

Testing for Unit Roots in Seasonal Time Series with R: The `uroot` Package

Ignacio Díaz-Emparanza * Javier López-de-Lacalle †
Universidad del País Vasco Universidad del País Vasco
Ignacio.Diaz-Emparanza@ehu.es javlacalle@yahoo.es

19th June 2006

Abstract

Traditional econometric methods require the data to be stationary. Since most of the macroeconomic time series are non-stationary, several tests have been developed to find out the most suitable way of transforming data into stationary. In a univariate framework, the `uroot` R-package implements tests for the null hypotheses of unit root and stationarity in seasonal time series. This document focuses on macroeconomic seasonal time series, nevertheless, the analysis can be extended to other areas of research.

A guidance on the use of the `uroot` R-package is given. The package performs graphical and statistical analysis and provides a graphical user interface. Some of the prominent test statistics are introduced and a protocol for detecting unit roots is described. The use of the package is illustrated with an example.

Keywords: Time series, unit root, seasonality, R.

*The first author acknowledges financial support from the Universidad del País Vasco to the research group 9/UPV-00038.321/2001 and to the Ministerio de Ciencia y Tecnología and FEDER to the project BEC2003-02028.

†The second author acknowledges predoctoral fellowship provided by the Universidad del País Vasco.

Contents

1	Introduction	3
I	Theoretical Review	4
2	Stationary and Integrated Processes	4
3	Long Run Unit Root Tests	6
3.1	The ADF test	6
3.2	The KPSS test	8
4	Seasonal Unit Root Tests	9
4.1	The HEGY test	9
4.2	The CH test	10
5	Protocol	11
II	Applied Analysis	14
6	The <code>uroot</code> Package	15
6.1	Installing and running <code>uroot</code>	15
6.2	Loading data	16
7	An Example	16
7.1	Graphics	17
7.1.1	Original series preview	17
7.1.2	Seasonal operator	19
7.1.3	Seasonal paths	20
7.2	Unit root tests	20
7.2.1	Deterministic components	21
7.2.2	Long run unit root tests	23
7.2.3	Seasonal unit root tests	25

1 Introduction

Traditional econometric methods require the data to be stationary.¹ Most of the macroeconomics time series, instead, display a trend and heteroskedasticity,² failing to fulfil stationarity conditions. As a consequence, time series must be modelled taking into account non-stationary features detected in the data. Equations below show some processes in which, due to a trend in the data, the mean is not constant. In spite of having a trend, its nature is different in each case.

$$y_t = y_{t-1} + \epsilon_t, \quad (1)$$

$$y_t = \alpha + \beta t + \epsilon_t, \quad (2)$$

where ϵ_t is *white noise*, $iid(0, 1)$. Equation (1) is a *random walk*. It may be easily showed that this process is generated by an accumulation or integration of shocks in ϵ_t which shapes a stochastic trend. Process in Equation (2) is guided by an intercept, α , and a deterministic trend, in particular, a linear trend with a slope equals to β .

With regard to seasonal pattern let us see the following processes:

$$y_t = y_{t-S} + \epsilon_t, \quad (3)$$

$$y_t = \sum_{i=1}^S \gamma_i D_{i,t} + \epsilon_t, \quad (4)$$

where ϵ_t is *white noise*, $iid(0, 1)$, S is the periodicity of the time series, and $D_{i,t}$ are seasonal dummies. Data generating process in Equation (3) is guided by seasonal accumulation or integration of shocks in such a way that, each season, as well as the long run, evolves according to a random walk. The process described in (4), instead, shapes a deterministic seasonal pattern, since γ_i are constant for all the period.

Note that when non-stationary features arise from a deterministic source, Equations (2) and (4), the data become stationary once the linear trend or seasonal dummies are estimated and removed from the original data. When the trend or seasonal component is stochastic, Equations (1) and (3), to take first or seasonal differences transforms the data into stationary.

Durlauf and Phillips (1988) warn about the dangers of testing a linear trend significance in the presence of a unit root. They prove that if the data

¹ARMA modelling, for instance, is based on Wold's Theorem, which applies to stationary data.

²This term refers to a non-constant variance. It is a common phenomenon in macroeconomic time series, suggesting that trend, seasonal, and irregular components add them up in a multiplicative way.

contain a unit root such as in Equation (1), the distribution of the t -statistic for the trend, t_β , is not the standard distribution since it would overestimate the significance of this component. The same applies to seasonal dummies.

There is a trade-off between the following two matters. On the one hand, if a unit root is omitted, estimates are inconsistent. On the other hand, to assume more unit roots than those that are actually in the data will increase the variance of the estimator. The first mistake invalidates inference outcomes, whereas the second increases the error prediction variance, making forecast less accurate.

In the face of this fact, most of the applied researches take a conservative position, limiting the analysis of stationarity to the paradigm of the $\Delta\Delta^S \log(y_t)$ filter. In practice, however, it may happen that only some of all the roots in that filter are equal to unity. In that case the regular and seasonal differences filter do not fit the data. Graphics and tests implemented in `uroot` allow the researcher to decide which roots in the filter $\Delta\Delta^S$ should be regarded equal to unity.

The remaining of this document consists of two differentiated parts. The first one reviews theoretical issues concerning the functionalities implemented in the `uroot` package. Section 2 describes some concepts related to unit root tests. Section 3 reviews ADF and KPSS tests, whereas section 4 is devoted to HEGY and CH seasonal unit root tests. Section 5 introduces a protocol for detecting unit roots. The second part provides a guidance in the use of the package. Section 6 explains how to run the `uroot` package. Section 7 shows an application in which the main facilities of the package are illustrated.

Part I

Theoretical Review

2 Stationary and Integrated Processes

Traditional statistical and econometric methods assume that the data arise from a *stationary* process. a stochastic process is *stationary* if all of its random variables are identically distributed. This condition implies that all of the statistical moments of the variable are identically distributed.

In practice, these conditions are too demanding, and only first and second moments are usually required to be identical. In this case, the condition is called *second order* or *weak stationarity* and entails that mean, variance, covariances, and correlations do not depend on the sample period in which

they are quantified, but they are constant.

Table 1 summarizes statistical properties of the process (5) when it is stationary, $I(0)$, that is, $|\phi| < 1$, against the case in which it is integrated, $I(1)$, for $\phi = 1$.

$$y_t = \phi y_{t-1} + \epsilon_t. \tag{5}$$

Table 1: Stationary process, $I(0)$, versus integrated process, $I(1)$

	Stationary ARMA	Integrated process
Variance	Finite	Infinite
Autocorrelation function	$\rho_k = \phi^k$	$\rho_k = \sqrt{\frac{t-k}{t}}$
Returning to mean estimated period	Finite	Infinite
Memory	Temporal	Permanent

Since most of the macroeconomic time series evolve along a trend and display heteroskedasticity, they are generally not stationary even in the weak sense. Nevertheless, these facts don't detract from the traditional econometric statements. Graphics and tests described in this article help to determine non-stationarity features in the data and to modelize them properly.

Since the article of Nelson and Plosser (1982) several tests have been developed to establish whether the long run of a time series is better described by a deterministic or a stochastic component. To this effect, Augmented Dickey-Fuller test [ADF], (Dickey and Fuller, 1981), as well as Kwiatkowski et al. (1992) test [KPSS] are implemented in `uroot`.

With regard to seasonal cycles, they have often been understood as a *noise* that hides outstanding information for economic analysis. This interpretation arose from the difficulty to explain fluctuations that depart from the trend-cycle component. Researches argued deterministic events to explain the short run fluctuations: holidays, meteorology, and other events that occurs regularly in the economy were understood as the main source of seasonal patterns.

Although it seems at first plausible, facts don't add up. Several articles, (Dickey et al., 1984), Hylleberg et al. (1990), and Canova and Hansen (1995), among others), have revealed that seasonal cycles are often bigger and less regular than if they are a consequence of phenomenons like those mentioned above.

