t-1} - \sum_{j=1}^{J_i} \hat{\phi}_{ij} \Delta y_{i,t-j}. \tag{14}$$

Remark 6. *We give no formal proof of the consistency of the bootstrap tests, W_a^* and W_b^*, which might be conjectured from Chang (2004) and Palm et al. (2008).*

Remark 7. *The selection of lag orders of the approximated autoregressions (11) and (14) can be based on any of the well-known selection criteria such as AIC and BIC. We use modified AIC procedure by Ng and Perron (2001).*

Remark 8. *The initial values of ϵ_{it}^* in (12) and y_{it}^* in (13) are simply set equal to zero. We run the recursion for 50 initial observations before using ϵ_{it}^* and y_{it}^* to mitigate the effect of initial conditions.*

Remark 9. *The TPM can be easily modified to incorporate weights, w_i, into the analysis as $W = \prod_{i=1}^{N} p_i^{w_i I(p_i \leq \tau)}$, thus allowing tests of more precision to play a larger role.*

4 Monte Carlo Study

In this section we explore small-sample performance of the proposed TPMs, W_m, W_a^* and W_b^*. We consider both "strong" and "weak" cross-section dependence, with the former driven by a common factor and the latter due to spatial dependence.

4.1 The Design of Monte Carlo

Initially we consider dynamic panels with the cross-section dependence driven by a common factor. The data generating process (DGP) in this case is given by

$$y_{it} = (1-\alpha_i)\mu_i + \alpha_i y_{i,t-1} + \epsilon_{it}, \tag{15}$$

where

$$\epsilon_{it} = \gamma_i f_t + \xi_{it}, \tag{16}$$

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 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 distribution as $\mu_i \sim$ i.i.d. $N(0,1)$, $f_t \sim$ i.i.d. $N(0,\sigma_f^2)$ and $\xi_{it} \sim$ i.i.d. $N(0,1)$.

The factor structure in equation (16) has been widely used in the literature with the dependence being driven by a single factor in the error terms, eg. Moon and Perron (2004) and Pesaran (2007). 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 10. *Note that the DGP in (15) and (16) covers the case of non-stationary common factor and idiosyncratic errors and is nested within Bai and Ng's (2004) framework. To illustrate, consider the following DGP:*

$$\begin{aligned} y_{it} &= \gamma_i F_t + e_{it}, \\ F_t &= \varphi F_{t-1} + f_t, \\ e_{it} &= \delta_i e_{i,t-1} + \xi_{it}, \end{aligned} \tag{17}$$

where γ_i, f_t and ξ_{it} are defined as before. Under the null hypothesis of common and idiosyncratic unit roots, that is, $H_0 : \varphi = 1$, and $\delta_i = 1$

for all i, the DGP setup in (17) is equivalent to our DGP in (15) and (16) 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, that is, $H_0 : \varphi = 1$, and $\delta_i \sim U[0.8, 1]$ for all i. See, for example, Banerjee et al. (2004) for the 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 (16) are generated either as an AR(1) process (*DGP 3*)

$$\xi_{it} = \rho_i \xi_{i,t-1} + e_{it}, \tag{18}$$

or as an MA(1) process (*DGP 4*)

$$\xi_{it} = e_{i,t} + \lambda_i e_{i,t-1}, \tag{19}$$

where $e_{it} \sim$ i.i.d. $N(0,1)$. 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 behavior of our tests under different types of serial correlation.

Finally we consider spatial dependence as an alternative means of capturing cross-section dependence in the panel. Following Baltagi et al. (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 (15) can be expressed as

$$\epsilon_t = \theta_1 W_N \epsilon_t + \upsilon_t = (I_N - \theta_1 W_N)^{-1} \upsilon_t, \tag{20}$$

where W_N is an $N \times N$ known spatial weights matrix, θ_1 is the spatial autoregressive parameter and the error component υ_t is assumed to be distributed independently across 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

$$\epsilon_t = \theta_2 W_N \upsilon_t + \upsilon_t = (I_N + \theta_2 W_N) \upsilon_t, \tag{21}$$

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 $\theta_2 = 0.8$. Following Kelejian and Prucha (1999), 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 having nonzero elements in positions $i+1$ and $i-1$. Each row of this matrix is normalized such that all its non-zero elements are equal to 1/2.

