Shih-Hsun Hsu[1]

Disentangling the source of non-stationarity in a panel of seasonal data

[1] Department of Economics, National Chengchi University, Taipei 11605, Taiwan, Tel.: +886-2-2939-3091 ext: 51667, Fax: +886-2-2939-0344, E-mail: shhsu@nccu.edu.tw. https://orcid.org/0000-0002-8564-4846.

Abstract:
In dealing with a panel of seasonal data with cross-section dependence, this paper establishes a common factor model to investigate whether the seasonal and non-seasonal non-stationarity in a series is pervasive, or specific, or both. Without knowing a *priori* whether the data are seasonal stationary or not, we propose a procedure for consistently estimating the model; thus, the seasonal non-stationarity of common factors and idiosyncratic errors can be separately detected accordingly. We evaluate the methodology in a series of Monte Carlo simulations and apply it to test for non-stationarity and to disentangle their sources in panels of worldwide real exchange rates and of consumer price indexes for 37 advanced economies.

Keywords: common factor, consumer price index, pooled test, purchasing power parity, seasonal non-stationarity, seasonal panels, seasonal unit roots

DOI: 10.1515/snde-2018-0075

1 Introduction

Whether or not a series is stationary is important for statistical inference, and it is currently quite standard for an empirical study to employ the unit root tests such as the (augmented) Dickey-Fuller test (Dickey and Fuller 1979; 1981) or the Phillips-Perron test (Phillips and Perron, 1988) to detect possible non-stationarity for the series of interest. Besides, in order to improve the testing power, many panel unit root tests have also been proposed by pooling information over cross-section units; they are constructed under the assumption of independence across cross-section units (e.g. Levin and Lin, 1993; Levin, Lin and Chu, 2002; , Im, Pesaran and Shin, 2003, henceforth IPS) or established by allowing for cross-section dependence (e.g. Bai and Ng, 2004; Chang, 2004; Moon and Perron, 2004; Breitung and Das, 2005; Pesaran, 2007); for more reviews on (non-seasonal) panel unit root tests, see Hlouskova and Wagner (2006) or Breitung and Pesaran (2008), among others.

Besides the conventional unit root +1 at long-run (or zero) frequency, unit roots at different seasonal frequencies may also induce non-stationarity;[1] these seasonal unit roots are especially important when the seasonal data are of interest. To detect various kinds of non-stationarity induced by seasonal unit roots for an univariate seasonal series, the test of Dickey, Hasza, and Fuller (1984), the HEGY approach of Hylleberg et al. (1990), Kunst's (2009) nonparametric test, the test of Osborn et al. (1988), etc., are thus proposed. While taking advantages of the progress of non-seasonal panel unit root tests, there have been few non-stationary unit root tests for seasonal panels in the last decade. With a cross-section independence assumption, Dreger and Reimers (2005), Otero, Smith, and Giulietti (2005), and Ucar and Guler (2010) extended the IPS test to pooling individual HEGY test statistics associated with cross-section units (HEGY-IPS test hereafter); while taking possible cross-section dependence into account, Otero, Smith and Giulietti (2007, 2008) and Kunst and Franses (2011) further extended the HEGY-IPS test by integrating Pesaran's (2007) method, whereas Ho (2008), Lee and Shin (2006), and Shin and Oh (2009) proposed tests based on instrument variable (IV) methods.

This paper also focuses on stochastic non-stationarity in a panel of seasonal data with cross-section dependence. However, unlike the existing works, we go beyond them to propose a framework to disentangle the source of seasonal non-stationarity induced by seasonal unit roots. We establish a factor model for the seasonal panel by extending PANIC (Panel Analysis of Nonstationarity in Idiosyncratic and Common components) approach of Bai and Ng (2004) to the case where possible seasonal non-stationarity is permitted. None of the stationarity assumptions on components in the model is necessary, and the cross-section dependence among units is delicately captured by common factors in this model. Based on the proposed approach, we can discover whether the seasonal non-stationarity induced by seasonal unit roots in a series is pervasive, unit-specific, or both, and detect the number of common trends in the panel, at the seasonal frequency of interest. To increase the testing power, we also introduce a pooled test by using the summation of minus two times the logarithm

Shih-Hsun Hsu is the corresponding author.

of the p-values (resulting from the individual HEGY tests) across individuals. Moreover, when there are multiple common factors, we also propose a procedure which involves successive tests to determine the number of (seasonal) non-stationary factors. The simulations show that the proposed estimation and testing procedures perform very well in most configurations considered. To the best of our knowledge, for a panel data with possible seasonal unit roots, no such analysis exists thus far in the literature.

For empirical applications, we applied the proposed framework to investigate the possible non-stationarity in the panel of worldwide real exchange rates and in the panel of consumer price indexes (CPIs) for 37 advanced economies. Since the existence of non-stationary real exchange rates may be viewed as a violation of purchasing power parity (PPP hereafter),[2] and seasonal patterns (and non-stationarity) are expected to be observed frequently in CPI because of its construction, the sources of of non-stationarity in each panel thus are of our particular interest. Two balanced panels of quarterly data were constructed from the International Financial Statistics (IFS hereafter) dataset of the International Monetary Fund (IMF hereafter). Based on the proposed factor model associated with the estimation and testing procedures, we tested for the existence of seasonal and non-seasonal non-stationarity, and then disentangled the source of non-stationarity in these two panels.

The rest of this paper is organized as follows. We present the definition of seasonal unit roots and the considered factor model in Section 2, the proposed estimation and test procedures in Section 3, the main asymptotic properties of the proposed approach in Section 4, and results for some simulation designs in Section 5. An empirical application is illustrated in Section 6 and we conclude this paper in Section 7.

2 Seasonal unit roots and factor model

Before proceeding to specify the proposed approach, for a seasonal process w_t, it is worth briefly introducing non-stationary seasonality. First, w_t is seasonally integrated of order d, $SI(d)$ say, if $(1-L^S)^d w_t \equiv \Delta_S^d w_t$ is a stationary process, where L is the lag operator. When $d=1$, $\Delta_S \equiv \Delta_S^1$ is known as the first order annual differencing operator or the seasonal differencing filter. For the quarterly data ($S=4$), w_t is $SI(1)$ implies that $(1-L^4)w_t = \Delta_4 w_t$ is stationary. Since $(1-L^4) = (1-L)(1+L)(1+L^2)$, the spectral density of w_t has infinite spectral power at the zero frequency (corresponding to the conventional unit root of +1) and at the seasonal frequencies of π and $\pi/2$ (corresponding to the semi-annual unit root of -1 and annual unit roots of $\pm i$, respectively).[3] Besides, we claim that some seasonal variables are seasonally cointegrated at some seasonal frequency, if each of them has a unit root but their combination is stationary at this seasonal frequency; see Hylleberg et al. (1990), Wells (1997), and Ghysels and Osborn (2001), among others.

2.1 Model specification

Let $\{y_{it}\}$ denote the quarterly panel data which satisfies the common factor structure as:

$$y_{it} = \mathscr{D}_{it} + \boldsymbol{\lambda}_i' \boldsymbol{F}_t + \epsilon_{it}, i = 1, \dots, N, t = -3, \dots, 0, \dots, T, \tag{1}$$

where $\mathscr{D}_{it}$ is the deterministic component of unit i at time t, $\boldsymbol{F}_t = [F_{1t}, F_{2t}, \cdots, F_{q^*t}]'$ is a $q^* \times 1$ vector of common factors where $q^* \ll N$, $\boldsymbol{\lambda}_i$ is a corresponding $q^* \times 1$ vector of factor loadings, and ϵ_{it} is an idiosyncratic error for unit i. In order to clearly discern between the stationary factor(s) and non-stationary one(s), if necessary, we may further denote $\boldsymbol{F}_t^* = \boldsymbol{\Gamma}' \boldsymbol{F}_t$ as a transformation of $\boldsymbol{F}_t$ with a $q^* \times q^*$ rotation matrix $\boldsymbol{\Gamma} = [\boldsymbol{\gamma}_1\ \boldsymbol{\gamma}_2\ \dots\ \boldsymbol{\gamma}_{q^*}]$ such that the first q_1 factors are non-stationary while the remaining $(q^* - q_1)$ factors are stationary.

Given this specification, the cross-section dependence among units is allowed and directly captured by the common factors $\boldsymbol{F}_t$. Besides, if a series y_{it} is seasonal stationary or not depends on the combinations of common factors $\boldsymbol{F}_t$ and idiosyncratic error ϵ_{it}. If one or more of the common factors are seasonal non-stationary, it indicates that for all units i with non-zero loadings, y_{it} is seasonal non-stationary, no matter how ϵ_{it} is, since the idiosyncratic errors are independent of common factors by definition. On the other hand, if all common factors are seasonal stationary, then y_{it} is (resp. not) seasonal stationary only when its idiosyncratic error ϵ_{it} is (resp. not). Moreover, at some seasonal frequency, if one of the common factors (and thus y_{it}) has the corresponding seasonal unit root but all ϵ_{it} do not, the linear relationships among y_{it} and $\boldsymbol{F}_t$ in the model specification (1) can be treated as a kind of seasonal cointegration at this seasonal frequency. Last but not least, after employing the typical seasonal unit root tests on these components in model (1), we can tell whether the non-stationarity at some seasonal frequency in the seasonal series is pervasive (due to common factors), or unit-specific (due to idiosyncratic error), or both.

3 Proposed estimation and testing procedures

According to model specification (1) and the aims of this paper, it is easy to see that the common factor structure plays a crucial role. However, the factors and the loadings are unobserved. Besides, because y_{it} in model (1) may be (seasonal) non-stationary, it is thus impossible to estimate $\boldsymbol{F}_t$ and λ_i as well as q^* (the number of common factors) directly based on y_{it}. As a consequence, this paper proposes an estimation procedure by extending the approach of Bai and Ng (2004) to take (possible) seasonal unit roots into account. Based on the estimates, we then introduce the augmented HEGY test of Hylleberg et al. (1990), its pooled version for the panel data, and propose a new procedure which involves successive tests for determining the number of common stochastic trends.

3.1 Estimation procedure

To estimate the model (1), we first compute the principal components (PC) for the annual-differenced series $\Delta_4 y_{it}$. By using these principal components, we then estimate the number of common factors, and the associated estimates of loadings and $\boldsymbol{F}_t$.