Causes and economic content of seasonal cycles have been also discussed. Nowadays, seasonality in economic time series is defined as Hylleberg, 1992, pg. 4 says:

Seasonality is the systematic, although not necessarily regular, intra-year movement caused by the changes of the weather, the calendar, and timing of decisions, directly or indirectly through the production and consumption decisions made by agents of the economy. These decisions are influenced by endowments, the expectations and preferences of the agents, and the production techniques available in the economy.

For a better understanding of some basic concepts in this article we will decompose the seasonal operator into partial polynomials. Let L be the lag operator and S the periodicity of the data, then $(1 - L^S)$ is the seasonal operator. In the case of quarterly data, $S = 4$, and monthly, $S = 12$, it yields:

$$\begin{aligned} (1 - L^4) &= (1 - L)(1 + L)(1 + L^2) \\ (1 - L^{12}) &= (1 - L)(1 + L)(1 + L^2)(1 + L + L^2)(1 - L + L^2) \\ &\quad (1 + \sqrt{3}L + L^2)(1 - \sqrt{3}L + L^2) \end{aligned}$$

Table 2 summarizes long run and seasonal cycles in quarterly and monthly series, as well as the roots of each polynomial. When the module of the roots in these polynomials are equal to unity, the corresponding cycle is shaped as an accumulation of shocks. When the root lies outside the unit circle, shocks have transient effect, instead.

Sections 3 and 4 introduce some statistics to test whether the module of long run and seasonal roots are equal to unity as it is the case in the polynomials reported in Table 2.

3 Long Run Unit Root Tests

3.1 The ADF test

On the basis of an AR(1) process, Equation (6), Fuller (1976) and Dickey and Fuller (1981) obtain the critical values to test if a unit root exists as well as the significance of deterministic components.

$$y_t = \rho y_{t-1} + \epsilon_t, \tag{6}$$

Table 2: Long run and seasonal frequencies

Frequency	Period	Cycles/ year	Root	Filter	
Monthly series					
0	Long run	∞	0	1	$(1 - L)$
$\frac{\pi}{6}, \frac{11\pi}{6}$	Annual	12; 1.09	1; 11	$\frac{1}{2}(\sqrt{3} \pm i)$	$(1 - \sqrt{3}L + L^2)$
$\frac{\pi}{3}, \frac{5\pi}{3}$	Semiannual	6; 1.2	2; 10	$\frac{1}{2}(1 \pm \sqrt{3}i)$	$(1 - L + L^2)$
$\frac{\pi}{2}, \frac{3\pi}{2}$	Quarterly	4; $\frac{4}{3}$	3; 9	$\pm i$	$(1 + L^2)$
$\frac{2\pi}{3}, \frac{4\pi}{3}$		3; 1.5	4; 8	$-\frac{1}{2}(1 \pm \sqrt{3}i)$	$(1 + L + L^2)$
$\frac{5\pi}{6}, \frac{7\pi}{6}$	Bimonthly	2.4; 1.7	5; 7	$-\frac{1}{2}(\sqrt{3} \pm i)$	$(1 + \sqrt{3}L + L^2)$
π		2	6	-1	$(1 + L)$
Quarterly series					
0	Long run	∞	0	1	$(1 - L)$
$\frac{\pi}{2}, \frac{3\pi}{2}$	Annual	4; $\frac{4}{3}$	1; 3	$\pm i$	$(1 + L^2)$
π	Semiannual	2	2	-1	$(1 + L)$

where the null hypothesis is $\rho = 1$. It is convenient rewrite the Equation (6) as follows:

$$\Delta y_t = \phi y_{t-1} + \epsilon_t. \quad (7)$$

The null hypothesis for a unit root is $\phi = 0$, and the corresponding t -statistic is as usual, $\tau_\phi = \frac{\hat{\phi}}{\hat{\sigma}_\phi}$. The alternative hypothesis is: $-2 < \phi < 0$. As $\phi > 0$ implies $\rho > 1$, that is, an explosive behaviour, it may be detected just in the graphics of the original data. Anyway, such an explosive behaviour is not usual in macroeconomic time series.

The distribution of the t -statistic depends on the deterministic components included in the auxiliary regression. In particular, Dickey and Fuller (1981) consider the following versions:

$$\begin{aligned} \Delta y_t &= \phi y_{t-1} + \epsilon_t, \\ \Delta y_t &= \alpha + \phi y_{t-1} + \epsilon_t, \\ \Delta y_t &= \alpha + \beta t + \phi y_{t-1} + \epsilon_t, \end{aligned}$$

Finally, it must be taken into account that the data generating process simulated to obtain the critical values is an AR(1) process where the residuals are white noise. Since including lags of the dependent variable does not affect to the t -statistic distribution, they should be added to the auxiliary regression. Several methods for selecting the order of lags that provides white noise

residuals are available in `uroot`. These methods are based on criteria such as: AIC, BIC, significant lags, and Ljung-Box statistics.

3.2 The KPSS test

Kwiatkowski et al. (1992) develop a statistic in which the null hypothesis is a stationary process, whereas the alternative is an integrated process. The statistic focuses on the variance of the residuals in the following two frameworks:

$$\begin{aligned} y_t &= \psi t + r_t + \epsilon_t, \\ r_t &= r_{t-1} + u_t, \quad u_t \sim iid(0, \sigma_u^2), \end{aligned}$$

where the initial value, r_0 is fixed, playing the role of an intercept. The time series is stationary when $\sigma_u^2 = 0$, otherwise the long run follows a random walk around a deterministic trend of slope ψ .

The second model is specified as follows:

$$\begin{aligned} y_t &= r_t + \epsilon_t \\ r_t &= r_{t-1} + u_t, \quad u_t \sim iid(0, \sigma_u^2), \end{aligned}$$

where under the null hypothesis is $\sigma_u^2 = 0$, hence, the series is guided by a level stationary process without a deterministic trend, that is, $\psi = 0$.

Depending on the null hypothesis desired to test the residuals of one of the following regressions are needed to obtain the statistic:

$$y_t = \alpha + \beta t + e_t \Rightarrow e_t = y_t - \hat{\alpha} - \hat{\beta} t, \quad (8)$$

for the null hypothesis of trend stationarity,

$$y_t = \alpha + e_t \Rightarrow e_t = y_t - \frac{1}{T} \sum_{t=1}^T y_t, \quad (9)$$

for the null hypothesis of level stationarity.

In both cases, the statistic is calculated as:

$$\mu = n^{-2} \sum_{t=1}^n \frac{S_t^2}{s^2(l)}, \quad (10)$$

where $S_t = \sum_{i=1}^t e_i$, $t = 1, 2, \dots, n$. Finally, $s^2(l)$, defined in Equation (11), is a consistent estimator of the residuals variance that fulfils regularity conditions described in Phillips, 1987, assumption 2.1. In this way, residuals are

allowed to belong to a less restricted set than $id \sim N(0, \sigma_\epsilon^2)$. In particular some degree of correlation is allowed.

$$s^2(l) = n^{-1} \sum_{t=1}^n e_t^2 + 2T^{-1} \sum_{s=1}^l w(s, l) \sum_{t=s+1}^T e_t e_{t-s}. \quad (11)$$

The weighting function $w(s, l)$ is the one proposed by Newey and West (1987) and is defined as $w(s, l) = 1 - \frac{s}{l+1}$. The lag truncation parameter, l , must tend to infinite as the sample size goes to infinite for the estimated covariances matrix to be consistent.