For all of DGPs considered here, we choose

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

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 analyze 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 7.9 using $M = 2000$ simulations. Within each simulation, additional $B = 1000$ bootstrap replications are performed. All of the above parameters, μ_i, α_i, γ_i, λ_i, ρ_i are generated independently of each other, and of the error ξ_{it} and e_{it}, and also of the factor f_t. Moreover, f_t is generated independently of ξ_{it} and e_{it}.

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). The null distributions of the ADF t-statistics generally converge to functionals of the Brownian motion and thus analytic expressions of the distribution functions are not available. Now it is a fairly standard practice to obtain p-values of unit root tests using response surface regressions. We use p-values of the ADF tests as provided by MacKinnon (1996). [5]

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

$$\bar{\hat{\rho}} = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij}, \tag{22}$$

[5] MacKinnon (1996) derives asymptotic p-values by taking account of heteroscedasticity and shows that, for the ADF t tests, the differences between asymptotic and finite-sample p-values are quite modest. One can also use the finite-sample p-values of the ADF tests that controls the lag order as provided by Cheung and Lai (1995).

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

$$CD = \sqrt{\frac{2T}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij}. \quad (23)$$

Under the null of no cross-section dependence, $CD \sim N(0, 1)$. If the CD statistic does not reject, we use Zaykin et al. (2002)'s original W test; but if the CD statistic rejects, we use the proposed tests, W_m, W_a^* and W_b^*.

4.2 Monte Carlo Results

This section reports the size and power of the modified TPM (denoted by W_m), difference-based bootstrap TPM (denoted by W_a^*) and residual-based bootstrap TPM (denoted by W_b^*). For comparison, we also include some other commonly used second-generation panel unit root tests. More specifically, we consider Demetrescu et al. (2006)'s modified inverse normal test (denoted by Z), Hanck (2008)'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^*). To gauge the effect of cross-section dependence on first-generation panel unit root tests, we also include Maddala and Wu (1999)'s Fisher test (denoted by P) and Zaykin et al. (2002)'s original TPM (denoted by W).

Table 1 gives the average cross-section correlation coefficient $\bar{\hat{\rho}}$ and the average CD statistic for $N = 20$, $T = 50$ and $N = 50$, $T = 100$, using 2000 replications. 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 DGPs considered here are rather general, representing a wide range of cross-section dependence in practice. For all of the cases except DGP 1, the CD statistics reject the null of no cross-section dependence, and therefore we use the proposed TPMs under DGPs 2-6.

[Table 1 about here.]

In the absence of clear guidance regarding the choice of τ, we try 10 different cut-off values, ranging from 0.05, 0.1, 0.2, $\cdots$, 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. In our paper we select $\tau = 0.1$. We also note that our simulation results are similar as τ varies between 0.05 and 0.2. It seems that a small truncation point is preferable in practice. [6]

The results in Table 2 are obtained for the case of cross-section independence for a benchmark comparison. Table 3 reports the case of cross-section dependence driven by a single common factor without residual serial correlation. Tables 4 and 5 report the results from positive AR serial correlation and negative MA serial correlation, respectively.[7] Tables 6 and 7 report the results with spatial dependence.

We focus on sizes of the tests first. With *no* cross-section dependence (Table 2), P, S, W and CIPS tests yield good empirical size, close to the 5% nominal level. The rest of the tests are slightly undersized.

In the presence of strong cross-section dependence (Tables 3, 4 and 5), P test shows severe size distortions. W_a^* and W_b^* tests behave similarly and exhibit good size properties. K_a^* and K_b^* tests generally perform well but are severely oversized when $N = 50$ and $T = 20$. These results are not unexpected as both tests are designed for the cases of small N and large T. CIPS and S tests show size distortions for the case of negative serial correlation but perform reasonably well for other cases. Z and W_m tests are slightly oversized but P_e^c test is conservative in most of time.