To be precise, it is instructive to start by considering the case where $\mathscr{D}_{it}$ contains an intercept or seasonal dummies. In this case we have $\Delta_4\mathscr{D}_{it} = 0$ and from (1), for $t = 1, \ldots, T$, $i = 1, \ldots, N$,

$$\begin{aligned} \Delta_4 y_{it} &= \boldsymbol{\lambda}_i' \Delta_4 \boldsymbol{F}_t + \Delta_4 \epsilon_{it}, \\ &:= \boldsymbol{\lambda}_i' \boldsymbol{f}_t^S + \varepsilon_{it}^S, \end{aligned} \tag{2}$$

where $\boldsymbol{f}_t^S = \Delta_4 \boldsymbol{F}_t$ and $\varepsilon_{it}^S = \Delta_4 \epsilon_{it}$ are annual-differenced variables of $\boldsymbol{F}_t$ and ϵ_{it}, respectively. Note that, because $\Delta_4 y_{it}$ in (2) is stationary for all i no matter what the nature of the original series y_{it} is, the space spanned either by $\boldsymbol{f}_t^S$ or by λ_i can be consistently estimated now by applying the conventional method of principal components to $\Delta_4 y$ when N and T are large enough.[4] On the other hand, if $\mathscr{D}_{it}$ contains a linear trend, then $\Delta_4\mathscr{D}_{it}$ is a non-zero constant and would be present on the right-hand side of model (2). As suggested in Bai and Ng (2004), if we replace $\Delta_4 y_{it}$ with its demeaned version in model (2), the loadings λ_i and the factors $\boldsymbol{f}_t^S$ can still be consistently estimated under the assumption that $\mathbb{E}[\boldsymbol{f}_t^S] = 0$.

Let $\Delta_4 \boldsymbol{Y}$ denote the $N \times T$ matrix in which the (i, t)-th element is $\Delta_4 y_{it}$, $i = 1, \ldots, N$, $t = 1, \ldots T$, and $\boldsymbol{\Lambda}(q) = [\boldsymbol{\lambda}_1(q)\ \boldsymbol{\lambda}_2(q)\ \ldots\ \boldsymbol{\lambda}_i(q)\ \ldots\ \boldsymbol{\lambda}_N(q)]'$ the corresponding $N \times q$ matrix of loadings, and $\boldsymbol{F}^S(q) = [\boldsymbol{f}_1^S(q)\, \boldsymbol{f}_2^S(q)\ \ldots\ \boldsymbol{f}_t^S(q)\ \ldots\ \boldsymbol{f}_T^S(q)]'$ the $T \times q$ matrix of factors, when there are q common factors. Under the normalization that $\boldsymbol{\Lambda}(q)'\boldsymbol{\Lambda}(q)/N = I_q$, a q-dimension identity matrix, and $\boldsymbol{F}^S(q)'\boldsymbol{F}^S(q)$ being diagonal, the resulting PC estimate of $\boldsymbol{\Lambda}(q)$, $\hat{\boldsymbol{\Lambda}}(q)$ say, is $\sqrt{N}$ times the eigenvectors corresponding to the q largest eigenvalues of the $N \times N$ matrix $\Delta_4 \boldsymbol{Y} \Delta_4 \boldsymbol{Y}'$, and $\hat{\boldsymbol{F}}^S(q) = \Delta_4 \boldsymbol{Y}' \hat{\boldsymbol{\Lambda}}(q)/N$ is the associated PC estimate of $\boldsymbol{F}^S(q)$.

Since the true number of common factors q^* is unknown in most cases, we can further estimate it by using the information criteria summarized in Bai and Ng (2008). The estimate for the number of factors is defined as:

$$\hat{q} = \arg \min_{0 \le q \le \bar{q}} \ln(S(q)) + q \cdot \Gamma(N, T), \tag{3}$$

where $\bar{q}$ is a pre-specified value;

$$S(q) = (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\Delta_4 y_{it} - \hat{\boldsymbol{\lambda}}_i'(q) \hat{\boldsymbol{f}}_t^S(q) \right]^2$$

is the sum of squared residuals when $\hat{\boldsymbol{f}}_t^S(q)$ and $\hat{\boldsymbol{\lambda}}_i(q)$, respectively, are the estimates of q factors and loadings based on the method of PC, and $\Gamma(N, T)$ is a penalty weight which satisfies the regularity condition that $\Gamma(N, T) \to 0$ and $C_{NT}^2 \cdot \Gamma(N, T) \to \infty$ as $N, T \to \infty$ with $C_{NT} = \min(\sqrt{N}, \sqrt{T})$. Three valid functions suggested in Bai and Ng (2008) will be considered for the simulations and two empirical applications below; they are:

$$\Gamma_1(N,T) = \frac{N+T}{NT}\ln\left(\frac{NT}{N+T}\right),$$
$$\Gamma_2(N,T) = \frac{N+T}{NT}\ln(C_{NT}^2),$$
$$\Gamma_3(N,T) = \frac{\ln(C_{NT}^2)}{C_{NT}^2}.$$

Besides, we also follow Bai and Ng**'s** (2008) suggestion to consider an additional penalty weight as:

$$\Gamma_4(N,T) = \frac{N+T-q}{NT}\ln(NT).$$

$\Gamma_4(N,T)$ has good properties when the errors are cross correlated even though it may fail the condition $\Gamma(N,T) \to 0$ under the some particular configurations of N and T. More details about the properties of the estimation of factor models by using PC can be found in Stock and Watson (2002b) and Bai and Ng (2008), among others.

Let $\hat{\boldsymbol{f}}_t^S := \hat{\boldsymbol{f}}_t^S(\hat{q})$ and $\hat{\boldsymbol{\lambda}}_i := \hat{\boldsymbol{\lambda}}_i(\hat{q})$ denote the corresponding estimates when $\hat{q}$ is determined by (3); the consistent estimate of ε_{it}^S is then obtained by $\hat{\varepsilon}_{it}^S = \Delta_4 y_{it} - \hat{\boldsymbol{\lambda}}_i'\hat{\boldsymbol{f}}_t^S$. As a consequence, the estimates of $\boldsymbol{F}_t$ and ϵ_{it}, $t = 1, \ldots, T$, $i = 1, \ldots, N$, are, respectively,

$$\hat{\boldsymbol{F}}_t = \sum_{k=0}^{\lfloor t/4 \rfloor} \hat{\boldsymbol{f}}_{t-4k}^S \text{ and } \hat{\epsilon}_{it} = \sum_{k=0}^{\lfloor t/4 \rfloor} \varepsilon_{it-4k}^S, \tag{4}$$

where $\lfloor m \rfloor$ denotes the largest integer not greater than m. $\hat{\boldsymbol{F}}_t$ and $\hat{\epsilon}_{\mathrm{it}}$ are simply the sample counterparts of $\boldsymbol{F}_t$ and ϵ_{it}, respectively, since:

$$\boldsymbol{F}_t = (1-L^4)^{-1}\boldsymbol{f}_t^S = (1+L^4+L^8+\cdots)\boldsymbol{f}_t^S,$$
$$\epsilon_{it} = (1-L^4)^{-1}\varepsilon_{it}^S = (1+L^4+L^8+\cdots)\varepsilon_{it}^S.$$

Note that these estimates in (4) differ from what was proposed in Bai and Ng (2004) which deals with the panel data with possible non-seasonal non-stationarity.

When there is only one common factor, we can directly infer its (seasonal) non-stationary dynamics from its corresponding estimate. However, when there are multiple factors, individually testing each of them for the possible seasonal unit roots at different frequencies would overstate the (seasonal) non-stationary dynamics in general since we can only estimate the space spanned by those factors instead of the factors themselves; any linear combinations of the factor(s) with seasonal unit root(s) and other stationary one(s) would yield another non-stationary factor. As a consequence, we further construct the transformed estimates $\hat{\boldsymbol{F}}_t^*$ by estimating the $\hat{q} \times \hat{q}$ rotation matrix $\hat{\boldsymbol{\Gamma}}$ such that:

$$\hat{\boldsymbol{F}}_t^* := [\hat{F}_{1t}^*, \hat{F}_{2t}^*, \ldots, \hat{F}_{\hat{q}t}^*]' = \hat{\boldsymbol{\Gamma}}'\hat{\boldsymbol{F}}_t, \tag{5}$$

where $\hat{\boldsymbol{\Gamma}} = [\hat{\boldsymbol{\gamma}}_1\ \hat{\boldsymbol{\gamma}}_2\ \ldots\ \hat{\boldsymbol{\gamma}}_{\hat{q}}]$ is the matrix of the eigenvectors associated with the descending eigenvalues of $T^{-2}\sum_{t=1}^{T}\hat{\boldsymbol{F}}_t\hat{\boldsymbol{F}}_t'$. Since the stationary factor has bounded variance while the variance of the non-stationary factor is explosive, if we suspect that the data of interest is driven by some stationary and some (seasonal) non-stationary factors, the transformed factor $\hat{F}_{\hat{q}t}^* = \hat{\boldsymbol{\gamma}}_{\hat{q}}'\hat{\boldsymbol{F}}_t$ is the most likely estimate for the stationary one, whereas $\hat{F}_{1t}^* = \hat{\boldsymbol{\gamma}}_1'\hat{\boldsymbol{F}}_t$ could be non-stationary with the highest probability. Similar arguments on distinguishing the stationary component(s) from the non-stationary one(s) can also be found in the literature on identifying cointegration relationships among variables (Stock and Watson, 1988; Harris, 1997) or detecting the rank of non-stationary space of common factors (Bai and Ng, 2004).

3.2 Testing procedure

Given the estimates of q^*, $\boldsymbol{F}_t$ and ϵ_{it}, we are now in a position to test their seasonal unit roots by introducing the so-called "augmented HEGY test" of Hylleberg et al. (1990) for a univariate quarterly series and their extensions to seasonal panel data.

3.2.1 Augmented HEGY test for univariate series

To test seasonal unit root(s) at different seasonal frequencies and seasonal integration of idiosyncratic error for a cross-section unit i, or of the common factors, the augmented HEGY test is employed as follows. Let w_{it} be an univariate quarterly series at time t;[5] the augmented HEGY test then considers a regression:

$$\Delta_4 w_{it} = \pi_1 w^{(1)}_{it-1} - \pi_2 w^{(2)}_{it-1} - \pi_3 w^{(3)}_{it-2} - \pi_4 w^{(3)}_{it-1} + \sum_{l=1}^{l_i} \rho_l \Delta_4 w_{it-l} + \eta_{it}, \tag{6}$$

where $w^{(1)}_{it-1} = (1+L)(1+L^2)w_{it}$, $w^{(2)}_{it-1} = (1-L)(1+L^2)w_{it}$, $w^{(3)}_{it-1} = (1-L)(1+L)w_{it}$, and l_i, the lag length of $\Delta_4 w_{it}$, is chosen such that η_{it} is uncorrelated white noise across t. Given the model (6) and OLS estimates for the corresponding parameters, we may detect:

(a) w_{it} has a conventional unit root +1 by testing the null $H_o^{(0)} : \pi_1 = 0$;

(b) w_{it} has a semi-annual unit root -1 by testing the null $H_o^{(\pi)} : \pi_2 = 0$;

(c) w_{it} has annual unit roots $\pm i$ by testing the null $H_o^{(\pi/2)} : \pi_3 = \pi_4 = 0$;

(d) w_{it} has all seasonal unit roots by testing the null $H_o^{(S)} : \pi_2 = \pi_3 = \pi_4 = 0$;

(e) w_{it} is $SI(1)$ by testing the null $H_o^{(SI)} : \pi_1 = \pi_2 = \pi_3 = \pi_4 = 0$.