4 Seasonal Unit Root Tests

4.1 The HEGY test

According to the ADF test, Dickey et al. (1984) obtained the critical values of ϕ under the following data generating process.

$$(1 - L^S)y_t = \phi y_{t-S} + \epsilon_t, \quad \epsilon_t \sim iid(0, \sigma_\epsilon^2), \quad (12)$$

where S is peridocity of the data: 4 in the case of quarterly series and 12 for monthly series. Under the null hypothesis ϕ is equal to 1 and the process contains all the roots in Table (2), page 7. In practice, however, it may be some, but not all, seasonal unit roots. Hence, it would be convenient to specify a model in which the regressors allow to test for individual roots. That is what Hylleberg et al. (1990) achieve in the case of quarterly series.

The seasonal operator can be decomposed into the following polynomials: $(1 - L^4) = (1 - L)(1 + L)(1 - iL)(1 + iL)$, where the roots are $(\pm 1, \pm i)$. Assuming y_t is generated by an AR(p) process, Hylleberg et al. (1990) show that y_t can be represented as

$$\varphi(L) \Delta^4 y_t = \pi_1 y_{1,t-1} + \pi_2 y_{2,t-1} + \pi_3 y_{3,t-2} + \pi_4 y_{3,t-1} + \epsilon_t, \quad (13)$$

where an intercept, linear trend, and/or seasonal dummies may be included. The $\varphi(L)$ polynomial consist of those lags of the dependent variable needed to ensure that the residuals are white noise. They may be chosen according with criteria such as: AIC, BIC, significant lags, and Ljung-Box. The regressors are defined as follows:

Each one of the regressors, $y_{i,t}$, retains a single unit root filtering the remaining roots. They are defined as follows:

- $y_{1,t} = (1 + L)(1 + L^2) y_t = (1 + L + L^2 + L^3) y_t,$

- $y_{2,t} = -(1 - L)(1 + L^2) y_t = -(1 - L + L^2 - L^3) y_t$,
- $y_{3,t} = -(1 - L)(1 + L) y_t = -(1 - L^2) y_t$.

A test for the null hypothesis of the unit root 1 is carried out with the t -statistics for $\pi_1 = 0$, the hypothesis of the unit root -1 entails the test $\pi_2 = 0$, whereas the same hypothesis for the unit roots $\pm i$ may be tested with a t -statistic for $\pi_3 = 0$ and $\pi_4 = 0$, or jointly with an F -type test. Hylleberg et al. (1990) collect the suitable critical values for each case.

Burridge and Taylor (2001) show that if the residuals are correlated, after including lagged seasonal differences the t -statistics for the unit roots 1 and -1 , as well as the F -type tests for the harmonic frequencies keep the same limiting null distributions, but the t -statistics for the harmonic frequencies do not, instead. Hence, we will consider in this article the F -type tests for those frequencies.

Beaulieu and Miron (1993), Franses (1990), and Franses and Hobijn (1997) get the critical values of the corresponding statistics for other periodicities of the data, monthly time series among others.

4.2 The CH test

In the same way as the KPSS test reverse the null and the alternative hypotheses of the ADF test, Canova and Hansen (1995) reverse HEGY's hypothesis. As in the KPSS test, the CH framework is an state space representation of an unobserved components model:

$$y_i = \mu + x_i' \beta + f_i' \gamma_i + e_i, \quad (14)$$

$$\text{con } A' \gamma_i = A' \gamma_{i-1} + u_i, \quad i = 1, 2, \dots, n \quad (15)$$

with γ_0 fixed, $u_i \sim iid$, $e_i \sim (0, \sigma_e^2)$, and x_i are lags of the original series. Matrix f_i and γ_i are built as follows:

$$\gamma_i = \begin{bmatrix} \gamma_{1i} \\ \cdot \\ \cdot \\ \gamma_{qi} \end{bmatrix}, \quad f_i = \begin{bmatrix} f_{1i} \\ \cdot \\ \cdot \\ f_{qi} \end{bmatrix} \quad (16)$$

where $q = S/2$, and S is the periodicity of the series. If $j < q$, $f_{ji}' = (\cos(j/q)\pi i, \sin(j/q)\pi i)$ and if $j = q$, $f_{qi}' = (\cos(\pi i), \sin(\pi i)) = (-1)^i, 0$. Matrix $A_{(s-1) \times a}$ selects a elements from γ_i to be tested.

Since no lags of the dependent variable are included, to avoid that residual correlation distorts the distribution of the statistics a consistent variances and

covariances matrix is estimated. This matrix may be calculated as Newey and West (1987) suggest:

$$\hat{\Omega} = \sum_{k=-m}^m w\left(\frac{k}{m}\right) \frac{1}{n} \sum_i d_{i+k} \hat{e}_{i+k} d'_{i+k} \hat{e}_{i+k}. \quad (17)$$

Intuitively, the test works as follows. If the variances and covariances matrix $E(u_i u'_i) = \tau^2 G$ is null, that is, $\tau^2 = 0$, the seasonal pattern of those cycles selected in A are not guided by the random walk described in the state Equation (15) but they are constant.

The matrix G is defined as:

$$G = (A' \Omega^f A)^{-1}, \quad \text{where} \quad (18)$$

$$\Omega^f = R'_1 \hat{\Omega} R_1, \quad \text{with} \quad (19)$$

$$R_1 = [f'_1, \dots, f'_s]. \quad (20)$$

The test for the null hypothesis, $H_0 : \tau^2 = 0$, against the alternative, $H_a : \tau^2 > 0$, is carried out with the statistic defined below:

$$\begin{aligned} L &= \frac{1}{n^2} \sum_{i=1}^n \hat{F}_i A (A' \hat{\Omega}^f A)^{-1} A' \hat{A}' \hat{F}_i \\ &= \frac{1}{n^2} \text{tr}((A' \hat{\Omega}^f A)^{-1} A' \sum_{i=1}^n \hat{F} \hat{F}' A), \end{aligned} \quad (21)$$

where F_i arise from f_i , defined in (16), as $\hat{F}_i = \sum_{t=1}^i f_t \hat{e}_t$, \hat{e}_t are the residuals obtained in Equation (14), $\hat{\Omega}^f$ is defined in (19), and $\text{tr}(Q)$ is the trace of Q .

Finally, Canova and Hansen (1995) explain that: $H_0 : L \xrightarrow{d} VM(a)$. That is to say, under the null hypothesis of stationarity in the a selected cycles, the statistic in expression (21) follows a *Von Mises* distribution with a degrees of freedom.

5 Protocol

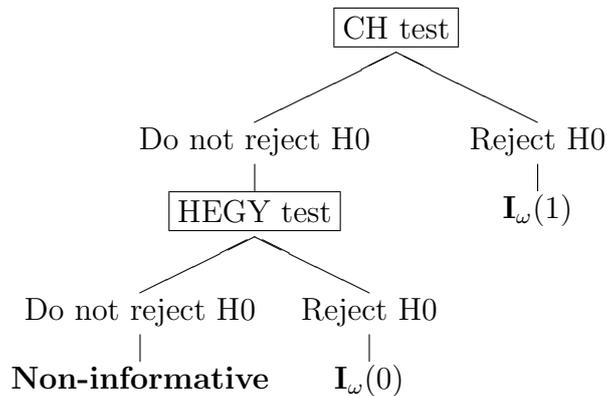
This section describes a protocol for detecting unit roots. As we have seen in Sections 3 and 4, the null hypothesis of ADF and HEGY tests are the alternative of KPSS and CH tests, and vice versa. To be brief, the remaining of the section refers to HEGY and CH tests. Nevertheless, taking into account that ADF and KPSS tests are alike to HEGY and CH, respectively, the conclusions in this section apply straightforward to long run unit root tests.

With regard to seasonal unit roots, Hylleberg (1995) shows some simulation exercises concluding that both tests, HEGY and CH, complement each

other. A variety of processes are considered and, depending on their features, it can be seen that when one of the tests fails to detect the data generating process, the other often leads to the right outcome. This conclusion is supported on the fact that the data generating process is a known one.