Under SAR specification (Table 6), 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. Ignoring such a weak correlation leads to over-rejection as well, as shown by the result from P test. K_a^* and K_b^* tests are severely undersized. W_a^* and W_b^* tests are undersized for small T, but the size distortion reduces as T becomes large. S test performs reasonably well. Under SMA specification (Table 7), while all bootstrap tests suffer from downward size distortions, P, CIPS and W_m tests are slightly oversized. The size of other tests are, however, reasonably close to the nominal size in most of the cases.

We turn to finite sample powers of the tests. In general, the size-unadjusted power increases with T. All the tests become more powerful as N increases if δ is fixed, which justifies the use of panel data in unit root

[6]To save space, the complete simulation results for the TPM with all candidate truncation points are not reported here, but are available upon request.

[7]The results from negative AR serial correlation are qualitatively similar to those in Table 5 and the results from positive MA serial correlation are qualitatively similar to those in Table 4 and thus they are not reported here.

tests. The tests become more powerful when the proportion of stationary series increases in the panel, consistent with the findings in Karlsson and Löthgren (2000).

Given the substantial size distortions of P test under cross-section dependence, we are only interested in the comparison of other tests in terms of size-unadjusted power. A general finding is that there is no uniformly most powerful test. The powers depend on the sample size, the proportion of stationary series and the DGPs. S test becomes most powerful sometimes when only very few series in the panel are stationary ($\delta = 0.1$), but is outperformed by other tests when the proportion of stationary series increases. While CIPS test tends to have the best power under the factor structure, P_e^c test performs clearly better than other tests under spatial dependence in most cases. K_a^* and K_b^* tests behave similarly and have good power under factor structure but not under SAR specification. Without exception, W_b^* test is slightly more powerful than W_a^* test, consistent with the findings in Palm et al. (2008). W_m and Z tests have the power comparable to W_b^* test in most cases and all of them perform quite well under factor structure with negative serial correlation and spatial dependence.

[Table 2 about here.]

[Table 3 about here.]

[Table 4 about here.]

[Table 5 about here.]

[Table 6 about here.]

[Table 7 about here.]

5 Empirical Application

In their analysis of the term structure of macroeconomic forecasts, Lahiri and Sheng (2008) 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 and real GDP, we test this hypothesis directly.

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

$$a_{ith} = a_{it,h+1} + b_{ith}, \tag{24}$$

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. In Bloom's (2009) terminology, b_{ith} is called "uncertainty shocks".

The data in this study are taken from Survey of Professional Forecasters (SPF), 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 and real GDP. Although the SPF began in 1968, for several reasons as stated in Engelberg et al. (2009), we restrict attention to data collected from the first quarter of 1992 to the second quarter of 2009. We study the density forecasts for the annual real GDP and 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 forecasts are 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 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 percent of the time. This leaves us with 24 individuals, whose identification numbers are listed in Table 8.[8] The precision a_{ith} is calculated as the reciprocal of the variance of the density forecast reported by individual i.[9]

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

[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

We first estimate individual DF regressions. As is well known, the DF 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. An alternative 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 compute the pairwise cross-section correlation coefficient of the residuals from the above individual DF regressions. In our sample the average of these correlation coefficients, $\bar{\hat{\rho}}$, is estimated to be 0.07 and 0.09 for inflation and real GDP, respectively. The CD statistics, 9.41 for inflation and 11.70 for real GDP, strongly reject the null of no cross-section correlation for both variables.

[Table 8 about here.]

Now turning to panel unit root tests that account for this positive cross-section correlation.[10] The joint null and alternative hypotheses are specified as in (3) and (4). For inflation forecasts, Hanck (2008)'s S test, Maddala and Wu (1999)'s P test and the modified TPM, W_m, reject the joint null hypothesis of non-stationarity in forecasters' precision updating process at the 5% significance level, but Demetrescu et al. (2006)'s Z test fails to reject the null. As for real GDP forecasts, S and W_m tests show strong evidence of rejection, but P and Z tests do not reject. To understand the mixed evidence against the null, recall that Z test uses all p-values and tends to lose power when there are some very large p-values. 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, thus providing more power for W_m. 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 seems to show that in predicting real GDP and inflation, some professional forecasters do not

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 high certainty than a precision of 101.

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

update their forecast precision in an optimal 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 evolution of forecasters' subjective uncertainty as new information arrives.

