ARTICLE IN PRESS

Computational Statistics and Data Analysis ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Numerical distribution functions for seasonal unit root tests☆

Ignacio Diaz-Emparanza*

University of the Basque Country UPV/EHU. Avenida Lehendakari Aguirre, 83. E48015 Bilbao, Spain

HIGHLIGHTS

- A generalisation of HEGY seasonal unit root tests was presented in Smith, Taylor and del Barrio Castro (2009).
- We use a response surface regressions approach to calculate *P*-values for the HEGY statistics.
- They can be used for any seasonal periodicity, sample size and autoregressive order.
- A Gretl function package is provided for applying the tests and calculating their *P*-values.

ARTICLE INFO

Article history:
Received 9 January 2012
Received in revised form 7 March 2013
Accepted 7 March 2013
Available online xxxx

Keywords:
Seasonality
Unit roots
Response surface analysis
HEGY tests

ABSTRACT

It is often necessary to test for the presence of seasonal unit roots when working with time series data observed at intervals of less than a year. One of the most widely used methods for doing this is based on regressing the seasonal difference of the series over the transformations of the series by applying specific filters for each seasonal frequency. This provides test statistics with non-standard distributions. A generalisation of this method for any periodicity is presented and a response surface regressions approach is used to calculate the *P*-values of the statistics whatever the periodicity and sample size of the data. The algorithms are prepared with the Gretl open source econometrics package and two empirical examples are presented.

1. Introduction

Unit roots may cause severe problems in a regression model if they are not properly dealt with: this may imply inconsistent coefficient estimators and non-standard distributions for significance tests and for forecast intervals. There have been many papers on testing for unit roots since the book by Fuller (1976), which introduced the test currently known as the *Augmented Dickey–Fuller test*, ADF (see also Dickey and Fuller, 1981). Apart from the ADF test, other noteworthy tests are Phillips and Perron (1988), the KPSS test for stationarity by Kwiatkowski et al. (1992) and the ADF–GLS test by Elliott et al. (1996), all of which have become widely used by empirical economists. However, when working with time series data observed at intervals shorter than a year, the presence of unit roots should be tested for, not only in the long run but also in seasonal cycles. Over the last thirty years various methods have been proposed for testing for seasonal unit roots. For example, Hasza and Fuller (1982) and Dickey et al. (1984) propose joint tests for all seasonal unit roots, and in later papers Osborn et al. (1988) and in particular Hylleberg et al. (1990) (hereinafter HEGY) propose tests that would deal with each seasonal and zero frequency root to be considered separately. There are also interesting tests of seasonal stability by Canova and Hansen (1995), which also consider each frequency individually. The HEGY tests are not very difficult to implement and have therefore become widely used among empirical economists.

☆ Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.csda.2013.03.006.

* Tel.: +34 946013732; fax: +34 946013754.
E-mail address: ignacio.diaz-emparanza@ehu.es.

0167-9473/$ – see front matter
http://dx.doi.org/10.1016/j.csda.2013.03.006

One of the problems with most of the unit root tests mentioned above is that their statistics have non-standard distributions, so in practice the values in the tables published need to be interpolated to compare them with the values calculated or simulate the empirical distributions for exactly the same model and the same sample size that is being used. MacKinnon (1994) uses simulation methods and response surface regressions to estimate the asymptotic distributions of a large number of unit roots and cointegration tests at zero frequency (long run). MacKinnon (1996) then extends these results, providing a way of approximating small sample distributions too.

Harvey and van Dijk (2006) apply MacKinnon's method, using response surface regressions to provide a simple way of obtaining critical values and P values for any sample size and any order of lags of the endogenous variable in the regression for the HEGY tests mentioned above. As in the original HEGY article, all this is for quarterly data.

In the 21st century it seems quite anachronistic to have to use statistical tables. Empirical economists normally use computers for calculations, so ideally they should have a computer algorithm to calculate P-values instead of having to look at statistical tables. The main objective of this paper is to obtain a generalisation of the method by Harvey and van Dijk for calculating the P values of the HEGY statistics whatever the seasonal periodicity, S, and sample size T of the data. Seasonal periodicity S is defined as the number of values of the series observed in each year (it is sometimes convenient to change the reference period from one year, for example, to one day if data are hourly, $S = 24$, or one week if data are daily, $S = 7$, but for ease of exposition I will continue to refer to the reference period as the 'year').