In practice, the structure of the data generating process is unknown, and obviously, it cannot be found out which the more suitable test is. Díaz-Emparanza and López-de Lacalle (2004) perform some simulations where, regardless to the data generating process, both tests are applied to reach a conclusion. This section describes the protocol for detecting unit roots used in those simulations. The arguments that justify the protocol are set out as well.

Figure 1: CH-HEGY sequence



The null hypothesis of the CH test is stationarity.

The null hypothesis of the HEGY test is that a unit root exists.

When both test statistics fail to reject their null hypotheses, it is advisable consider that a unit root exists.

According to the null and the alternative hypothesis defined in both tests, HEGY and CH, it may be expected that when one of them rejects its null hypothesis, the other will maintain its corresponding null hypothesis, and vice versa. However, when the data generating process departs from the model specified in the null hypotheses, the size and power of these tests become distorted. This is the case when there are some, but not all, seasonal unit roots, when some of the roots are close to unity, or when the data are affected by outliers, especially level shifts.

As a consequence, with real data, it may happen that both tests either maintain or reject their corresponding null hypotheses. In the first case, the data are non-informative since neither of the tests rejects its null hypothesis, perhaps due to a short sample. In the second case, conclusions are not coherent and a further analysis would be necessary, an outliers analysis, for instance, would be well advice.

This circumstance raises the following question. How does it affect the sequence in which both statistics are interpreted? What is the most suitable sequence for applying these tests? Figure 1 depict the decision tree for the CH-HEGY sequence.

It must be noted that on the basis of that tree, and given a significance level and power of the tests, it cannot be guessed the probability of reaching each one of the outcomes. The reason is that both statistics are not independent. Furthermore, we cannot determine the significance level entailed by the whole sequence of tests. Put in other words, the sequence must be chosen on the basis of other criteria. Anyway, a further insight into the case in which both statistics are independent will be first useful.

As a reference, under the fictitious case in which both statistics are independent, the corresponding conclusions would be obtained according to the results in Table 3. Given a 5% significance level in both tests, this table shows the probability of reaching each one of the conclusions when the data generating process is integrated or stationary. Both sequences, HEGY-CH and CH-HEGY, are considered.

Under the previous assumption, when a unit root exists, the CH-HEGY sequence detects this fact a μ per cent of all the cases, bigger than 0.95μ in the case of HEGY-CH. When the series is stationary, the sequence HEGY-CH is the one which detects this fact with more success, $\psi > 0.95 \psi$.

Even in that fictitious case it is an awkward matter to decide on one of the sequences, since it entails choosing which mistake we wish to minimize. As we brought forward in the first section, there is a trade-off between the following two matters that must be weighed up. On the one hand, if a unit root is omitted, estimates are inconsistent (Durlauf and Phillips, 1988). On the other hand, to assume more unit roots than those that are actually in the data will increase the variance of the estimator. The former mistake invalidates inference outcomes, whereas the latter increases error prediction variance, making forecast less accurate.

Judging the drawback of omitting a unit root more serious than the converse mistake, we advocate using the CH-HEGY sequence. It is also advised to consider the presence of a unit root in those cases in which the series is non-informative.

If a unit root exists, it would be better to conclude the analysis rejecting

Table 3: Test sequences

$I_\omega(1)$ series		
Hypothesis	CH-HEGY	HEGY-CH
$I_\omega(0)$	$0.05(1 - \mu)$	0.05
$I_\omega(1)$	μ	0.95μ
<i>Non-info.</i>	$0.95(1 - \mu)$	$0.95(1 - \mu)$

$I_\omega(0)$ series		
Hypothesis	CH-HEGY	HEGY-CH
$I_\omega(0)$	0.95ψ	ψ
$I_\omega(1)$	0.05	$0.05(1 - \psi)$
<i>Non-info.</i>	$0.95(1 - \psi)$	$0.95(1 - \psi)$

Cells report the percentage of cases in which each one of the

hypotheses is maintained.

μ is the power of CH test.

ψ is the power of HEGY test.

stationarity hypothesis rather than maintain the unit root hypothesis. As the CH-HEGY sequence begins testing the stationarity null hypothesis it takes position on favour of this goal. It also avoids that, depending on the power of the HEGY test, the analysis ends up rejecting a unit root when it actually exists.

Therefore, the CH-HEGY sequence solves the dilemma mentioned above taking a conservative approach. Although this sequence is expected to minimize the cases in which a unit root is omitted, this goal is achieved to the detriment of getting success in detecting stationary cases. Hence, the analysis focuses on avoiding inconsistent estimates allowing a possible increase in the variance of the estimator as little as possible.

Burridge and Taylor (2004) Burridge et al. (2003)

Part II

Applied Analysis

6 The `uroot` Package

6.1 Installing and running `uroot`

This paper documents the version 1.4 of the `uroot` package, which was built under R 2.1.0. The source code and binaries of the package are available at The Comprehensive R Archive Network [CRAN] ([http://cran.r-project.org/web/packages/uroot/index.html](#)). To install the package, copy the binaries in the subdirectory `library` where R is installed.³ Alternatively, download the source package and install it with R `CMD INSTALL uroot_1.4.tar.gz`. To install it from an R-console type `install.packages("uroot")` and select a mirror near to your location.

The package provides a graphical user interface [GUI] which is launched by the `urootgui()` function. The GUI displays each time series, their transformations, and selected subsamples by means of a tree widget. It also displays dialog boxes describing the arguments of the main functions implemented in `uroot`, in this way, these functions can be easily run. Nevertheless, not all the capabilities of those functions can be carried out from the GUI and some knowledge on the command-line functions is recommended.

The GUI is independent of the package and all the functions and data are available without launching it. It is based on the `tcltk` R-package and `BWidget`, an extension for `tcltk` that consists of Tcl scripts. `BWidget` is not included in the standard `tcltk` package. The version 1.7.0 of `BWidget`⁴ is attached to `uroot`. `urootgui()` automatically sets up this extension adding a new Tcl path directory to the `uroot` directory, where `BWidget` is stored.

Finally, the `xtable` R-package is suggested to convert objects to an `xtable` object, which can be printed as a \LaTeX or `HTML` object. The `urt.xtable` and `save.xtable` methods are based on this package and create a table with the main results of the unit root tests. As we will see, there are also some other useful functions that take advantage of this package to export results to a \LaTeX file.

³Type `R.home()` in an R-console to find out the home directory.

⁴An updated version can be downloaded from [http://www.gna.org/projects/bwidget/](#).

6.2 Loading data

The package provides sample data that consist of selected macroeconomic time series from the Basque Country,⁵ `mccapv`, and Spain,⁶ `mces`, as well as the data employed in the book Franses (1996), `mcmisc`. These data sets are stored in `.rda` file which contains a list object with the time series. A particular time series can be loaded from the menu **Data - Data bank** in the GUI. The complete list for each data set can be loaded as usual, `data("mces")`, then one or more time series can be picked up from the whole list. See the standard help pages for a description and sources of the data.

The GUI displays a dialog box for importing data stored in a `.csv` or `.txt` file. The user is asked for a label, periodicity and starting date of the sample, so that the suitable attributes can be added to the object with the data. In addition to the `ts` object, an item is created in the tree-widget of the GUI. If there are several items in the tree-widget menu, the data regarded as the input for any of the functions is the one that is highlighted. The user can change the working series clicking with the cursor of the mouse over the items.

In the example of the next section, the Real Gross National Product in Germany is analysed. These data are in the list `mcmisc` and the series is labeled as `gergnp`. We can start the example loading the package and the data:

```
> library(uroot)
> data("mcmisc")
> gergnp <- mcmisc$gergnp
```

7 An Example

This section illustrates with an example how to use the functions implemented in `uroot` for detecting long run and seasonal unit roots. All the functions used in this example are easy to run just with the right click over the menu in the GUI. However, the main commands for the console are showed so that the reader become acquainted with the names and usage of the functions. A description on the usage of these functions is also provided by the `uroot` standard help files.