6 Conclusion

In this paper we extend Zaykin et al. (2002)'s original TPM to allow for cross-section dependence in panel units. Three tests are developed: (i) modified TPM relying on a constant correlation assumption, W_m; (ii) difference-based bootstrap TPM, W_a^*; and (iii) residual-based bootstrap TPM, W_b^*. By construction, W_m test allows for unbalanced panel, and W_a^* and W_b^* tests are robust to general forms of cross-section dependence in the panel. As illustrated by the empirical example on studying forecasters' precision updating process, the proposed tests gain power in cases where there are some very large p-values.

We conduct a systematic comparison of the proposed tests with other commonly used panel unit root tests. Monte Carlo evidence shows that W_a^* and W_b^* tests yield good empirical size especially under factor structure with positive serial correlation and under spatial autoregressive specification and W_m test is slightly oversized. All three tests deliver satisfactory power under factor structure with negative serial correlation and spatial dependence, but W_b^* test appears to be slightly more powerful than W_a^* test.

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. This issue is currently under investigation by the authors. Furthermore, the proposed approach can also be applied to panel cointegration test.

References

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

Baltagi, B.H., Bresson, G. and Pirotte, A. (2007). Panel Unit Root Tests and Spatial Dependence. *Journal of Applied Econometrics* 22, 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* 7, 322-340.

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.

Bühlmann, P. (1997). Sieve Bootstrap for Time Series. *Bernoulli* 3, 123-148.

Chang, Y. and Park, J.Y. (2003). A Sieve Bootstrap for the Test of a Unit Root. *Journal of Time Series Analysis* 24, 379-400.

Chang, Y. (2004). Bootstrap Unit Root Tests in Panels with Cross-sectional Dependency. *Journal of Econometrics* 120, 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* 13, 277-280.

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.

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* 29, 111-145.

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

Hanck, C. (2008). Intersection Test for Panel Unit Roots. Forthcoming in *Econometric Reviews*.

Hanck, C. (2010). 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* 66, 249-255.

Kelejian, H.H. and Prucha, I.R. (1999). A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. *International Economic Review* 40, 509-533.

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

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.

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

Moskvina, V., Craddock, N., Holmans, P., Nikolov, I., Pahwa, J.S., Green, E., Wellcome Trust Case Control Consortium, Owen, M.J. and O'Donovan M.C. (2009). Gene-wide Analyses of Genome-wide Association Data Sets: Evidence for Multiple Common Risk Alleles for Schizophrenia and Bipolar Disorder and for Overlap in Genetic Risk. *Molecular Psychiatry* 14, 252-260.

Ng, S. and Perron, P. (2001). Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power. *Econometrica* 69, 1519-1554.

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

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

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

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* 34, 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* 5(9): e13022.

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.

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: Cross-section correlation for dynamic panels

N	T	DGP	Average correlation coefficient	Average CD statistic
20	50	DGP 1	0.000	-0.01
		DGP 2	0.797	77.66
		DGP 3	0.785	76.50
		DGP 4	0.793	77.29
		DGP 5	0.221	21.54
		DGP 6	0.074	7.18
50	100	DGP 1	0.000	-0.01
		DGP 2	0.802	280.61
		DGP 3	0.793	277.42
		DGP 4	0.796	278.71
		DGP 5	0.088	30.81
		DGP 6	0.029	10.15

DGP 1: cross-section independence
DGP 2: factor structure, no serial correlation
DGP 3: factor structure, positive AR serial correlation
DGP 4: factor structure, negative MA serial correlation
DGP 5: spatial AR dependence
DGP 6: spatial MA dependence