According to the estimation results for the regression model (6), the first two (one-sided to the left) tests are easily implemented by using the resulting t-statistics, and the last three (one-sided to the right) are done by employing the corresponding F-statistics. Besides, the joint tests for multiple seasonal unit roots, by considering any two of $H_o^{(0)}, H_o^{(\pi)}$ and $H_o^{(\pi/2)}$, are also available by constructing the associated F-statistics. In the following analysis, for different sample periods, the critical values and the p-values associated with these test statistics for seasonal unit roots are obtained from the simulated distributions constructed by 50,000 Monte Carlo replications; the data-generating process is $\Delta_4 w_{it} \sim i.i.d.N(0,1)$, and no augmented lags ($l_i = 0$) are considered in the regression model (6).

3.2.2 Pooled augmented HEGY tests for panels

Based on the introduced augmented HEGY test for univariate series w_{it}, we are now going to extend them to detect the seasonal non-stationarity (at different seasonal frequencies) for the panel of w_{it}, $i = 1, \dots, N$, $t = 1, \dots, T$.

Given OLS estimates of the model (6) for variable w_i ($i = 1, \dots, N$), let $p_i^{(\kappa)}$ denote the p-value associated with HEGY test statistic for the null $H_o^{(\kappa)}$, where $\kappa = 0, \pi, \pi/2, S, SI$, respectively, indicate the different types of (seasonal) unit roots; the proposed pooled test statistic for testing the null hypothesis that all variables in a panel have this $\kappa-$type (seasonal) unit root is expressed as:

$$\Upsilon^{(\kappa)}_{pool} = -2 \sum_{i=1}^{N} \ln p_i^{(\kappa)}. \tag{7}$$

If these p-values are uniformly distributed on the interval $[0,1]$ and independent across i, the distribution for the summation of these logarithm of p-values over i as proposed in (7) will be χ^2 random variable with $2N$ degrees of freedom, $\chi^2(2N)$ say, since we have $-2\ln p_i^{(\kappa)} \sim \chi^2(2)$. This type of test statistic was originally suggested by Fisher (1932) for meta analysis and was introduced to panel unit root tests by Maddala and Wu (1999). Note that the independence across series w_i, $i = 1, \dots, N$, is necessary for the test (7) to be valid. It immediately follows that the panel HEGY test for $w_{it} = y_{it}$ might be inappropriate since for all $i \neq j$, when y_{it} and y_{jt} driven by the same set of common factors $\boldsymbol{F}_t$ in model (1) are dependent.[6] However, as argued by Bai and Ng (2004), this independence requirement across series seems plausible for $w_{it} = \epsilon_{it}$ (and its estimate) because most of the cross-section dependence should be captured by common factors instead of idiosyncratic errors. For more discussions on this type of test statistics, refer to Maddala and Wu (1999), Choi (2001), and Bai and Ng (2004).

3.2.3 Determining the number of stochastic common trends

When there are multiple factors ($q^* > 1$), we propose the following successive procedure to determine the number of (seasonal) non-stationary factors (and thus the number of stationary ones) by using the transformed estimates of factors $\hat{\boldsymbol{F}}_t^* := [\hat{F}_{1t}^*, \hat{F}_{2t}^*, \dots, \hat{F}_{\hat{q}t}^*]'$ which are constructed by (5). In what follows, let $q_1^{(\kappa)}$ and $\hat{q}_1^{(\kappa)}$ denote the true and the estimated number of $\kappa-$type (seasonal) non-stationary factors, respectively, where $\kappa = 0, \pi, \pi/2, S, SI$, refers to the cases where the factor has a conventional unit root (+1), semi-annual unit root (−1), annual unit roots ($\pm i$), all seasonal unit roots, and the factor is $SI(1)$, respectively.

Given the $m-$th ($m = 1, 2, \dots, \hat{q}$) transformed estimates of factor $\hat{F}_{mt}^*$, and its p-value associated with HEGY test statistic for the null $H_o^{(\kappa)}$ ($\kappa = 0, \pi, \pi/2, S, SI$), $p_m^{(\kappa)}$ say, the proposed procedure involves sequentially testing $H_o : q_1^{(\kappa)} = n$ versus $H_1 : q_1^{(\kappa)} \le n-1$, for $n = \hat{q}, \hat{q}-1, \dots, 1$, at the significance level α. Under the null $H_o : q_1^{(\kappa)} = n$, the proposed test statistic is directly based on the p-value $p_n^{(\kappa)}$; we compare $p_n^{(\kappa)}$ with α and this null is rejected only if $p_n^{(\kappa)} \le \alpha$. The testing sequence terminates when the null is not rejected for the first time. As a consequence, the estimated number of $\kappa-$type (seasonal) non-stationary factors, $\hat{q}_1^{(\kappa)} = n^*$ say, is thus determined by:

$$\left\{p_{\hat{q}}^{(\kappa)} \le \alpha, p_{\hat{q}-1}^{(\kappa)} \le \alpha, \cdots, p_{n^*+1}^{(\kappa)} \le \alpha, p_{n^*}^{(\kappa)} > \alpha\right\}. \tag{8}$$

It should be noted that the involved successive tests are not independent to each other because they are applied to the same estimates of factors $\hat{\boldsymbol{F}}_t^*$. Therefore, it is supposed they share the same asymptotic properties with the conventional rank tests such as MQ_f and MQ_c tests of Bai and Ng (2004) for testing the rank of non-stationary space of factors, or the trace or eigenvalue tests of Johansen (1995) for determining the number of cointegrating vectors; given the significance level α in each step, the probability of choosing the true number $q_1^{(\kappa)}$ by using the above successive tests would approach $1-\alpha$; thus, the overall asymptotic type I error is α. Whether this proposed procedure holds this property will be investigated carefully by simulations in Section 5.2.

4 Assumptions and asymptotic properties

Given the model specification for y_{it} in (1), we now clearly specify the dynamic structures of the common factor $\boldsymbol{F}_t$ and idiosyncratic errors ϵ_{it}, They are:

$$(I - L^4)\boldsymbol{F}_t = \Delta_4 \boldsymbol{F}_t = \boldsymbol{B}_F(L)\boldsymbol{u}_t, \tag{9}$$

dollar

(9)

$$(1 - L^4)\epsilon_{it} = \Delta_4 \epsilon_{it} = B_{\epsilon i}(L) v_{it}, \tag{10}$$

where $\boldsymbol{B}_F(L) = \sum_{j=0}^{\infty} \boldsymbol{B}_{Fj} L^j$ and $B_{\epsilon i}(L) = \sum_{j=0}^{\infty} B_{\epsilon ij} L^j$. Note that theses specifications for $\boldsymbol{F}_t$ and ϵ_{it} extend from the PANIC model of Bai and Ng (2004) in which seasonal unit roots were not involved. Let r_s (= $+1, -1, +i$ and $-i$) denote a root of L for the polynomial $(1 - L^4) = 0$; if ϵ_{it} in model (10) has a (seasonal) unit root r_s, then $B_{\epsilon i}(r_s) \neq 0$; otherwise, $B_{\epsilon i}(r_s) = 0$. Similarly, for the $q^* \times 1$ vector $\boldsymbol{F}_t$, $\boldsymbol{B}_F(r_s)$ has rank δ_s $(0 \le \delta_s \le q^*)$ once there are δ_s factors with this unit root. These (rank) conditions associated with seasonal unit roots come from the Wald representation for the process; for more details please refer to the Appendix.

Denote $\|\boldsymbol{A}\| = \mathrm{trace}(\boldsymbol{A}'\boldsymbol{A})^{1/2}$; then following Bai and Ng (2004), the regularity conditions for the proposed estimation and testing procedures to be valid are:

A.1 The loadings λ_i, the errors $\boldsymbol{u}_t$ and v_{it} are three mutually independent groups.

A.2 $\mathbb{E}[\|\boldsymbol{\lambda}_i\|^4] < \infty$, and $(1/N)\sum_{i=1}^{N} \boldsymbol{\lambda}_i \boldsymbol{\lambda}_i' \xrightarrow{P} \Sigma_\Lambda$ as $N \to \infty$, where Σ_Λ is a $q^* \times q^*$ positive definite matrix.

A.3 (i) $\boldsymbol{u}_t \sim iid(0, \Sigma_u)$, $\mathbb{E}[\|\boldsymbol{u}_t\|^4] < \infty$; (ii) $\sum_{j=0}^{\infty} j\|\boldsymbol{B}_{Fj}\| < \infty$, $\mathrm{var}(\Delta_4 \boldsymbol{F}_t) = \sum_{j=0}^{\infty} \boldsymbol{B}_{Fj} \Sigma_u \boldsymbol{B}_{Fj}' > 0$; (iii) $\boldsymbol{B}_F(r_s)$ has rank $\delta_s, 0 \le \delta_s \le q^*, s = 1, \dots, 4$.

A.4 (i) For each i, $v_{it} \sim iid(0, \sigma_{v_i}^2)$, $\mathbb{E}[|v_{it}|^8] < \infty$; (ii)$\sum_{j=0}^{\infty} j|B_{\epsilon ij}| < \infty$, $\mathrm{var}(\Delta_4 \epsilon_{it}) = \sum_{j=0}^{\infty} B_{\epsilon ij}^2 \sigma_{v_i}^2 > 0$; (iii) v_{it} are independent over i;

A.5 For $\tau = -3, -2, -1, 0$, $\mathbb{E}[\|F_\tau\|] < \infty$, and for every $i = 1, \ldots, N$, $\mathbb{E}[|\epsilon_{i\tau}|] < \infty$.

As compared with Bai and Ng's (2004) assumptions for the validity of the PANIC approach, we further introduce A.3(iii) and A.5 to deal with the panel data with possible seasonal unit roots at different frequencies, and directly impose the independence assumption for v_{it} in A.4(iii) for the validity of the proposed panel test (7) by pooling the associated p-values computed from the HEGY tests for the idiosyncratic errors and their estimates.[7]

Given these assumptions, what immediately follows is below:

Theorem 4.1

Suppose the data driven by q^* common factors are generated by (1), (9) and (10), and Assumptions A.1 to A.5 hold. As $N, T \to \infty$ and $T/N \to 0$, we have:

a. $\hat{q} \xrightarrow{P} q^*$ *where* $\hat{q}$ *is determined by (3).*

b. *The space spanned by* $\hat{\boldsymbol{F}}_t$ *is a consistent estimate for the space spanned by* $\boldsymbol{F}_t$ *up to a location shift.*

c. $\hat{\epsilon}_{it} \xrightarrow{P} \epsilon_{it}$ *for all i.*

Theorem 4.1(a) directly follows Result C in Bai and Ng (2008) or Corollary 1 in Bai and Ng (2002) because the chosen penalty weights $\Gamma_1(N,T)$ to $\Gamma_3(N,T)$ in (3) are valid; Theorem 4.1(b) and (c) are the straightforward implications of Lemma 1 and Lemma 2 in Bai and Ng (2004), since we just generalize the PANIC model of Bai and Ng (2004) to seasonal panel unit root models.