The time series analyzed in this example is the Real Gross National Product in Germany, labeled as `gergnp`. Data are collected quarterly from 1960.1

⁵The original source is the IKERBIDE database developed by the Public Administration of the Basque Country.

⁶The original source is the Instituto Nacional de Estadística of Spain [INE].

to 1990.4. The series can be extracted from the `mcmisc` data set, as in the code above, or can be straightforwardly loaded from Franses (1996) data bank in the `Data` menu. The latter will create a list object called `gerpnp`, as well as an item in the tree widget.

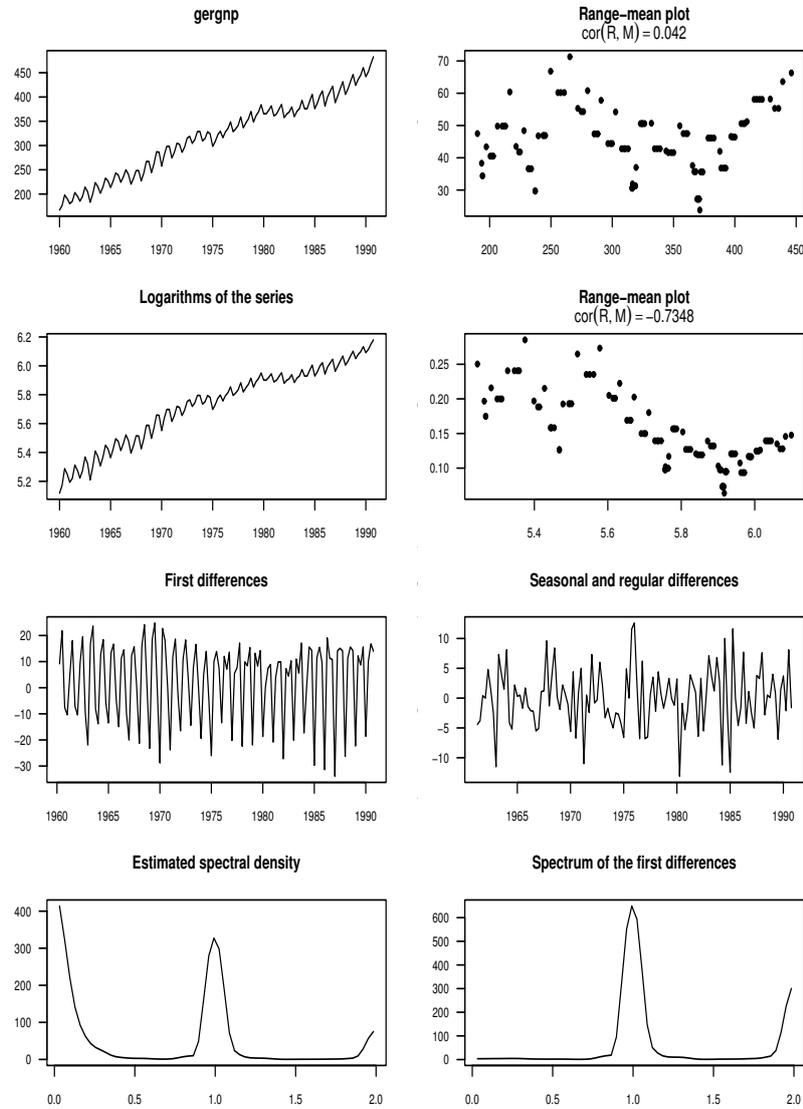
7.1 Graphics

7.1.1 Original series preview

The panel available in the `Utilities - Panel - Quarterly series - Selected graphics` menu is useful for a first glance at the data, see Figure 2. Graphics in this panel may be individually displayed in the `Graphics` menu. Looking at range-mean plots, Figure 2 shows that the variance remains relatively stable along the sample period.

There is no relationship between the range and the mean in the original series, since the cloud of points locates along a horizontal stripe. If logarithms are taken, the range-mean relationship becomes negative and large, -0.73 , instead. Hence, range-mean plots do not support taking logarithms. Taking logarithms is a common strategy when working with data, since it not only homogenizes the variance but also abates outliers and contributes to normalizing residuals. However, in this illustrative application we will be rigorous in interpreting the range-mean plot and the scale of the series will not be changed.

Figure 2: Real GNP in Germany. Selected graphics



The original series appears to have a prominent seasonal pattern. First differences plot, in the same Figure 2, seems to still retain this pattern, whereas adding the seasonal difference it vanishes.

Finally, estimated spectral density plots in Figure 2 confirm the presence of outstanding seasonal cycles, since that which completes one cycle per year explains a large proportion of the variance of the series. A lower peak at semiannual frequency is noticed as well.

Figure 3, available in the `Utilities - Panel - Correlograms` menu, displays autocorrelation functions [ACF] and partial autocorrelation func-

tions [PACF] for the original and differenced series. Broadly speaking, it may be noted that non-stationary features exist. ACF of the original series does not decrease exponentially. Nor does it the ACF of the first differences series, where seasonal correlations remains significant even for large orders.

Seasonally differenced series⁷, $\Delta^s(y)$, has a reasonable ACF for stationay data. After adding a regular difference, $\Delta \Delta^s(y)$, ACF and PACF do not suggests any structure and the data seem to be overdifferenciated in that case.

7.1.2 Seasonal operator

In the example below the seasonal operator is decomposed and plotted so that each one of the cycles collected in table 2, pg. 7, can be graphically analysed. The function `factorsdiff` is used for it. The argument `factors` is a 0-1 vector of length $S/2+1$, where S is the periodicity of the data. For quarterly series, the frequencies are ordered in the vector as follows: $0, \pi$, and $\pi/2$.⁸ A zero in the element i of the vector indicates that the cycle associated to the corresponding frequency will not be filtered. The cycles related to frequencies that are set equal to 1 are filtered.

```
> sdiff <- diff(gergnp, lag = 4)
> lrun <- factorsdiff(gergnp, factors = c(0, 1, 1))
> sann <- factorsdiff(gergnp, factors = c(1, 0, 1))
> ann <- factorsdiff(gergnp, factors = c(1, 1, 0))
> opar <- par(mfrow = c(2, 2), mar = c(2, 3, 3.5, 2), tcl = -0.5,
+ las = 1)
> plot(sdiff, main = "Seasonal difference")
> plot(lrun[[1]], main = "Long run")
> plot(sann[[1]], main = "Semiannual cycle")
> plot(ann[[1]], main = "Annual cycle")
> par(opar)
```

Figure 4 displays the seasonal differencing filter and its decomposition into long run cycle and sesonal cycles. The instability in the semiannual cycle, especially at the beginning of the sample, makes the semiannual frequency liable to contain a unit root. On the other hand, in spite of displaying a rather straight growing trend, the long run cycle undergoes changes in the level that depart it from a purely deterministic trend. A priori, those changes may be due to outliers such as level shifts or an additive outlier in 1975.I. However, to

⁷Remember that this filter removes a long run unit root as well.

⁸See `uroot` help files for monthly series.

keep inside the limits of this analysis, recursive testing is proposed in Section 7.2 in order to avoid that possible outliers distort the interpretation of the statistics.