For this approach to be practical it needs to be possible to implement it in a computer algorithm. As a complement to this paper an algorithm prepared in Gretl is provided (see http://gretl.sourceforge.net). Gretl is a cross-platform software package for econometric analysis. It is *open source*, free software: anyone may redistribute it and/or modify it under the terms of the GNU General Public License (GPL). This makes Gretl a very good econometric package in terms of the replicability of its calculations and results (Baiocchi, 2007).

2. Seasonal unit roots

The study of seasonality requires the use of some concepts of *spectral* or *frequency-domain analysis*. The fundamental goal of such analysis is to determine how important cycles of different frequencies are in accounting for the behaviour of the series [For a thorough introduction to frequency-domain analysis see, for example, Chapter 6 of Harvey (1993) or the equivalent chapter in Hamilton (1994)].

The period, $\mathbb{T}$, is defined as the time taken to complete one cycle. The angular frequency, $\omega = 2\pi/\mathbb{T}$, measures the frequency of the cycle in radians per unit of time. For a series with T observations, if T is even a total of $T/2$ complete cycles may be observed, with their periods being T/j with j being an integer and $j = 1, 2, \ldots, T/2$. If T is odd, $T/2$ is not an integer, and only $(T-1)/2$ complete cycles may be observed, with their periods being T/j with $j = 1, 2, \ldots, (T-1)/2$. Those of these cycles that can be observed within a year are called *seasonal cycles*. Series whose values are observed S times a year may, if S is even, show up to $S/2$ complete seasonal cycles every year, with periods S/j with j being an integer and $j = 1, 2, \ldots, S/2$. If S is odd, only $(S-1)/2$ cycles may be observed within a year and their periods are S/j with $j = 1, 2, \ldots, (S-1)/2$. For example if $S = 4$ the seasonal cycles have periods 4 and 2, which correspond respectively to cycles observed once and twice a year; if $S = 5$ two seasonal cycles are also observed, one with period 5 and the other with period 2.5. In summary, a time series in which observations are regularly collected S times a year can contain $\lfloor S/2 \rfloor$ different seasonal cycles, with $\lfloor\cdot\rfloor$ denoting the integer part of the number contained in brackets, i.e.

$$\lfloor S/2 \rfloor = \begin{cases} S/2 & \text{if } S \text{ is even} \\ (S-1)/2 & \text{if } S \text{ is odd} \end{cases}$$

and the angular frequencies corresponding to the seasonal cycles are $\omega_j = \frac{2\pi j}{S}$, with $j = 1, \ldots, \lfloor S/2 \rfloor$.

If a series is generated by an autoregressive (AR) process such as

$$\phi(L)x_t = \varepsilon_t \tag{1}$$

where L is the lag operator, such that $Lx_t = x_{t-1}$, $\phi(L)$ represents the polynomial $1 - \phi_1 L - \phi_2 L^2 - \cdots$ and ε_t is a white noise process with variance σ_ε^2, the cycles in x_t are associated with the roots of the polynomial equation $\phi(z) = 0$, where z is a number in the complex plane, i.e. $z = a + b \cdot i$ with $a, b \in \mathbb{R}$ and $i = \sqrt{-1}$. Each root implies a cycle. The real and imaginary parts of the roots determine the period and the frequency of the cycle. The angular frequency of a cycle is related to the roots of the AR polynomial by $\omega = \arctan(b/a)$. If $z > 0$ and real ($b = 0$) the root is on the horizontal right axis of the complex plane, $\omega = 0$, and the period is infinite. This root corresponds to the long run (or trend) of the series. If $z < 0$ and real, the root is on the left side of the horizontal axis, $\omega = \pi$ (called the *Nyquist frequency*) and the period is 2. All the complex solutions (with $b \neq 0$) of $\phi(z) = 0$ have conjugate solutions associated with them. In this case, the problem of *aliasing* arises (See Hamilton, 1994, pp. 160–161, for a detailed description of this problem). Basically the problem is this: based on a sample of T observations it is impossible to distinguish between the cycle associated with the root $z = a + b \cdot i$, with frequency ω, and the cycle associated with its complex conjugate root $z = a - b \cdot i$, with frequency $-\omega$. The only possible way of working with them is therefore to treat them jointly.

Hylleberg et al. (1990) study how to test for unit roots in seasonal time series. They take quarterly periodicity of data ($S = 4$) as their reference and assume that the series x_t is generated by a possibly infinite order autoregressive process,

Please cite this article in press as: Diaz-Emparanza, I., Numerical distribution functions for seasonal unit root tests. Computational Statistics and Data Analysis (2013), http://dx.doi.org/10.1016/j.csda.2013.03.006