Given the consistency of the estimates of the model, for ϵ_{it} and $\hat{\epsilon}_{it}$, $i = 1, \ldots, N$, to test the null $H_o^{(\kappa)}$ ($\kappa = 0, \pi, \pi/2, S, SI$), assumption A.4(iii) ensures that the associated p-values, $p_i^{(\kappa)}$, $i = 1, \ldots, N$, computed from the associated HEGY tests, are independent uniform random variables over the interval $[0,1]$; thus, the logarithm of the p-value times minus two is a χ^2 random variable with two degrees of freedom. Summarizing these logarithm of p-values over i for fixed N, we have $\Upsilon_{pool}^{(\kappa)}$ proposed in (7) which follows a χ^2 distribution with $2N$ degrees of freedom. Let $N \to \infty$; we have the following result by the Central Limit Theory.

Theorem 4.2

Suppose the assumptions of Theorem 4.1 hold, for $i = 1, \ldots, N$, given ϵ_{it} (or $\hat{\epsilon}_{it}$) in the panel, let $p_i^{(\kappa)}$ denote its p-value associated with HEGY test statistic for the $\kappa-$type null hypothesis $H_o^{(\kappa)}$, where $\kappa = 0, \pi, \pi/2, S, SI$, then for the pooled test statistic $\Upsilon_{pool}^{(\kappa)} = -2\sum_{i=1}^{N} \ln p_i^{(\kappa)}$, we have:

$$\frac{\Upsilon_{pool}^{(\kappa)} - 2N}{\sqrt{4N}} \xrightarrow{d} N(0,1),$$

the standardized pooled test statistic, under the given null hypothesis that all variables in the panel have this $\kappa-$type (seasonal) unit root, converges in distribution to a standard normal distribution.

5 Simulations

In the section we evaluate the small-sample properties of the proposed estimation and testing procedures. For $i = 1, \ldots, N$ and $t = 1, \ldots, T$, the data generating process (DGP) is:

$$\begin{aligned} y_{it} &= \boldsymbol{\lambda}_i' \boldsymbol{F}_t + \epsilon_{it}, & \boldsymbol{\lambda}_i &\sim MN(\boldsymbol{0}, I_{q^*}), \\ \boldsymbol{F}_t &= \boldsymbol{\alpha} \boldsymbol{F}_{t-4} + \boldsymbol{u}_t, & \boldsymbol{u}_t &\sim MN(\boldsymbol{0}, \sigma_F^2 I_{q^*}), \\ \epsilon_{it} &= \rho \epsilon_{it-4} + v_{it}, & v_{it} &\sim N(0,1), \end{aligned} \tag{11}$$

where $\boldsymbol{F}_t$ is an $q^* \times 1$ vector of common factors, $\boldsymbol{\lambda}_i$ is a corresponding $q^* \times 1$ vector of factor loadings, $\boldsymbol{\alpha}$ (which drives the dynamic patterns of $\boldsymbol{F}_t$) is a $q^* \times q^*$ diagonal matrix of coefficients with diagonal elements $\alpha_1, \alpha_2, \ldots, \alpha_{q^*}$. We assume that both $\boldsymbol{\lambda}_i$ and $\boldsymbol{u}_t$ follow zero-mean multivariate normal (MN) distributions, and v_{it} is a standard normal random variable. Note that this DGP is quite similar to what was proposed in the simulations of Bai and Ng (2004) except that the current common factor $\boldsymbol{F}_t$ and idiosyncratic errors ϵ_{it} ($i = 1, \ldots, N$) are directly affected, respectively, by $\boldsymbol{F}_{t-4}$ and ϵ_{it-4} (instead of $\boldsymbol{F}_{t-1}$ and ϵ_{it-1} in the PANIC framework).

While applying the HEGY test and/or its pooled version to the variables of interest, five null hypotheses considered are $H_o^{(\kappa)}$, $\kappa = 0, \pi, \pi/2, S, SI$, which refer to the null that the variable has a conventional unit root (+1), semi-annual unit root (-1), annual unit roots ($\pm i$), all seasonal unit roots, and the variable is $SI(1)$, respectively; cf., Section 3.2.1. Besides, for simplicity, no lags will be introduced while estimating the corresponding auxiliary regression (6). In all cases of simulations, three different sizes of panel data with $(N = 40, T = 100)$, $(N = 100, T = 100)$ and $(N = 200, T = 100)$ are generated; the nominal size is 5%, and the number of Monte Carlo replications is 5000.

5.1 Univariate common factor: $q^* = 1$

When $q^* = 1$, there is only one common factor in the model. In DGP (11), F_t (resp. ϵ_{it}) is seasonally integrated of order one, $SI(1)$, by setting $\alpha_1 = 1$ (resp. $\rho = 1$); otherwise, it is stationary. Besides, we consider $\sigma_F^2 = 1$ which is equal to the variance of idiosyncratic errors. The variables being tested are y_{it}, $\hat{F}_t$, $\hat{\epsilon}_{\mathrm{it}}$ and ϵ_{it}, where $\hat{F}_t$ and $\hat{\epsilon}_{\mathrm{it}}$ are estimated based on the proposed method in Section 3.1. The univariate HEGY test is conducted for testing $\hat{F}_t$ while pooled HEGY tests are employed for all the other variables in all the cases; for details, refer to Sections 3.2.1 and 3.2.2. Moreover, in order to investigate the validity of the proposed pooled test via statistic (7), the performance of the testing panel of ϵ_{it} is also revealed. Three classes of settings for α_1 and ρ are considered; only the idiosyncratic errors are $SI(1)$, only the common factor is $SI(1)$, and both F_t and ϵ_{it} are either $SI(1)$ or stationary. The corresponding rejection rates for the tests are reported in Table 1.

Table 1: Rejection rates for one common factor cases.

			N = 40, T = 100					N = 100, T = 100					N = 200, T = 100				
ρ	α_1	w_{it}	$\kappa = 0$	π	$\pi/2$	S	SI	$\kappa = 0$	π	$\pi/2$	S	SI	$\kappa = 0$	π	$\pi/2$	S	SI
(A) ϵ_{it} are $SI(1)$																	
1.00	0.00	y_{it}	0.834	0.830	0.399	0.656	0.804	0.985	0.987	0.608	0.893	0.966	0.999	1.000	0.775	0.973	0.995
		$\widehat{F}_t$	0.924	0.913	0.968	0.992	0.999	0.992	0.991	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\hat{\epsilon}_{it}$	0.056	0.056	0.055	0.063	0.063	0.052	0.051	0.059	0.062	0.060	0.057	0.048	0.066	0.067	0.060
		ϵ_{it}	0.051	0.055	0.051	0.056	0.055	0.057	0.054	0.051	0.055	0.050	0.065	0.059	0.055	0.057	0.051
1.00	0.50	y_{it}	0.551	0.534	0.126	0.242	0.347	0.841	0.824	0.182	0.408	0.568	0.966	0.955	0.252	0.563	0.752
		$\widehat{F}_t$	0.691	0.686	0.778	0.899	0.950	0.849	0.844	0.949	0.987	0.997	0.919	0.914	0.988	0.998	0.999
		$\hat{\epsilon}_{it}$	0.052	0.046	0.050	0.051	0.051	0.057	0.052	0.052	0.060	0.060	0.060	0.049	0.065	0.067	0.062
		ϵ_{it}	0.051	0.047	0.049	0.050	0.052	0.059	0.055	0.047	0.054	0.053	0.062	0.052	0.056	0.057	0.053
(B) F_t is $SI(1)$																	
0.00	1.00	y_{it}	0.960	0.962	0.991	0.997	0.998	0.992	0.991	1.000	1.000	1.000	0.999	0.998	1.000	1.000	1.000
		$\widehat{F}_t$	0.054	0.058	0.045	0.047	0.047	0.053	0.051	0.046	0.046	0.050	0.045	0.052	0.050	0.049	0.048
		$\hat{\epsilon}_{it}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		ϵ_{it}	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.50	1.00	y_{it}	0.805	0.801	0.859	0.956	0.972	0.856	0.843	0.946	0.994	0.995	0.883	0.880	0.984	0.999	1.000
		$\widehat{F}_t$	0.055	0.051	0.048	0.045	0.047	0.051	0.057	0.045	0.050	0.050	0.050	0.050	0.049	0.046	0.050
		$\hat{\epsilon}_{it}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		ϵ_{it}	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(C) Others																	
1.00	1.00	y_{it}	0.160	0.148	0.134	0.130	0.131	0.236	0.218	0.188	0.194	0.185	0.303	0.296	0.237	0.240	0.233
		$\widehat{F}_t$	0.053	0.050	0.053	0.047	0.047	0.048	0.041	0.052	0.047	0.048	0.050	0.047	0.051	0.053	0.053
		$\hat{\epsilon}_{it}$	0.053	0.048	0.046	0.048	0.050	0.062	0.056	0.054	0.059	0.055	0.063	0.053	0.053	0.059	0.054
		ϵ_{it}	0.051	0.053	0.044	0.052	0.050	0.062	0.054	0.051	0.054	0.053	0.063	0.054	0.051	0.057	0.052
0.50	0.50	y_{it}	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\widehat{F}_t$	0.973	0.976	1.000	1.000	1.000	0.973	0.976	0.999	1.000	1.000	0.979	0.975	0.999	1.000	1.000
		$\hat{\epsilon}_{it}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		ϵ_{it}	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

1. The data are generated as $\epsilon_{it} = \rho\epsilon_{it-4} + v_{it}$, $F_t = \alpha_1 F_{t-4} + u_t$, and $y_{it} = \lambda_i F_t + \epsilon_{it}$.
2. $\kappa = 0, \pi, \pi/2, S, SI$, refers to the null that the factor has a conventional unit root (+1), semi-annual unit root (−1), annual unit roots ($\pm i$), all seasonal unit roots, and the factor is $SI(1)$, respectively.
3. Rows follow y_{it}, $\hat{\epsilon}_{it}$, and ϵ_{it}, are rejection rates of corresponding pooled HEGY tests for different seasonal unit root tests; the finite sample distribution of $\Upsilon^{(\kappa)}_{pool}$ considered here is $\chi^2(2N)$. Rows follow $\widehat{F}_t$ are the rejection rates for its corresponding univariate HEGY tests. The nominal size is 5%.

When only the idiosyncratic errors are $SI(1)$, panel (A) in Table 1 reports the average rejection rates for the univariate/pooled HEGY tests over 5000 replications. First, the results for ϵ_{it} indicate that, without estimation bias, the performance of the proposed panel tests for these five hypotheses is quite good since all the rejection rates are around the nominal size 5%. Besides, the similar results for testing the estimates $\hat{\epsilon}_{\text{it}}$ implies that the proposed estimation method works very well, and that these results are not sensitive to the parameter α_1 which regulates the dynamics of F_t. While testing the estimate of common factor F_t, we observe that the powers of most tests are greater than 0.9, and they increase when N increases from 40 to 100 (or to 200); e.g. as in case that $\rho = 1$ and $\alpha_1 = 0.5$, the average rejection rates for $\kappa = S$ null is 0.899, 0.987, and 0.998, when $N = 40, 100$ and 200, respectively. On the other hand, every y_{it} in fact is $SI(1)$ since it is the combination of a stationary F_t (with random loading λ_i) and an $SI(1)$ ϵ_{it}. However, the panel tests for y_{it} do not successfully reveal this feature; the rejection rates for all panel tests are much larger than the nominal size. This phenomenon is expected since y_{it} following the common factor structure is correlated across N units; the cross-sectional correlations may lead the pooled HEGY test to over-reject the null hypothesis; see e.g. O'Connell (1998) and Bai and Ng (2004).