7.1.3 Seasonal paths

Buy-Ballot plot is a common way of displaying seasonal time series. It can be obtained with the function `bbplot`. This function depicts the path of each season. In the example below, quarterly paths for original and first differenced series are displayed. See Figure 5. After removing the effect of the long run over seasonal paths, that is, in the first differenced series, the graphic shows that the quarterly paths are not parallel. This suggests that the seasonal pattern is not constant, hence, some seasonal unit roots are expected to be found. In addition, a slight trend in the seasonal paths can be detected. The user can take a different view of this plot by means of the following functions: `bb3D`; for a perspective plot of surface over the season-year plane, `bbc`; which displays the contour lines of the previous plot, and `bbap`; to plot the annual path of selected years.

```
opar <- par(mfrow=c(1,2), las=1, mar=c(2,3,3.5,2))
bbplot(wts=gergnp)
bbplot(wts=diff(gergnp, lag=1))
par(opar)
```

7.2 Unit root tests

This section applies the tests expounded in Sections 3 and 4. Previously, upon the guidance provided by the graphical analysis, significance of the deterministic components: Intercept, linear trend, and seasonal dummies, is discussed. We will do it on the basis of the regressions of the ADF and HEGY tests. For it, the arguments of the corresponding functions are described below.

We have seen that the rationale of these tests is similar, consequently, both functions, `ADF.test` and `HEGY.test` contain the same arguments, namely, the working time series, `wts`, deterministic components, `itsd`, other regressor variables, `regvar`, and the lag selection method, `selectlags`.

With regard to the deterministic components, the `itsd` argument must be a vector with the following structure: `c(0,0,c(0))`. If the first and/or second elements are set equal to 1, it indicates that an intercept and/or linear trend, respectively, are included. The third element is a vector indicating which

seasonal dummies should be included. If no seasonal dummies are desired it must be set equal to zero. For example, `regular=c(1,0,c(1,2,3))` would include an intercept, no trend, and the first three seasonal dummies.

Any other regressor variable, for example a dummy variable previously defined, can be included indicating the name of the corresponding matrix object in the `regvar` argument. By default this argument is set equal to zero.

Finally, `selectlags` is a list object indicating the method for choosing lags, `mode`, and the maximum number of lags allowed in the regression, `Pmax`. Available methods are "aic", "bic", and "signf". The first two methods fit two regressions: one with the maximum number of lags and another one removing the lag of the highest order. Then, if the first regression has an AIC or BIC lower than the second, the lag of the highest order is kept, otherwise it is dropped. This procedure is repeated for all the lags. "signf" fits a regression with the maximum number of lags and drops those that are not significant at the 10% level of significance. Then, the regression is fitted again with the significant lags. This procedure is repeated until all the lags included in the regression are significant. The user can also define `mode` as a numeric vector indicating the desired lag orders.

7.2.1 Deterministic components

For the design of the auxiliary regression in unit root tests, the deterministic components that should be included into it must be decided before. There is a major problem when making this choice. The distribution of the corresponding statistics is not the standard one but it depends on which unit roots exist. Nevertheless, it is known that, in the presence of unit roots, the critical values of those distributions move towards the extreme sides, hence, if the statistic lies between the standard critical values the null hypothesis cannot be rejected.

We will start fitting the ADF regression with all the possible deterministic components; an intercept, trend, and three out of the four seasonal dummies, to avoid multicollinearity. It is worth looking at the extended report on the test, `summary`, where the estimates of all the coefficients is showed. Here, to save space, only the `regvarcoefs` slot containing the coefficients we are now concerned with is showed. The remaining of the output will be discussed later.

```
> adf.out <- ADF.test(wts = gergnp, itsd = c(1, 1, c(1, 2, 3)),
+   selectlags = list(mode = "signf", Pmax = 8))
> round(adf.out@regvarcoefs, 2)
```

	Estimate	Std. Error	t value	Pr(> t)
Intercept	29.56	8.27	3.57	0.00
Trend	0.33	0.09	3.57	0.00
SeasDummy.1	-8.14	2.60	-3.13	0.00
SeasDummy.2	0.41	2.90	0.14	0.89
SeasDummy.3	2.17	2.29	0.95	0.35

Dickey and Fuller (1981) obtain the distribution of the t -statistic for an intercept and a linear trend, as well as some joint F -statistics. Taking as reference Tables II and III in Dickey and Fuller (1981) the critical values for an intercept and a trend are $3.11 < 3.57$ and $2.79 < 3.57$, respectively, with a 5% significance level, hence non-significance hypothesis is rejected. With regard to seasonal dummies, it cannot be rejected non-significance for two of them. Only the t -statistic of the first seasonal dummy, -3.13 , is bigger than the critical value ± 1.96 , taking as reference the Normal distribution. Since critical values in Dickey and Fuller (1981) do not consider the effect of seasonal dummies, it is advisable to check the same estimates in the HEGY regression. This is done in the code below.

```
> hegy.out <- HEGY.test(wts = gergnp, itsd = c(1, 1, c(1, 2, 3)),
+   selectlags = list(mode = "signf", Pmax = 8))
> round(hegy.out@regvarcoefs, 2)
```

	Estimate	Std. Error	t value	Pr(> t)
Intercept	25.49	8.74	2.92	0.00
Trend	0.26	0.10	2.51	0.01
SeasDummy.1	-9.77	2.63	-3.71	0.00
SeasDummy.2	-3.95	3.31	-1.19	0.24
SeasDummy.3	-0.39	2.68	-0.15	0.88

It can be noticed that the t -statistics for the intercept and trend moves to the non-rejection area. The coefficients of the seasonal dummies are negative. This is what we could expect, since the Buys-Ballot plot showed that the level of the series in the fourth season, the omitted seasonal dummy, tends to be higher than in the other seasons, specially in the second half of the sample.

On the other hand, it should be noticed that the effect of an intercept upon a seasonally differenced series is similar to a linear trend in the original series. A linear trend entails a near-exponential growth, whereas seasonal dummies entail that each season evolves along a linear trend. Díaz-Empananza and

López-de Lacalle (2004) explain these results. In this sense, graphics in the previous section showed that the long run of the series evolves rather straight, barring some recessions. At the same time, Buys-Ballot plot revealed that seasonal paths display a slight trend.

Allowing for the graphical analysis and the results mentioned here, it is feasible to consider an intercept and seasonal dummies with no trend in the auxiliary regressions of the tests.

7.2.2 Long run unit root tests

The `KPSS.test` function performs both null hypotheses described in 7.2, one in which data are level stationary, and the other where data are stationary around a linear trend. By default, the lag truncation parameter to compute the variance of the residuals is set equal to $3\sqrt{n}/13$, where n is the number of observations. However, it is a good practice to check the test for a sequence of different values.

Although we have decided not to consider a deterministic linear trend, there is no harm in taking a look at both statistics. In order to it, two object are created, `kpss.level` and `kpss.trend`. In the example below different lag truncation parameters are considered, from 0 to 4. The functions `urt.xtable` and `save.table` may be useful to this end. Given an object of class `kpssstat`, these functions compute the KPSS test for lags three times lower than that in the `kpssstat` object and three higher than that. The results are printed to the console or to a \LaTeX file.

```
> kpss.level <- kpss.trend <- rep(NA, 5)
> for (i in 0:4) {
+   kpss.level[i + 1] <- KPSS.test(wts = gergnp, ltrunc = i)@levelst
+   kpss.trend[i + 1] <- KPSS.test(wts = gergnp, ltrunc = i)@trendst
+ }
> print(data.frame(Trunc = c(0:4), Level = round(kpss.level, 2),
+   Trend = round(kpss.trend, 2)))
```

	Trunc	Level	Trend
1	0	11.94	0.57
2	1	6.12	0.41
3	2	4.16	0.38
4	3	3.16	0.34
5	4	2.55	0.27

The critical values for the 5% level of significance are 0.46 and 0.15, respectively. Regardless to the lag truncation parameter, both statistics reject their null hypotheses at the 5% level of significance. This entails that the long run is guided by a random walk.