Table 2: Size and power of tests under DGP 1: cross-section independence

	N	T	P	Z	S	CIPS	P_e^c	K_a^*	K_b^*	W
δ=0	20	20	0.055	0.041	0.056	0.043	0.035	0.028	0.026	0.055
		50	0.049	0.042	0.053	0.045	0.035	0.040	0.037	0.044
		100	0.051	0.034	0.048	0.050	0.029	0.033	0.039	0.048
	50	20	0.044	0.020	0.055	0.030	0.022	0.031	0.031	0.054
		50	0.043	0.015	0.052	0.041	0.020	0.021	0.033	0.054
		100	0.048	0.019	0.052	0.049	0.032	0.033	0.035	0.045
δ=0.1	20	20	0.049	0.035	0.051	0.042	0.043	0.035	0.038	0.064
		50	0.084	0.063	0.060	0.049	0.052	0.069	0.094	0.086
		100	0.141	0.109	0.101	0.077	0.109	0.148	0.160	0.146
	50	20	0.052	0.019	0.054	0.044	0.037	0.030	0.025	0.058
		50	0.095	0.047	0.065	0.063	0.056	0.081	0.079	0.086
		100	0.264	0.142	0.124	0.098	0.160	0.239	0.272	0.244
δ=0.5	20	20	0.087	0.068	0.059	0.066	0.096	0.064	0.075	0.072
		50	0.334	0.245	0.103	0.176	0.316	0.410	0.409	0.213
		100	0.900	0.751	0.316	0.562	0.864	0.965	0.965	0.797
	50	20	0.140	0.069	0.064	0.064	0.124	0.052	0.055	0.085
		50	0.637	0.376	0.092	0.224	0.596	0.504	0.533	0.407
		100	0.999	0.951	0.368	0.758	0.995	1.000	1.000	0.985
δ=0.9	20	20	0.140	0.101	0.061	0.083	0.179	0.108	0.104	0.101
		50	0.703	0.324	0.146	0.396	0.751	0.813	0.808	0.440
		100	1.000	0.849	0.497	0.980	1.000	1.000	1.000	0.987
	50	20	0.233	0.116	0.064	0.088	0.297	0.066	0.069	0.136
		50	0.970	0.338	0.121	0.566	0.981	0.939	0.928	0.713
		100	1.000	0.960	0.585	1.000	1.000	1.000	1.000	1.000

Note: Rejection rates of panel unit root tests at nominal level α=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 (2010)'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 and W is Zaykin et al. (2002)'s original TPM.

Table 3: Size and power of tests under DGP 2: factor structure and no serial correlation

	N	T	P	Z	S	CIPS	P_e^c	K_a^*	K_b^*	W_m	W_a^*	W_b^*
	20	20	0.239	0.078	0.034	0.036	0.033	0.047	0.027	0.081	0.050	0.065
		50	0.222	0.072	0.030	0.050	0.030	0.033	0.031	0.077	0.068	0.075
δ=0		100	0.236	0.071	0.034	0.045	0.040	0.029	0.032	0.074	0.070	0.080
	50	20	0.271	0.076	0.030	0.037	0.025	0.217	0.196	0.079	0.063	0.063
		50	0.292	0.071	0.022	0.042	0.019	0.101	0.080	0.073	0.066	0.073
		100	0.287	0.068	0.034	0.050	0.022	0.083	0.075	0.070	0.071	0.075
	20	20	0.228	0.084	0.043	0.032	0.024	0.033	0.026	0.093	0.061	0.067
		50	0.257	0.087	0.037	0.256	0.020	0.073	0.058	0.092	0.064	0.081
δ=0.1		100	0.280	0.094	0.070	0.621	0.024	0.135	0.142	0.100	0.073	0.084
	50	20	0.304	0.094	0.035	0.020	0.016	0.231	0.171	0.102	0.062	0.075
		50	0.311	0.080	0.037	0.312	0.020	0.147	0.133	0.085	0.072	0.083
		100	0.356	0.091	0.082	0.748	0.030	0.261	0.262	0.118	0.089	0.098
	20	20	0.254	0.110	0.040	0.008	0.058	0.035	0.030	0.120	0.057	0.065
		50	0.375	0.151	0.069	0.473	0.105	0.125	0.089	0.188	0.100	0.117
δ=0.5		100	0.634	0.345	0.191	0.980	0.135	0.330	0.346	0.408	0.240	0.328
	50	20	0.326	0.124	0.036	0.000	0.091	0.211	0.166	0.142	0.067	0.071
		50	0.456	0.152	0.063	0.583	0.158	0.264	0.224	0.184	0.105	0.145
		100	0.706	0.367	0.182	0.997	0.217	0.600	0.592	0.429	0.262	0.349
	20	20	0.296	0.117	0.051	0.018	0.088	0.060	0.039	0.120	0.071	0.092
		50	0.476	0.164	0.081	0.356	0.321	0.158	0.152	0.176	0.169	0.200
δ=0.9		100	0.832	0.460	0.275	0.958	0.637	0.463	0.461	0.467	0.445	0.532
	50	20	0.358	0.118	0.047	0.007	0.125	0.291	0.231	0.124	0.076	0.093
		50	0.556	0.177	0.083	0.466	0.456	0.467	0.449	0.189	0.164	0.210
		100	0.877	0.445	0.244	0.990	0.747	0.841	0.852	0.456	0.438	0.516