On the other hand, when only the common factor is $SI(1)$, the rejection rates in panel (B) show that the performance of testing ϵ_{it} and $\hat{\epsilon}_{\text{it}}$ are the same (powers are equal to one in all cases), the average rejection rates of univariate HEGY tests for F_t are around 5%, and the panel tests tend to over-reject all the null hypotheses for y_{it}. When both F_t and ϵ_{it} are $SI(1)$ ($\rho = \alpha_1 = 1$), panel (C) shows that over-rejections for testing y_{it} are still observed but less serious. When both F_t and ϵ_{it} are stationary, all the rejection rates for y_{it}, ϵ_{it} and $\hat{\epsilon}_{\text{it}}$ are equal to one while the rejection rates for F_t are above 0.97 for all tests; these results can be expected since all the tests for variables of interest, except F_t, are pooled tests which aim at increasing the testing power.

In brief, when $q^* = 1$, these simulation results show that the performances of testing the estimated common factor and idiosyncratic errors in general are good, without knowing a *priori* whether or not the series are seasonal stationary. These results are encouraging; as the direct panel tests on the data with common factor structure may be misleading, our proposed approach successfully indicates the sources of the seasonal non-stationarity via testing the consistent estimates of the common factor and idiosyncratic errors instead. The proposed procedure is an extension of the PANIC framework of Bai and Ng (2004) to the seasonal panel, and works very well when there is one common stochastic trend.

5.2 Multiple common factors: $q^* > 1$

When there are q^* common factors, we generate $q_1^{(SI)}$ $SI(1)$ factors by setting $\alpha_j = 1, j = 1, \ldots, q_1^{(SI)}$, and $q^* - q_1^{(SI)}$ stationary factors with coefficient $\alpha_j = 0.5$ for $j = q_1^{(SI)} + 1, \ldots, q^*$, in DGP (11). In the simulations, we set the total number of common factors $q^* = 3$ and consider the cases in which $q_1^{(SI)}$ varies from 0 to 3. Besides, the idiosyncratic errors are allowed to be $SI(1)$ or stationary by setting $\rho = 1$ or 0.5, and the variation of factors could be equal to, or larger than, those of the idiosyncratic errors by letting $\sigma_F^2 = 1$ or 10; there are four classes of simulations associated with different combinations of ρ and σ_F^2. These simulation settings mimic what was introduced in Bai and Ng (2004), whereas we consider seasonal integrated common trends and errors in the seasonal panel.

Table 2: Probability of selecting the correct number of common factors.

		$N = 40, T = 100$				$N = 100, T = 100$				$N = 200, T = 100$			
q^*	$q_1^{(SI)}$	Γ_1	Γ_2	Γ_3	Γ_4	Γ_1	Γ_2	Γ_3	Γ_4	Γ_1	Γ_2	Γ_3	Γ_4
(A) $\sigma_F^2 = 1, \rho = 1$													
3	0	1.000	1.000	1.000	1.000	1.000	1.000	0.996	1.000	1.000	1.000	1.000	1.000
3	1	1.000	1.000	1.000	1.000	1.000	1.000	0.997	1.000	1.000	1.000	1.000	1.000
3	2	1.000	1.000	1.000	1.000	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000
3	3	1.000	1.000	1.000	0.999	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000
(B) $\sigma_F^2 = 1, \rho = 0.5$													
3	0	1.000	1.000	1.000	0.999	1.000	1.000	0.698	1.000	1.000	1.000	1.000	1.000
3	1	1.000	1.000	1.000	0.997	1.000	1.000	0.700	1.000	1.000	1.000	1.000	1.000
3	2	1.000	1.000	1.000	0.995	1.000	1.000	0.722	1.000	1.000	1.000	1.000	1.000
3	3	1.000	1.000	0.999	0.992	1.000	1.000	0.719	1.000	1.000	1.000	1.000	1.000
(C) $\sigma_F^2 = 10, \rho = 1$													
3	0	1.000	1.000	1.000	1.000	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000
3	1	1.000	1.000	1.000	1.000	1.000	1.000	0.997	1.000	1.000	1.000	1.000	1.000
3	2	1.000	1.000	1.000	1.000	1.000	1.000	0.996	1.000	1.000	1.000	1.000	1.000
3	3	1.000	1.000	1.000	1.000	1.000	1.000	0.996	1.000	1.000	1.000	1.000	1.000

(D) $\sigma_F^2 = 10, \rho = 0.5$													
3	0	1.000	1.000	1.000	1.000	1.000	1.000	0.694	1.000	1.000	1.000	1.000	1.000
3	1	1.000	1.000	1.000	1.000	1.000	1.000	0.706	1.000	1.000	1.000	1.000	1.000
3	2	1.000	1.000	1.000	1.000	1.000	1.000	0.710	1.000	1.000	1.000	1.000	1.000
3	3	1.000	1.000	1.000	1.000	1.000	1.000	0.729	1.000	1.000	1.000	1.000	1.000

1. The data are generated as $\epsilon_{it} = \rho\epsilon_{it-4} + v_{it}$, $\boldsymbol{F}_t = \boldsymbol{\alpha}\boldsymbol{F}_{t-4} + u_t$, and $y_{it} = \boldsymbol{\lambda}_i'\boldsymbol{F}_t + \epsilon_{it}$.
2. q^* and $q_1^{(SI)}$ refer to the total number of common factors and *SI* factors, respectively. Γ_1 to Γ_4 refer to the criteria from $\Gamma_1(N,T)$ to $\Gamma_4(N,T)$ introduced in Section 3.1.
3. The entries are the probabilities of selecting the correct number of common factors based on the information criterion (3).

First, we evaluate the performance of estimating the total number of common factors (q^*) by using criterion (3) for the annual-differenced series $\Delta_4 y_{it}$. Given $\bar{q} = 6$ (the maximum number of factors) in (3) and four introduced penalty functions from $\Gamma_1(N,T)$ to $\Gamma_4(N,T)$, the probability of selecting the correct number of common factors via 5000 replications are summarized in Table 2. Except for some cases with $\Gamma_3(N,T)$ as the penalty function, all the results imply that the introduced selection criterion which generates the case of PANIC model of Bai and Ng (2004) to the proposed quarterly factor model works very well; the true number of common factors, $q^* = 3$, can be precisely estimated by using the annual-differenced data, even without knowing whether the original data is stationary or $SI(1)$.[8]

Table 3: Average trace-R^2 for multiple common factors cases.

		N = 40, T = 100			**N = 100, T = 100**			**N = 200, T = 100**		
q^*	$q_1^{(SI)}$	$\tilde{q} = 2$	3	4	$\tilde{q} = 2$	3	4	$\tilde{q} = 2$	3	4
(A) $\sigma_F^2 = 1, \rho = 1$										
3	0	0.523	0.719	0.761	0.613	0.849	0.875	0.661	0.915	0.930
3	1	0.824	0.958	0.966	0.819	0.980	0.984	0.811	0.989	0.991
3	2	0.850	0.982	0.986	0.845	0.992	0.994	0.844	0.996	0.997
3	3	0.843	0.990	0.992	0.853	0.996	0.997	0.852	0.998	0.998
(B) $\sigma_F^2 = 1, \rho = 0.5$										
3	0	0.691	0.974	0.975	0.707	0.990	0.990	0.715	0.995	0.995
3	1	0.862	0.996	0.996	0.839	0.999	0.999	0.821	0.999	0.999
3	2	0.869	0.998	0.998	0.854	0.999	0.999	0.844	1.000	1.000
3	3	0.848	0.999	0.999	0.854	1.000	1.000	0.855	1.000	1.000
(C) $\sigma_F^2 = 10, \rho = 1$										
3	0	0.676	0.950	0.960	0.700	0.980	0.984	0.710	0.990	0.992
3	1	0.861	0.994	0.995	0.838	0.998	0.998	0.820	0.999	0.999
3	2	0.866	0.998	0.998	0.858	0.999	0.999	0.848	1.000	1.000
3	3	0.849	0.999	0.999	0.852	1.000	1.000	0.854	1.000	1.000
(D) $\sigma_F^2 = 10, \rho = 0.5$										
3	0	0.703	0.997	0.997	0.712	0.999	0.999	0.718	1.000	1.000
3	1	0.863	1.000	1.000	0.839	1.000	1.000	0.821	1.000	1.000
3	2	0.870	1.000	1.000	0.853	1.000	1.000	0.848	1.000	1.000
3	3	0.850	1.000	1.000	0.852	1.000	1.000	0.856	1.000	1.000

1. The data are generated as $\epsilon_{it} = \rho\epsilon_{it-4} + v_{it}$, $\boldsymbol{F}_t = \boldsymbol{\alpha}\boldsymbol{F}_{t-4} + u_t$, and $y_{it} = \boldsymbol{\lambda}_i'\boldsymbol{F}_t + \epsilon_{it}$.
2. q^* and $q_1^{(SI)}$ refer to the total number of common factors and *SI* factors, respectively.
3. Given the number of estimated factors $\tilde{q}$, the trace R^2 is computed as

$$\text{trace-}R^2(\tilde{q}) = \frac{\text{Trace}\left(\boldsymbol{F}'\boldsymbol{Z}(\tilde{q})\left(\boldsymbol{Z}(\tilde{q})'\boldsymbol{Z}(\tilde{q})\right)^{-1}\boldsymbol{Z}(\tilde{q})'\boldsymbol{F}\right)}{\text{Trace}\left(\boldsymbol{F}'\boldsymbol{F}\right)},$$

where $\boldsymbol{F} = [\boldsymbol{F}_1, \boldsymbol{F}_2, \ldots, \boldsymbol{F}_T]'$ is a $T \times q^*$ matrix of true factors, $\boldsymbol{Z}(\tilde{q}) = [\boldsymbol{1}_T, \widehat{\boldsymbol{F}}(\tilde{q})]$ is the $T \times (\tilde{q}+1)$ matrix with $\boldsymbol{1}_T$, a $T \times 1$ vector of ones, and $\widehat{\boldsymbol{F}}(\tilde{q}) = [\widehat{\boldsymbol{F}}_1(\tilde{q}), \widehat{\boldsymbol{F}}_2(\tilde{q}), \ldots, \widehat{\boldsymbol{F}}_T(\tilde{q})]'$, a $T \times \tilde{q}$ matrix of $\tilde{q}$ estimated factors.