As a complement, the null hypothesis of a random walk can be tested with the `ADF.test` function. We have already described the arguments of these functions. Here, it is clearer to show the output according to the `show` format.

```
> adf.out <- ADF.test(wts = gergnp, itsd = c(1, 0, c(1, 2, 3)),
+   selectlags = list(mode = "signf", Pmax = 8))
> show(adf.out)
```

```
-----
Augmented Dickey & Fuller test
-----

Null hypothesis: Unit root.
Alternative hypothesis: Stationarity.

-----
ADF statistic:

      Estimate Std. Error t value Pr(>|t|)
adf.reg    0.001    0.006   0.098    0.1

Lag orders: 1 2 4 7
Number of available observations: 116
```

Fuller (1976) (pp.373) collects critical values for the null hypothesis of a long-run unit root. It must be noted again that neither of these tables refer to the case in which an intercept and seasonal dummies are included. Nevertheless, the value of the statistic, 0.098, is close to zero and hence, relatively departed from either of the critical values; -1.95 , when none deterministic components are included, -2.89 , when an intercept is included, and -3.45 , if an intercept and an linear trend are included, for a significance level equals to 5%. Therefore, the unit root hypothesis cannot be rejected, confirming the results of the KPSS statistics.

The p-values for the test statistics are based on the tables from the original papers of each test. The `approx` function in the `stats` package is used to

interpolate the p-value according with the sample size. The critical values considered allow for significance levels ranged between 0.01% and 0.10%. If the interpolated p-value lies beyond these confidence levels a warning message is reported.

The p-values for regressors different from the test statistics reported in the original table, that is, deterministic components and lags, are the default values provided by `lm` and assume a Normal distribution for large samples.

Finally, note that lags of order 1, 2, 4, and 7 has been included, yielding a relatively parsimonious model and being reasonable for quarterly series. Taking other methods for choosing lags, the conclusions do not change. The reader can check that after taking first differences both statistics lead to conclude that there is not another long run unit root.

7.2.3 Seasonal unit root tests

In this subsection, following the protocol described in Section 5, CH and HEGY statistics are computed to detect seasonal unit roots. As it was mentioned, taking into account that outliers such as level shifts may distort the distribution of the statistics, recursive testing is proposed to assess the robustness of the results.

The user will find it useful the `recursive testing` option available in the menu of the GUI for these tests. This will save the user from having to write a loop in which the test is computed for different sample sizes. Some functions are also available from the console to have a greater control over this procedure. First, we will see how to run `CH.test` and `HEGY.test` and the way in which the results are displayed.

The example below tests the null hypothesis of stationarity for seasonal cycles by means of the CH joint test. As usual, `wts` is a `ts` object, in our case `gergnp`. The argument `frec` is a 0-1 vector of the same length as the number of seasonal frequencies, in quarterly series: $\frac{\pi}{2}$ and π , and in monthly series: $\frac{\pi}{6}$, $\frac{\pi}{3}$, $\frac{\pi}{2}$, $\frac{2\pi}{3}$, $\frac{5\pi}{6}$, and π , in this order. The cycles of the corresponding frequencies that are set to 1 in the vector `frec` are tested for stationarity.

As we have detected a long-run unit root, a first order lag is included in the regression, `f0=1`. A deterministic trend turned out to be non-significant in our analysis, hence, a linear trend is not considered, `DetTr=FALSE`. Finally, the lag truncation parameter for estimating the residuals covariance matrix is the default value, $S(n/100)^{0.25}$, where S is the periodicity and n the number of observations.

```
> ch.out <- CH.test(wts = gergnp, frec = c(1, 1), f0 = 1, DetTr = FALSE,
+   ltrunc = NULL)
> ch.out
```

```
----- - ----- ----  
Canova & Hansen test  
----- - ----- ----
```

```
Null hypothesis: Stationarity.  
Alternative hypothesis: Unit root.  
Frequency of the tested cycles: pi/2 , pi ,
```

```
L-statistic: 2.279  
Lag truncation parameter: 4
```

```
Critical values:
```

```
0.10 0.05 0.025 0.01  
0.846 1.01 1.16 1.35
```

The CH statistic, 2.279, must be compared with a Von Mises distribution with three degrees of freedom.⁹ For a 5% level of significance the critical value is $1.01 < 2.279$, hence, the null hypothesis of stationary seasonal cycles is rejected and the unit roots -1 and $\pm i$ exist, according to this test.

To look for more evidence in favour of the previous result, we will test for seasonal roots by means of the recursive procedure. According to this strategy, the statistics are computed for different subsamples, so that the effect of prospective outliers on the results can be assessed.

The `CH.rectest` and `HEGY.rectest` functions are used in the code below. In addition to the arguments in `CH.test` and `HEGY.test`, the type and length of the subsamples are defined and some information on the iteration progress is requested as well.

The number of observations in each subsample or in the starting subsample, depending on the type of subsamples, is set by the argument `nsub`. Three types of subsamples can be considered: “backw”, the statistic is computed for the last `nsub` observations and then one year backwards is added until the beginning of the sample; “forw”, the statistic is computed for the first `nsub` observations and then one year forwards is added until the end of the sample; “moving”, the statistic is computed over moving subsamples of length `nsub`.

Three levels of information can be printed during the procedure. The argument `trace` is a list object indicating what trace of the iteration progress

⁹Remember that three roots are tested -1 and $\pm i$.

must be printed:¹⁰ if `remain` is equal to 1, the percentage of the whole procedure that has been completed is printed; if `plot` is equal to 1, a plot with the computed statistics is displayed; if `elaps` is equal to 1, how much time was required to complete the whole procedure is reported.

```
chrec.pi <- CH.rectest(wts=gergnp, type="backw", nsub=32,
+ frec=c(0,1), f0=1, DetTr=FALSE, ltrunc=NULL,
+ trace=list(remain=1, plot=0, elaps=1))
hegyrec.pi2 <- HEGY.rectest (wts=gergnp, type="backw", nsub=32,
+ itsd=c(1,0,c(1,2,3)), selectlags=list(mode="signf", Pmax=NULL),
+ trace=list(remain=1, plot=0, elaps=1))
```

The first example tests for stationary cycle of frequency π and the second example tests for the seasonal unit roots $\pm i$. To see the output of the recursive test, `show` and `plot` methods can be used. In the examples, 32 observations have been taken for the initial subsample. When more regressors are included in the model, for example in the HEGY regression for monthly data with seasonal dummies and lags of high orders, that size may not be enough data to fit the model. In that case, larger subsamples should be considered or a lower maximum lag, `Pmax`, could be tried. Here, to save space, results are collected in Tables 4 and 5.

These tables can be built by means of the `recch.save.xtable` and `rechegy.save.xtable` functions,¹¹ which are a mixture of the functions for recursive testing and the functions `xtable` and `print.xtable` in the `xtable` package. In fact, the arguments are the union of the arguments in those functions.

```
recch.save.xtable (wts=gergnp, nsub=32, frec=c(0,1), f0=1, DetTr=FALSE,
+ ltrunc=NULL, trace=list(remain=1, plot=0, elaps=1),
+ caption="CH recursive test. Cycle of frequency  $\pi$ ",
+ label="Tchrec", align="lcrll", type="latex", table.placement="ht",)

rechegy.save.xtable (wts=gergnp, nsub=32, itsd=c(1,0,c(1:3)),
+ refstat="Fpi3:4", selectlags=list(mode="signf", Pmax=NULL),
+ trace=list(remain=1, plot=0, elaps=1), label="Thegyrec",
+ caption="HEGY recursive test. Cycle of frequency  $\pi/2$ ",
+ align="lcrll", type="latex", file=NULL, table.placement="ht")
```

¹⁰In windows operative systems, the option `Buffered output` in the `Misc` menu must be active for this information to be printed while the process is running.

¹¹These functions are implement for the ADF and KPSS tests as well. See `recadf.save.xtable` and `reckpss.save.xtable` in the help pages of the package.