Note: Rejection rates of panel unit root tests at nominal level α=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 (2010)'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 relying on a constant correlation assumption, W_a^* is difference-based bootstrap TPM and W_b^* is residual-based bootstrap TPM.

Table 4: Size and power of tests under DGP 3: factor structure and positive AR serial correlation

	N	T	P	Z	S	CIPS	P_e^c	K_a^*	K_b^*	W_m	W_a^*	W_b^*
	20	20	0.184	0.095	0.030	0.030	0.050	0.044	0.025	0.089	0.053	0.051
		50	0.197	0.083	0.030	0.023	0.032	0.056	0.051	0.084	0.057	0.063
δ=0		100	0.173	0.073	0.028	0.033	0.032	0.082	0.068	0.072	0.060	0.060
	50	20	0.265	0.100	0.038	0.025	0.033	0.226	0.181	0.100	0.047	0.054
		50	0.242	0.081	0.026	0.040	0.019	0.173	0.161	0.087	0.053	0.055
		100	0.241	0.076	0.025	0.045	0.027	0.200	0.192	0.077	0.060	0.068
	20	20	0.198	0.095	0.031	0.026	0.042	0.044	0.023	0.093	0.045	0.055
		50	0.222	0.086	0.035	0.089	0.026	0.078	0.063	0.089	0.061	0.063
δ=0.1		100	0.239	0.094	0.062	0.301	0.038	0.216	0.196	0.108	0.068	0.075
	50	20	0.259	0.096	0.034	0.019	0.027	0.194	0.164	0.097	0.050	0.054
		50	0.289	0.102	0.036	0.103	0.021	0.237	0.213	0.106	0.074	0.076
		100	0.325	0.102	0.055	0.401	0.040	0.328	0.414	0.112	0.068	0.076
	20	20	0.229	0.116	0.044	0.007	0.072	0.062	0.035	0.136	0.047	0.059
		50	0.320	0.130	0.050	0.272	0.113	0.196	0.149	0.158	0.068	0.111
δ=0.5		100	0.558	0.305	0.134	0.882	0.163	0.491	0.454	0.363	0.227	0.260
	50	20	0.274	0.106	0.032	0.002	0.101	0.200	0.146	0.125	0.049	0.053
		50	0.404	0.146	0.048	0.257	0.160	0.385	0.364	0.176	0.086	0.103
		100	0.654	0.334	0.120	0.963	0.249	0.786	0.783	0.411	0.261	0.303
	20	20	0.246	0.103	0.037	0.017	0.120	0.073	0.060	0.108	0.055	0.064
		50	0.428	0.158	0.060	0.152	0.386	0.281	0.215	0.165	0.140	0.166
δ=0.9		100	0.761	0.377	0.187	0.818	0.738	0.595	0.590	0.385	0.369	0.434
	50	20	0.327	0.129	0.034	0.013	0.158	0.244	0.201	0.137	0.068	0.089
		50	0.505	0.154	0.052	0.272	0.563	0.588	0.538	0.167	0.142	0.175
		100	0.829	0.393	0.169	0.967	0.855	0.935	0.934	0.410	0.425	0.455

Note: See the note of Table 3.