Since the total number of common factors can be precisely estimated, as verified, the consistency result claimed in Theorem 4.1(b) where the space spanned by $\widehat{\boldsymbol{F}}_t$ is a consistent estimate for the space spanned by the true factor $\boldsymbol{F}_t$ up to a location shift, is then verified as follows. When the considered number of factors is $\tilde{q}$, we compute the trace-R^2 as:

$$\text{trace-}R^2(\tilde{q}) = \frac{\text{Trace}\left(\boldsymbol{F}'\boldsymbol{Z}(\tilde{q})\left(\boldsymbol{Z}(\tilde{q})'\boldsymbol{Z}(\tilde{q})\right)^{-1}\boldsymbol{Z}(\tilde{q})'\boldsymbol{F}\right)}{\text{Trace}\left(\boldsymbol{F}'\boldsymbol{F}\right)}, \quad (12)$$

where $\boldsymbol{F} = [\boldsymbol{F}_1, \boldsymbol{F}_2, \ldots, \boldsymbol{F}_T]'$ is a $T \times q^*$ matrix of true factors, $\boldsymbol{Z}(\tilde{q}) = [\boldsymbol{1}_T, \hat{\boldsymbol{F}}(\tilde{q})]$ is the $T \times (\tilde{q}+1)$ matrix with $\boldsymbol{1}_T$, a $T \times 1$ vector of ones, and $\hat{\boldsymbol{F}}(\tilde{q}) = [\hat{\boldsymbol{F}}_1(\tilde{q}), \hat{\boldsymbol{F}}_2(\tilde{q}), \ldots, \hat{\boldsymbol{F}}_T(\tilde{q})]'$ is a $T \times \tilde{q}$ matrix of estimated factors. The trace-R^2 is a multivariate version of the typical R^2 of the regression of the true factors on a constant and the estimated factors, and the number close to one indicates a good approximation of the true common factors; for e.g. see Stock and Watson (2002a). For each simulation setting, Table 3 presents the average trace-R^2 over 5000 replications when the number of factors is under-estimated ($\tilde{q} = 2$), exact-estimated ($\tilde{q} = 3$), and over-estimated ($\tilde{q} = 4$). As expected, increasing N or σ_F^2 cannot help to improve the fitting performance once the number of factors is under-estimated as $\tilde{q} = 2$, all the average trace-R^2 are lower than 0.88. By contrast, when $\tilde{q} = 3$, almost all the average trace-R^2 are greater than 0.95 and they get close to 1 as N or σ_F^2 increases. Besides, except for the cases in panel (A), introducing an additional estimated factor could not improve the fitting performance; the average trace-R^2 of $\tilde{q} = 4$ is quite similar to that of $\tilde{q} = 3$ for all the configurations.

Table 4: Probability of selecting the correct number of common stochastic trends.

		$N = 40, T = 100$					$N = 100, T = 100$					$N = 200, T = 100$				
q^*	$q_1^{(SI)}$	$\kappa = 0$	π	$\pi / 2$	S	SI	$\kappa = 0$	π	$\pi / 2$	S	SI	$\kappa = 0$	π	$\pi / 2$	S	SI
(A) $\sigma_F^2 = 1, \rho = 1$																
3	0	0.338	0.327	0.473	0.707	0.837	0.627	0.594	0.829	0.951	0.987	0.754	0.745	0.938	0.988	0.997
3	1	0.544	0.535	0.683	0.828	0.891	0.751	0.742	0.883	0.938	0.951	0.850	0.824	0.935	0.953	0.952
3	2	0.750	0.743	0.840	0.894	0.904	0.852	0.847	0.911	0.913	0.914	0.872	0.877	0.920	0.921	0.911
3	3	0.866	0.871	0.855	0.826	0.809	0.874	0.878	0.858	0.839	0.813	0.872	0.873	0.861	0.841	0.818
(B) $\sigma_F^2 = 1, \rho = 0.5$																
3	0	0.929	0.918	0.995	1.000	1.000	0.921	0.922	0.995	1.000	1.000	0.922	0.924	0.995	1.000	1.000
3	1	0.905	0.903	0.958	0.957	0.955	0.905	0.912	0.952	0.951	0.953	0.915	0.906	0.953	0.956	0.957
3	2	0.894	0.903	0.925	0.918	0.909	0.907	0.902	0.924	0.918	0.913	0.903	0.902	0.925	0.919	0.909
3	3	0.863	0.870	0.861	0.835	0.807	0.858	0.870	0.856	0.824	0.801	0.868	0.869	0.856	0.833	0.812
(C) $\sigma_F^2 = 10, \rho = 1$																
3	0	0.848	0.858	0.981	0.998	1.000	0.891	0.879	0.991	0.999	1.000	0.902	0.896	0.992	1.000	1.000
3	1	0.875	0.876	0.948	0.951	0.954	0.892	0.891	0.954	0.951	0.950	0.901	0.897	0.950	0.955	0.951
3	2	0.891	0.886	0.926	0.920	0.914	0.900	0.909	0.919	0.916	0.909	0.903	0.900	0.922	0.916	0.907
3	3	0.870	0.875	0.858	0.843	0.813	0.875	0.875	0.863	0.841	0.812	0.865	0.875	0.870	0.848	0.820
(D) $\sigma_F^2 = 10, \rho = 0.5$																
3	0	0.922	0.916	0.994	0.999	1.000	0.917	0.913	0.994	1.000	1.000	0.913	0.905	0.995	1.000	1.000
3	1	0.904	0.918	0.954	0.952	0.956	0.903	0.904	0.947	0.947	0.946	0.911	0.897	0.953	0.954	0.956
3	2	0.905	0.899	0.921	0.919	0.915	0.898	0.904	0.924	0.920	0.911	0.902	0.902	0.925	0.921	0.916
3	3	0.873	0.866	0.859	0.837	0.811	0.871	0.874	0.858	0.835	0.818	0.874	0.870	0.867	0.842	0.817

1. The data are generated as $\epsilon_{it} = \rho\epsilon_{it-4} + v_{it}$, $\boldsymbol{F}_t = \boldsymbol{\alpha}\boldsymbol{F}_{t-4} + u_t$, and $y_{it} = \boldsymbol{\lambda}_i'\boldsymbol{F}_t + \epsilon_{it}$.
2. q^* and $q_1^{(SI)}$ refer to the total number of common factors and *SI* factors, respectively.
3. $\kappa = 0, \pi, \pi/2, S, SI$, refers to the cases that the factor has a conventional unit root (+1), semi-annual unit root (−1), annual unit roots ($\pm i$), all seasonal unit roots, and the factor is *SI*(1), respectively.
4. The entries are the probabilities of selecting the correct number of *SI*(1) factors based on the proposed statistic in the successive procedure; the nominal size is 5%.

After evaluating the space spanned by the estimated factors, Table 4 further reports the probabilities of correctly selecting the number of $\kappa-$type common stochastic trends over 5000 replications by using the successive testing procedure proposed in Section 3.2.3 under different settings of ρ and σ_F^2.[9] Panel A reports the results when the idiosyncratic errors are $SI(1)$ and the variance of the common factor equals that of the errors. When $N = 40$, the accuracy rates of selecting the true number of factors with conventional ($\kappa = 0$), semi-annual ($\kappa = \pi$) and annual ($\kappa = \pi/2$) unit roots are not high (below 0.6) when $q_1^{(SI)} = 0$ or 1. However, this situation improves if N increases from 40 to 100 (even to 200), and/or σ_F^2 increases from 1 to 10 (in panel C), and/or the errors are stationary (in panel B). Including more cross-sectional units or a larger variation of common trends relative to that of errors would help us to estimate the number of the common trends with different stochastic types indicated by κ. In general, except for the few cases in panel (A), the proposed easy-to-implement successive testing procedure works well in the configurations considered; it correctly selects their right number with a probability over 0.8, and in most cases, the accuracy rates are even over 0.9, as shown in Table 4.

6 Empirical applications

In this section, we first investigate the possible non-stationarity in the panel of worldwide real exchange rates by employing the proposed approach, since the existence of non-stationarity implies that PPP may not provide a good long-run approximation for determining the dynamics of real exchange rates. In addition, for 37 advanced economies identified by IMF, we also analyze the panel of their consumer price indexes (CPIs) where the seasonal unit roots are expected to be observed more frequently. Two balanced panels of data were constructed from the IFS dataset of IMF: one is 94 real effective exchange rates (based on CPI) from 2005 Q1 to 2016 Q4, and the other is 37 CPIs of advanced economies from 2003 Q1 to 2017 Q4.[10]

The process of investigating the possible non-stationarity for these two panels of data is as follows. All the data are first standardized to have mean zero and unit variance; we then perform the augmented HEGY seasonal unit root tests for these standardized data. If most of the data in the given panel are non-stationary (at some specified seasonal frequency), we further disentangle the source of non-stationarity for them based on the proposed framework. More precisely, we estimate their corresponding factor model (1) by using the procedure proposed in Section 3.1, where the number of common factors is determined by using the penalty $\Gamma_4(N,T)$ in (3).[11] In consequence, the source of non-stationarity (at each specified seasonal frequency) can be disentangled by performing augmented HEGY seasonal unit root tests on the transformed factors and the estimated idiosyncratic errors induced from the factor model; the non-stationarity is pervasive if transformed common factors are the only source of non-stationarity; otherwise, it is both pervasive and economy-specific.

Throughout these two empirical studies, the optimal lag length in the augmented HEGY seasonal unit root tests is determined by Bayesian Information Criteria (BIC) with maximum lags equal to eight,[12] and the significant level is fixed at 5%.

6.1 Real exchange rates

Table 5: Detected $\kappa-$ type unit roots in real exchange rates by augmented HEGY tests.

Economies	$\kappa-$type unit roots	Economies	$\kappa-$type unit roots	Economies	$\kappa-$type unit roots
Algeria	0	Finland	0	Pakistan	0
Antigua and Barbuda	0	France	0	Papua New Guinea	0
Armenia		Gabon	0	Paraguay	0
Australia	0	Gambia	0	Philippines	0
Austria	0	Georgia	0	Poland	0
Bahamas	0	Germany	0	Portugal	0
Bahrain	0	Ghana	0	Romania	
Belgium	$0^{\dagger}$	Greece	0	Russian Federation	0
Belize	0	Grenada	0	Samoa	0
Bolivia	0	Guyana	$0^{\dagger}$	Saudi Arabia	0

Brazil	0		Hungary	$0^{\dagger}$		Sierra Leone	0	
Bulgaria		π	Iceland	0	π	Singapore	0	
Burundi	0		Iran	0		Slovak		
Cameroon	0		Ireland	0		Solomon Islands	0	
Canada	0		Israel	$0^{\dagger}$		South Africa	0	
Central African Republic	0		Italy	0		Spain	0	
Chile	0		Japan	0		St. Kitts and Nevis		
China	0		Latvia	0		St. Lucia	0	
Colombia	0		Lesotho	0		St. Vincent and the Grenadines	0	
Congo			Luxembourg	0		Sweden	0	
Costa Rica	0		Macedonia	0		Switzerland	0	
Cote d Ivoire	0		Malawi	0		Togo	$0^{\dagger}$	
Croatia	0		Malaysia	0		Trinidad and Tobago	0	
Cyprus	0		Malta	$0^{\dagger}$		Tunisia	0	
Czech Republic	0		Mexico	0		Uganda	0	
Denmark	0		Moldova	0		Ukraine	0	
Dominica	0		Morocco	0		UK	0	
Dominican Republic	0		Netherlands	$0^{\dagger}$		US	0	
Ecuador	0		New Zealand	0		Uruguay	0	
Equatorial Guinea	0		Nicaragua	0		Zambia		
Euro Area	0		Nigeria	0				
Fiji	0		Norway	0				
Transformed Factor	$\kappa-$type unit roots		Transformed Factor	$\kappa-$type unit roots		Transformed Factor	$\kappa-$type unit roots	
$\widehat{F}_1^*$	0		$\widehat{F}_2^*$					

1. The real effective exchange rates (based on consumer price index) of 94 economics are from IFS dataset of IMF. The sample period is from 2005 Q1 to 2016 Q4.
2. $\kappa = 0, \pi, \pi/2$, refers to the null that the series of interest has a conventional unit root (+1), semi-annual unit root (−1), annual unit roots ($\pm i$), respectively.
3. † indicates that the detected type of non-stationarity is solely due to the estimated common factor.

Table 5 first summarizes the results of the augmented HEGY test for the panel of real exchange rates and for the transformed common factors. First, for these 94 series of real exchange rates, only 6 of them are stationary (the rates of Armenia, the Congo, Romania, Slovak, St. Kitts and Nevis, and Zambia) while the others are non-stationary. Second, for those non-stationary exchange rates, 86 of them include conventional unit roots at zero frequency only, one of them (the rate of Bulgaria) has a pure seasonal unit root −1, and Iceland is the only country whose exchange rate involves both a conventional (+1) and a seasonal unit root (−1). It is obviously that the seasonal non-stationarity at frequency π could not be pervasive in the panel of these real exchange rates. In addition, the pooled tests (7) reject all the $\kappa-$type null hypotheses except the null that every series has a conventional unit root.

Given the panel of these 94 standardized real effective exchange rates, we then estimate their corresponding factor model (1) with two common factors. These two factors are further transformed by (5), $\widehat{F}_1^*$ and $\widehat{F}_2^*$ say, and their corresponding results of augmented HEGY tests are summarized in the last row of Table 5; the first common transformed factor has a unit root at zero frequency, while the second one is stationary. Besides, for those 87 real exchange rates with unit root +1, 80 of their estimated idiosyncratic errors induced from the factor model still have a conventional unit root; it means that the most observed non-stationarity of real exchange rates at zero frequency are due to both the common factor and the idiosyncratic errors; they are neither purely pervasive nor purely economy-specific. In sum, for the considered panel of exchange rates, all the above results indicate that there is no substantial evidence on the validity of PPP in the long run; many conventional unit roots and a few seasonal unit roots were detected in the data, and part of the violation of PPP without mean-reversion is pervasive.

6.2 Consumer price indexes

Table 6: Detected $\kappa-$ type unit roots in cpis by augmented HEGY tests.

Economies	$\kappa-$type unit roots			Economies	$\kappa-$type unit roots	Economies	$\kappa-$type unit roots	
Australia	0			Greece	0	Norway	0	
Austria	0			Iceland	0	Portugal	0	π
Belgium	0			Ireland	0	San Marino	0	π
Canada				Israel	0	Singapore	0	
China, P.R.:Hong Kong	0	π	π / 2	Italy	0	Slovak Republic	0	
China, P.R.:Macao	$0^{\dagger}$			Japan	0	Slovenia	0	π
Cyprus	0			Korea	0	Spain	0	
Czech Republic	0			Latvia	0	Sweden	0	
Denmark	0			Lithuania	0	Switzerland	0	π
Estonia	$0^{\dagger}$			Luxembourg	0	UK	0	
Finland	0			Malta	0	US	0	
France	0			Netherlands	0			
Germany	0			New Zealand	0			
Transformed Factor	$\kappa-$type unit roots			Transformed Factor	$\kappa-$type unit roots	Transformed Factor	$\kappa-$type unit roots	
$\widehat{F}_1^*$	0			$\widehat{F}_2^*$	0	$\widehat{F}_3^*$	0	
$\widehat{F}_4^*$								

1. The consumer price indexes (base year is 2010) of 37 advanced economics are from IFS dataset of IMF. The sample period is from 2003Q1 Q1 to 2017 Q4.
2. $\kappa = 0, \pi, \pi/2$, refers to the null that the series of interest has a conventional unit root (+1), semi-annual unit root (-1), annual unit roots $(\pm i)$, respectively.
3. † indicates that the detected type of non-stationarity is solely due to the estimated common factor.

For the panel of CPIs of the advanced economies, Table 6 presents the following results of the augmented HEGY Tests. First, for these 37 series of CPI, only CPI of Canada is stationary while all the others are not. Second, for those 36 non-stationary exchange rates, all of them have a conventional unit root at zero frequency, 5 of them have an additional seasonal unit root with frequency π, and Hong Kong is the only economy whose CPI involves unit roots $+1, -1$ and $\pm i$. Third, the ratio of detected seasonal non-stationary CPIs at frequency π to the panel is $5/37$, which is more than twice the significance level (5%) of the HEGY tests; we observe seasonal non-stationarity more frequently as expected. Fourth, four common factors of the factor model (1) are estimated for this panel of CPIs, and the non-stationarity of every transformed factor is investigated by augmented HEGY tests. The last two rows in Table 6 present the corresponding results; the first three transformed common factor has a unit root at zero frequency, while the last one is stationary. It immediately implies that the detected seasonal non-stationarity at frequency π in this panel of CPIs are economy-specific; they are not pervasive. Moreover, for those 36 CPIs with unit root +1, HEGY tests for their estimated idiosyncratic errors, which resulted from the factor model, show that 34 of those errors still have a conventional unit root; it means that the most observed non-stationarity of CPIs at zero frequency are neither purely pervasive nor purely economy-specific.

7 Conclusion

By extending the PANIC framework of Bai and Ng (2004) to a seasonal panel with factor structure, this paper proposes an estimation procedure to consistently estimate the number of common factors, the space spanned by factors, factor loadings, and idiosyncratic errors, without knowing a *priori* whether or not the series are seasonal stationary. By employing the common factor structure, certain types of seasonal non-stationarity of common factors and of the idiosyncratic errors can thus be tested separately. This feature is quite different from the existing literature on seasonal non-stationarity in panels; e.g. Otero, Smith, and Giulietti (2005, 2007, and 2008), Dreger and Reimers (2005) and Ucar and Guler (2010).

Within this proposed framework, we constructed the pooled tests for seasonal panels by using the p-values of the test statistic of individual tests associated with each cross-section unit, and provided a procedure involving successive tests to determine the number of independent stochastic trends (at each seasonal frequency) when there are multiple common trends. When there is one common factor, our simulations show that the direct (seasonal) panel unit root tests on the observed data tend to over-reject the null of (seasonal) unit root(s); when the panel data are driven by a common factor structure, the proposed works very well instead. When there are multiple factors, the finite-sample performance of the proposed successive testing procedure is good in most configurations considered here. Since there is no procedure yet in the literature to show how to determine the number of common stochastic trends with certain non-stationary types in a panel of seasonal data with possible factor structures, the proposed successive testing procedure can serve as a candidate for resolving this issue. For the empirical applications, we applied the proposed framework to investigate the possible non-stationarity in the panel of worldwide real exchange rates and in the panel of CPIs for 37 advanced economies. For the worldwide exchange rates, the seasonal non-stationarity is seldom observed, and one stationary common factor and one non-stationary factor (with unit root +1) are identified for the panel. Most of the non-stationary exchange rates with unit root +1 are further found to be driven by both the non-stationary factor and the non-stationary idiosyncratic errors. On the other hand, three non-stationary common factors with unit root +1 and one stationary factor are identified as the common driving forces of the CPIs of advanced economies; seasonal non-stationarity (at frequency π) is still not common enough to be pervasive even when it is detected more often than the exchange rates. Besides, most of the CPIs with unit root at zero frequency are observed to be neither purely pervasive nor purely economy-specific.

Acknowledgement

The Research Support from the National Science Council of the Republic of China (Taiwan), NSC 101-2410-H-004-030, is gratefully acknowledged.

Appendix: wald representation and rank condition for seasonal unit roots

Given the dynamics model (9) of factors, we are going to show why the rank of $\boldsymbol{B_F}(r_s)$ equals the number of factors with root r_s in this appendix. To illustrate, we first consider an univariate quarterly series w_t, which potentially has unit roots at zero and all seasonal frequencies; $(1-L^4)w_t$ is thus a stationary process, but may have a zero on the unit circle. Hylleberg et al. (1990) showed that its Wald representation is:

$$(1-L^4)w_t = B_w(L)\eta_t, \tag{13}$$

where $B_w(L)$ is the polynomial for the lag operator L and η_t is a white noise process with zero mean and finite variance σ_η^2. Let r_s $(= +1, -1, +i$ and $-i)$ denote a root of z for the polynomial $(1-z^4) = 0$; then for each s, we must have $B_w(r_s) \neq 0$, provided that w_t has the seasonal unit root r_s. To illustrate this point more clearly, let $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$,respectively, denote the real part and imaginary part of a complex number z; then in representation (13):

$$\begin{aligned} B_w(L) = {} & \Psi_1[1+L+L^2+L^3] + \Psi_2[1-L+L^2-L^3] + (\Psi_3+\Psi_4 L)[1-L^2] \\ & + B_w^{**}(L)(1-L^4), \end{aligned}$$

where $\Psi_1 = B_w(1)/4$, $\Psi_2 = B_w(-1)/4$, $\Psi_3 = \mathrm{Re}(B_w(i))/2$, $\Psi_4 = \mathrm{Im}(B_w(i))/2$ and $B_w^{**}(L)$ is the lag polynomial such that $B_w^{**}(1) \neq 0$, $B_w^{**}(-1) \neq 0$, $B_w^{**}(i) \neq 0$ and $B_w^{**}(-i) \neq 0$. Based on this representation, if $B_w(1) = 0$, it implies that $\Psi_1 = B_w(1)/4 = 0$, and w_t does not have this unit root since the representation (13) can be rearranged as $(1+L+L^2+L^3)w_t = [\Psi_2(1+L^2) + (\Psi_3+\Psi_4 L)(1+L) + B_w^{**}(L)(1+L+L^2+L^3)]\eta_t$ after canceling the factor $(1-L)$ in all terms of both sides of (13). Therefore, if w_t has the unit root +1, it necessitates that $B_w(1) \neq 0$. Generating this result to the multivariate case such as $\boldsymbol{F}_t$ in model (9), the rank of $\boldsymbol{B}_F(1)$ is merely the number of the factors with unit root +1. Similar arguments about the existence of other seasonal unit roots can apply, too. For more details, refer to Hylleberg et al. (1990).