Table 5: Size and power of tests under DGP 4: factor structure and negative MA serial correlation

	N	T	P	Z	S	CIPS	P_e^c	K_a^*	K_b^*	W_m	W_a^*	W_b^*
	20	20	0.280	0.087	0.063	0.134	0.039	0.030	0.018	0.094	0.070	0.077
		50	0.278	0.090	0.079	0.200	0.034	0.015	0.012	0.093	0.085	0.091
δ=0		100	0.276	0.082	0.096	0.213	0.035	0.015	0.014	0.081	0.080	0.088
	50	20	0.315	0.078	0.057	0.151	0.023	0.205	0.211	0.084	0.071	0.083
		50	0.344	0.085	0.097	0.256	0.021	0.053	0.042	0.089	0.082	0.087
		100	0.341	0.074	0.111	0.312	0.032	0.034	0.044	0.078	0.081	0.095
	20	20	0.281	0.092	0.055	0.080	0.027	0.031	0.020	0.098	0.066	0.079
		50	0.315	0.100	0.107	0.488	0.025	0.039	0.026	0.108	0.081	0.105
δ=0.1		100	0.337	0.110	0.156	0.819	0.026	0.082	0.090	0.135	0.082	0.106
	50	20	0.325	0.082	0.062	0.094	0.021	0.201	0.189	0.086	0.064	0.075
		50	0.352	0.080	0.118	0.589	0.027	0.092	0.079	0.094	0.078	0.082
		100	0.386	0.103	0.198	0.938	0.034	0.168	0.152	0.133	0.094	0.112
	20	20	0.305	0.110	0.070	0.010	0.065	0.047	0.025	0.129	0.048	0.089
		50	0.428	0.180	0.154	0.607	0.121	0.084	0.065	0.211	0.119	0.161
δ=0.5		100	0.682	0.405	0.324	0.995	0.144	0.262	0.266	0.468	0.301	0.412
	50	20	0.374	0.136	0.071	0.004	0.105	0.178	0.157	0.155	0.060	0.092
		50	0.502	0.184	0.173	0.757	0.170	0.176	0.155	0.225	0.119	0.175
		100	0.744	0.408	0.399	1.000	0.191	0.438	0.443	0.481	0.291	0.404
	20	20	0.338	0.124	0.075	0.035	0.065	0.046	0.041	0.132	0.085	0.103
		50	0.563	0.218	0.176	0.509	0.253	0.102	0.095	0.221	0.189	0.256
δ=0.9		100	0.837	0.525	0.438	0.978	0.525	0.344	0.362	0.519	0.457	0.554
	50	20	0.391	0.119	0.081	0.031	0.102	0.269	0.233	0.132	0.092	0.113
		50	0.605	0.218	0.236	0.715	0.358	0.362	0.316	0.229	0.191	0.236
		100	0.885	0.543	0.532	0.996	0.584	0.676	0.682	0.541	0.459	0.568

Note: See the note of Table 3.

Table 6: Size and power of tests under DGP 5: spatial AR dependence

	N	T	P	Z	S	CIPS	P_e^c	K_a^*	K_b^*	W_m	W_a^*	W_b^*
	20	20	0.126	0.076	0.049	0.115	0.092	0.009	0.006	0.092	0.019	0.021
		50	0.135	0.070	0.046	0.118	0.073	0.008	0.007	0.079	0.036	0.040
δ=0		100	0.118	0.071	0.044	0.138	0.082	0.004	0.003	0.067	0.039	0.054
	50	20	0.117	0.051	0.052	0.096	0.071	0.013	0.013	0.092	0.012	0.014
		50	0.129	0.056	0.058	0.109	0.062	0.002	0.002	0.080	0.033	0.037
		100	0.125	0.056	0.047	0.097	0.088	0.001	0.001	0.080	0.038	0.048
	20	20	0.124	0.078	0.056	0.120	0.094	0.011	0.011	0.114	0.026	0.027
		50	0.147	0.092	0.060	0.131	0.091	0.011	0.011	0.130	0.041	0.052
δ=0.1		100	0.228	0.134	0.104	0.188	0.127	0.014	0.014	0.196	0.067	0.088
	50	20	0.145	0.067	0.053	0.081	0.076	0.011	0.012	0.137	0.019	0.021
		50	0.181	0.083	0.051	0.124	0.102	0.006	0.006	0.157	0.044	0.064
		100	0.318	0.154	0.106	0.172	0.173	0.007	0.006	0.308	0.148	0.166
	20	20	0.174	0.098	0.061	0.132	0.148	0.014	0.011	0.138	0.032	0.042
		50	0.361	0.195	0.086	0.240	0.265	0.035	0.034	0.227	0.122	0.156
δ=0.5		100	0.781	0.548	0.266	0.520	0.506	0.141	0.116	0.611	0.535	0.579
	50	20	0.208	0.095	0.071	0.105	0.155	0.017	0.018	0.165	0.029	0.036
		50	0.564	0.275	0.103	0.266	0.456	0.032	0.030	0.317	0.197	0.244
		100	0.970	0.822	0.314	0.667	0.849	0.186	0.210	0.847	0.842	0.891
	20	20	0.225	0.129	0.064	0.153	0.201	0.025	0.023	0.161	0.031	0.051
		50	0.621	0.270	0.133	0.454	0.583	0.088	0.071	0.273	0.235	0.284
δ=0.9		100	0.984	0.721	0.434	0.922	0.957	0.474	0.415	0.721	0.850	0.887
	50	20	0.296	0.140	0.070	0.142	0.290	0.021	0.020	0.182	0.033	0.058
		50	0.880	0.321	0.122	0.544	0.878	0.109	0.116	0.319	0.405	0.487
		100	1.000	0.879	0.478	0.995	1.000	0.748	0.752	0.872	0.991	0.996