Notes

1 As is well-known, seasonal non-stationarity may come from stochastic components (such as seasonal unit roots), deterministic components (such as seasonal dummies) and structural changes, etc. This paper only focuses on investigating the stochastic non-stationarity induced by seasonal unit roots in a seasonal panel.
2 In the literature, many studies have tried to search for stationarity among them via controlling the seasonality or the cross-section dependence in order to support PPP; see for e.g. Papell (1997), Wu and Wu (2001), and Ho (2008).
3 In this paper, we focus on the quarterly data to introduce the proposed framework, but generalizing the proposed to other seasonal data is straightforward. In general, if w_t is $SI(1)$, it means that w_t has a conventional unit root at the zero frequency, and $S-1$ seasonal unit roots at different seasonal frequencies.
4 As is well known in factor models, the loadings λ_i and the factors $\boldsymbol{f}_t^S$ (or $\boldsymbol{F}_t$) are not directly identifiable, since $\boldsymbol{\lambda}_i^* = \boldsymbol{H}^{-1}\boldsymbol{\lambda}_i$ and $\boldsymbol{f}_t^{S^*} = \boldsymbol{H}\boldsymbol{f}_t^S$ (or $\boldsymbol{F}_t^* = \boldsymbol{H}\boldsymbol{F}_t$) can induce the equivalent factor models when introducing an arbitrary invertible matrix $\boldsymbol{H}$. Therefore, we can only consistently estimate the space spanned by the factors and the space spanned by the loadings.
5 w_{it} is a generic variable that could be the factor in $\boldsymbol{F}_t$, ϵ_{it}, or their estimated counterparts.
6 The invalidity of the panel HEGY test for y_{it} will be shown by simulations in Section 5.
7 The independence assumption for v_{it} in A.4(iii) is stronger than the commonly required one as only the consistency of the estimators of the factor model is of interest. The weak cross-correlation in the errors is usually allowed in the framework of factor models; see for e.g. Chamberlian and Rothschild (1983) and Bai and Ng (2008).
8 Other than the results reported in Table 2, we also considered the cases with $N = 10, 20, 40$, and $T = 100, 200, 300, 400$. In general, the simulation results showed that, except few cases for Γ_4, the data with $N \geq 20$ actually gives us enough information about the factors (disentangling from the noise) and helps us to estimate them more precisely, all the resulting probabilities of correctly select the true number of common factors are greater than 0.99. To conserve space, we do not report those results here; they are available on request.
9 For cases considered in Table 4, the corresponding simulation results show that pooled HEGY tests on y_{it} generally over-reject the null hypotheses for all cases when the idiosyncratic error and/or some factor(s) are $SI(1)$. This is similar to the phenomenon observed in cases of single common factor. Instead, the pooled HEGY tests on $\hat{\epsilon}_{\text{it}}$ still perform well; most of the rejection rates are around the nominal size 5% (a few cases are little over-sized) and all the powers are 1 when $\rho = 0.5$ in all cases no matter what $q_1^{(SI)}$ is. To conserve space, we do not report the results for pooled HEGY tests for y_{it}, $\hat{\epsilon}_{\text{it}}$ here; they are available on request.
10 All the quarterly data are recorded as the average of their monthly observed values by IMF. For the real effective exchange rates, there are 96 series globally available in the IFS dataset; we dropped two (the rates of Netherlands Antilles and of Venezuela) of them to keep the panel balanced. For the CPIs, we consider all 37 advanced economies identified by IMF; the panel of CPIs starts from 2003 Q1 to keep the panel balanced. For more details, please refer to the web site: http://data.imf.org/?sk=4C514D48-B6BA-49ED-8AB9-52B0C1A0179B.
11 For these two cases of applications, when the maximum number of factors $\bar{q}$ is pre-specified from 4 to 8, the criterion $\Gamma_4(N, T)$ always yields the same results ($\hat{q} = 2$ for real exchange rates and 4 for CPIs). However, the other three criteria: $\Gamma_1(N, T)$, $\Gamma_2(N, T)$ and $\Gamma_3(N, T)$ are not robust to $\bar{q}$; all three criteria always select $\hat{q} = \bar{q}$ for different specified values of $\bar{q}$.
12 For the series of interest, the augmented HEGY test is performed by first estimating the largest linear regression model with intercept, trend and seasonal dummies as deterministic regressors; the trend or/and seasonal dummies in the model are then dropped to yield a degenerated model if the corresponding estimated coefficients are insignificant. Accordingly, the results of the tests are based on the estimates of the considered model. The package used is the HEGY add-ins in Eviews written by Nicolas Ronderos in 2015.

References

Bai, J., and S. Ng. 2002. "Determining the Number of Factors in Approximate Factor Model." *Econometrica* 70: 191–221.
Bai, J., and S. Ng. 2004. "A Panic Attack on Unit Root and Cointegration." *Econometrica* 72: 1127–1177.
Bai, J., and S. Ng. 2008. "Large Dimensional Factor Analysis." *Foundations and Trends® in Econometrics* 3: 89–163.
Breitung, J., and S. Das. 2005. "Panel Unit Root Tests Under Cross Sectional Dependence." *Statistica Neerlandica* 55: 414–433.
Breitung, J., and M. H. Pesaran. 2008. "Unit Roots and Cointegration in Panels." In *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, edited by L. Matyas and P. Sevestre, 279–322. Berlin: Springer. Doi: 10.1007/978-3-540-75892-1_9.
Chamberlian, G., and M. Rothschild. 1983. "Arbitrage, Factor Structure and Mean-variance Analysis in Large Asset Markets." *Econometrica* 51: 1305–1324.
Chang, Y. 2004. "Bootstrap Unit Root Tests in Panels with Cross-Sectional Dependence." *Journal of Econometrics* 120: 263–293.
Choi, I. 2001. "Unit Root Tests for Panel Data." *Journal of International Money and Finance* 20: 249–272.
Dickey, D. A., and W. A. Fuller. 1979. "Distribution of the Estimators for Autoregressive Time Series with a Unit Root." *Journal of the American Statistical Association* 74: 427–431.
Dickey, D. A., and W. A. Fuller. 1981. "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root." *Econometrica* 49: 1057–1072.
Dickey, D. A., D. P. Hasza, and W. A. Fuller. 1984. "Testing for Unit Roots in Seasonal Time Series." *Journal of the American Statistical Association* 79: 355–367.
Dreger, C., and H. E. Reimers. 2005. "Panel Seasonal Unit Root Test: Further Simulation Results and An Application to Unemployment Data." *AStA Advances in Statistical Analysis* 89: 321–337.
Fisher, R. A. 1932. *Statistical Methods for Research Workers*. 4th ed. Edinburgh:Oliver & Boyd.
Ghysels, E., and D. R. Osborn. 2001. *The Econometric Analysis of Seasonal Time Series*. Cambridge:Cambridge University Press. Doi: 10.1017/CBO9781139164009.
Harris, D. 1997. "Principal Components Analysis of Cointegrated Time Series." *Econometric Theory* 13: 529–557.
Hlouskova, J., and M. Wagner. 2006. "The Performance of Panel Unit Root and Stationarity Tests: Results from a Large Scale Simulation Study." *Econometric Reviews* 25: 85–116.
Ho, T.-W. 2008. "Testing Seasonal Mean-reversion in the Real Exchange Rates: an Application of Nonlinear IV Estimator." *Economics Letters* 99: 314–316.

Hylleberg, S., R. F. Engle, C. W. J. Granger, and B. S. Yoo. 1990. "Seasonal Integration and Cointegration." *Journal of Econometrics* 44: 215–238.

Im, K., M. H. Pesaran, and Y. Shin. 2003. "Testing for Unit Roots in Heterogeneous Panels." *Journal of Econometrics* 115: 53–74.

Johansen, S. 1995. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press. Doi: 10.1093/0198774508.001.00.

Kunst, R. M. 2009. "A Nonparametric Test for Seasonal Unit Roots." *Economics Series*, 233, Institute for Advanced Studies. Doi: 10.1111/j.1468-0084.2010.00627.x.

Kunst, R. M., and P. H. Franses. 2011. "Testing for Seasonal Unit Roots in Monthly Panels of Time Series." *Oxford Bulletin of Economics and Statistics* 73: 469–488.

Lee, Y., and D. Shin. 2006. "Unit Root Tests for Cross-Sectionally Dependent Seasonal Panels." *Economics Letters* 93: 311–317.

Levin, A., and C. F. Lin. 1993. *Unit Root Tests in Panel Data: Asymptotic and Finite-sample Properties.* Unpublished manuscript. Doi: 10.1016/S0304-4076(01)00098-7.

Levin, A., C. Lin, and C. J. Chu. 2002. "Unit Root Tests in Panel Data: Asymptotic and Finite-sample Properties." *Journal of Econometrics* 108: 1–24.

Maddala, G. S., and S. Wu. 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, R., and B. Perron. 2004. "Testing for Unit Root in Panels with Dynamic Factors." *Journal of Econometrics* 122: 81–126.

O'Connell, P. 1998. "The Overvaluation of Purchasing Power Parity." *Journal of International Economics* 44: 1–19.

Osborn, D. R., A. P. L. Chui, J. P. Smith, and C. R. Birchenhall. 1988. "Seasonality and the Order of Integration for Consumption." *Oxford Bulletin of Economics and Statistics* 50: 361–377.

Otero, J., J. Smith, and M. Giulietti. 2005. "Testing for Seasonal Unit Roots in Heterogeneous Panels." *Economics Letters* 86: 229–235.

Otero, J., J. Smith, and M. Giulietti. 2007. "Testing for Seasonal Unit Roots in Heterogeneous Panels in the Presence of Cross Section Dependence." *Economics Letters* 97: 179–184.

Otero, J., J. Smith, and M. Giulietti. 2008. *Testing for Seasonal Unit Roots in Heterogeneous Panels Using Monthly Data in the Presence of Cross Sectional Dependence*. Unpublished manuscript.

Papell, D. 1997. "Searching for Stationarity: Purchasing Power Parity Under the Current Float." *Journal of International Economics* 43: 313–323.

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 P. Perron. 1988. "Testing for Unit Root in Time Series Regression." *Biometrica* 75: 335–346.

Shin, D., and M. Oh. 2009. "Tests for Seasonal Unit Roots in Panels of Cross-Sectionally Correlated Time Series." *Statistics* 43: 139–152.

Stock, J. H., and M. W. Watson. 1988. "Testing for Common Trends." *Journal of the American Statistical Association* 83: 1097–1107.

Stock, J. H., and M. W. Watson. 2002a. "Forecasting Using Principal Components from a Large Number of Predictors." *Journal of the American Statistical Association* 97: 1167–1179.

Stock, J. H., and M. W. Watson. 2002b. "Macroeconomic Forecasting Using Diffusion Indexes." *Journal of Business and Economic Statistics* 20: 147–162.

Ucar, N., and H. Guler. 2010. "Testing Stochastic Convergence Hypothesis in Seasonal Heterogeneous Panels." *Economic Modelling* 27: 422–431.

Wells, J. 1997. "Business Cycles, Seasonal Cycles, and Common Trends." *Journal of Macroeconomics* 19: 443–469.

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

Supplementary Material: The online version of this article offers supplementary material (DOI: https://doi.org/10.1515/snde-2018-0075).