Note: See the note of Table 3.

Table 7: Size and power of tests under DGP 6: spatial MA dependence

	N	T	P	Z	S	CIPS	P_e^c	K_a^*	K_b^*	W_m	W_a^*	W_b^*
	20	20	0.070	0.049	0.044	0.072	0.060	0.024	0.023	0.070	0.014	0.017
		50	0.075	0.050	0.052	0.089	0.046	0.013	0.015	0.074	0.016	0.023
δ=0		100	0.077	0.051	0.042	0.088	0.053	0.019	0.020	0.074	0.034	0.041
	50	20	0.081	0.039	0.058	0.047	0.044	0.024	0.024	0.089	0.009	0.009
		50	0.078	0.036	0.043	0.062	0.035	0.012	0.010	0.071	0.013	0.017
		100	0.076	0.031	0.048	0.060	0.046	0.012	0.013	0.074	0.022	0.038
	20	20	0.080	0.054	0.050	0.066	0.055	0.025	0.021	0.102	0.019	0.024
		50	0.101	0.062	0.053	0.077	0.077	0.029	0.034	0.111	0.036	0.034
δ=0.1		100	0.185	0.121	0.097	0.132	0.125	0.056	0.058	0.212	0.077	0.107
	50	20	0.076	0.034	0.062	0.068	0.045	0.032	0.025	0.109	0.009	0.014
		50	0.133	0.058	0.050	0.085	0.088	0.029	0.025	0.146	0.024	0.033
		100	0.281	0.144	0.118	0.127	0.197	0.069	0.071	0.298	0.133	0.179
	20	20	0.125	0.078	0.055	0.082	0.101	0.031	0.038	0.121	0.027	0.032
		50	0.346	0.221	0.083	0.209	0.283	0.127	0.153	0.243	0.113	0.155
δ=0.5		100	0.862	0.668	0.280	0.525	0.667	0.568	0.573	0.700	0.633	0.689
	50	20	0.155	0.076	0.050	0.077	0.139	0.024	0.042	0.150	0.008	0.016
		50	0.600	0.330	0.098	0.239	0.576	0.213	0.215	0.354	0.169	0.232
		100	0.995	0.910	0.350	0.720	0.987	0.890	0.898	0.918	0.942	0.955
	20	20	0.187	0.118	0.067	0.109	0.177	0.055	0.054	0.151	0.027	0.048
		50	0.652	0.278	0.114	0.399	0.653	0.378	0.367	0.257	0.254	0.326
δ=0.9		100	0.999	0.801	0.498	0.953	0.990	0.952	0.968	0.802	0.928	0.946
	50	20	0.290	0.150	0.079	0.106	0.316	0.040	0.038	0.198	0.024	0.052
		50	0.942	0.344	0.130	0.535	0.956	0.572	0.548	0.320	0.415	0.544
		100	1.000	0.930	0.543	1.000	1.000	1.000	1.000	0.930	1.000	1.000

Note: See the note of Table 3.

Table 8: 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 relying on a constant correlation assumption. 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.