This utility is not implemented in the GUI menu. The `Utilities - LaTeX output` menu refers to objects created after running the ADF, KPSS, HEGY, or CH tests. When these functions are run via the GUI, results are stored in the `.out` object. In this way, the main results of these functions can be exported to a `LaTeX` table.

Following with the example, after some minor changes the code above creates Tables 4 and 5.

Table 4: CH recursive test. Cycle of frequency π

Sample	Stats		Sample	Stats	
1960:1–1967:4	0.62	**	1983:1–1990:4	0.10	
1960:1–1968:4	0.66	**	1982:1–1990:4	0.10	
1960:1–1969:4	0.75	***	1981:1–1990:4	0.11	
1960:1–1970:4	0.91	***	1980:1–1990:4	0.21	
1960:1–1971:4	0.93	***	1979:1–1990:4	0.15	
1960:1–1972:4	0.92	***	1978:1–1990:4	0.11	
1960:1–1973:4	0.83	***	1977:1–1990:4	0.10	
1960:1–1974:4	0.64	**	1976:1–1990:4	0.09	
1960:1–1975:4	0.76	***	1975:1–1990:4	0.17	
1960:1–1976:4	0.84	***	1974:1–1990:4	0.33	
1960:1–1977:4	0.94	***	1973:1–1990:4	0.50	*
1960:1–1978:4	1.04	***	1972:1–1990:4	0.64	**
1960:1–1979:4	1.15	***	1971:1–1990:4	0.81	***
1960:1–1980:4	1.16	***	1970:1–1990:4	0.76	***
1960:1–1981:4	1.20	***	1969:1–1990:4	0.83	***
1960:1–1982:4	1.32	***	1968:1–1990:4	0.92	***
1960:1–1983:4	1.43	***	1967:1–1990:4	0.99	***
1960:1–1984:4	1.39	***	1966:1–1990:4	1.11	***
1960:1–1985:4	1.47	***	1965:1–1990:4	1.21	***
1960:1–1986:4	1.60	***	1964:1–1990:4	1.34	***
1960:1–1987:4	1.69	***	1963:1–1990:4	1.46	***
1960:1–1988:4	1.75	***	1962:1–1990:4	1.59	***
1960:1–1989:4	1.84	***	1961:1–1990:4	1.73	***
1960:1–1990:4	1.83	***	1960:1–1990:4	1.83	***

Table 4 reports the CH statistics computed recursively for the seasonal unit root -1 . To save space the table for the seasonal unit roots $\pm i$ is omitted. The significance levels are labeled as follows: “0.01”, “***”; “0.025”, “**”; “0.05”, “*”; and “0.10”, “.”. Stationarity in the semiannual cycle is rejected in most of the subsamples. The reader can check that the same happens for

the annual cycle. In spite of such a strong evidence, it must be noted that those observations at the beginning of the sample are the main source of non-stationarity. If the latest subsamples are considered, stationarity cannot be rejected for either of the seasonal cycles.

Keeping in mind the decision tree in Figure 1, page 12, we will get more information computing the HEGY statistics for the seasonal unit roots $\pm i$, Table 5. The table for the seasonal unit root -1 is omitted. The reader may check that the seasonal unit root -1 is not rejected in most of the subsamples at the 5% level of significance.

According to the pattern of the recursive HEGY statistics in Table 5, we may hesitate to consider the presence of the $\pm i$ unit roots, since the null hypothesis is rejected in some subsamples. To get a further insight into the analysis of these roots, we will compute the HEGY statistics including a dummy that stands for a structural shift halfway the sample, from 1968 to 1977.

The `Mdates` function helps to build such dummy. It searches for the position in the sample of a given date. Setting the argument `yso` to 1968.1 and 1977.1, this function looks these dates up in a quarterly calendar that collects all the observation dates in the original time series and stores them in the objects `obs1` and `obs2`. The dummy called `Mdumm` is then built just setting to 1 the elements from `obs1` to `obs2` in a column-matrix made up of zeros.

```
> obs1 <- Mdates(gergnp, yso = c(1968, 1))@obs
> obs2 <- Mdates(gergnp, yso = c(1977, 1))@obs
> Mdumm <- matrix(0, nrow = length(gergnp), ncol = 1)
> Mdumm[obs1:obs2, 1] <- 1
> hegy.out <- HEGY.test(wts = gergnp, itsd = c(1, 0, c(1, 2, 3)),
+   regvar = Mdumm)
> hegy.out
```

```
-----
HEGY test
-----
```

```
Null hypothesis: Unit root.
Alternative hypothesis: Stationarity.
```

```
-----
HEGY statistics:
```

Table 5: HEGY recursive test. Cycle of frequency $\pi/2$

Sample	Stats		Sample	Stats	
1960:1–1967:4	9.16	**	1983:1–1990:4	3.48	
1960:1–1968:4	10.85	***	1982:1–1990:4	3.28	
1960:1–1969:4	3.14		1981:1–1990:4	5.39	
1960:1–1970:4	11.15	***	1980:1–1990:4	6.35	.
1960:1–1971:4	1.55		1979:1–1990:4	6.74	*
1960:1–1972:4	6.66	*	1978:1–1990:4	1.76	
1960:1–1973:4	12.41	***	1977:1–1990:4	1.92	
1960:1–1974:4	17.98	***	1976:1–1990:4	13.02	***
1960:1–1975:4	14.93	***	1975:1–1990:4	6.25	.
1960:1–1976:4	9.98	***	1974:1–1990:4	20.87	***
1960:1–1977:4	8.19	**	1973:1–1990:4	20.89	***
1960:1–1978:4	1.58		1972:1–1990:4	24.43	***
1960:1–1979:4	1.98		1971:1–1990:4	26.65	***
1960:1–1980:4	1.67		1970:1–1990:4	25.63	***
1960:1–1981:4	1.91		1969:1–1990:4	10.33	***
1960:1–1982:4	2.18		1968:1–1990:4	7.31	*
1960:1–1983:4	2.11		1967:1–1990:4	6.03	.
1960:1–1984:4	2.82		1966:1–1990:4	6.88	*
1960:1–1985:4	4.50		1965:1–1990:4	5.72	.
1960:1–1986:4	5.11		1964:1–1990:4	5.64	.
1960:1–1987:4	4.48		1963:1–1990:4	6.07	.
1960:1–1988:4	4.77		1962:1–1990:4	6.01	.
1960:1–1989:4	4.94		1961:1–1990:4	5.12	.
1960:1–1990:4	5.28		1960:1–1990:4	5.28	.

	Stat.	p-value
tpi_1	0.603	0.1
tpi_2	-2.400	0.1
Fpi_3:4	5.250	0.1
Fpi_2:4	5.422	NA
Fpi_1:4	4.161	NA

Lag orders: 1

Number of available observations: 119

The results¹² show that a prospective shift between 1968 and 1977 does not change the results for the whole sample. The statistics for the seasonal roots -1 and $\pm i$ are -2.40 and 5.25 , respectively, and remain in the area where the null hypothesis is maintained. In fact, the generic dummy is not significant, as the t -statistic, 0.70 , suggests (See the slot `regvarcoefs` in `hegy.out` or `summary(hegy.out)`).

To summarize, the following model collects the features found in the analysis carried out through this example.

$$(1 - L)(1 + L)(1 + L^2) y_t \equiv \Delta y_t = \alpha + \sum_{i=1}^4 \gamma_i D_{i,t} + \epsilon_t.$$

The functions introduced in this paper allowed us to decompose the seasonal differencing filter and to test whether the modulus of each one of the roots are equal to unity. Likewise, tests for stationary cycles were carried out as well. A graphical analysis was useful for a first glance at the data and to decide the deterministic regressors in the auxiliary regressions of the tests.

¹²In this version, the p-values for the $F_{pi_2} : 4$ and $F_{pi_1} : 4$ statistics in the HEGY tests are not reported.

Figure 3: Real GNP in Germany. Correlograms

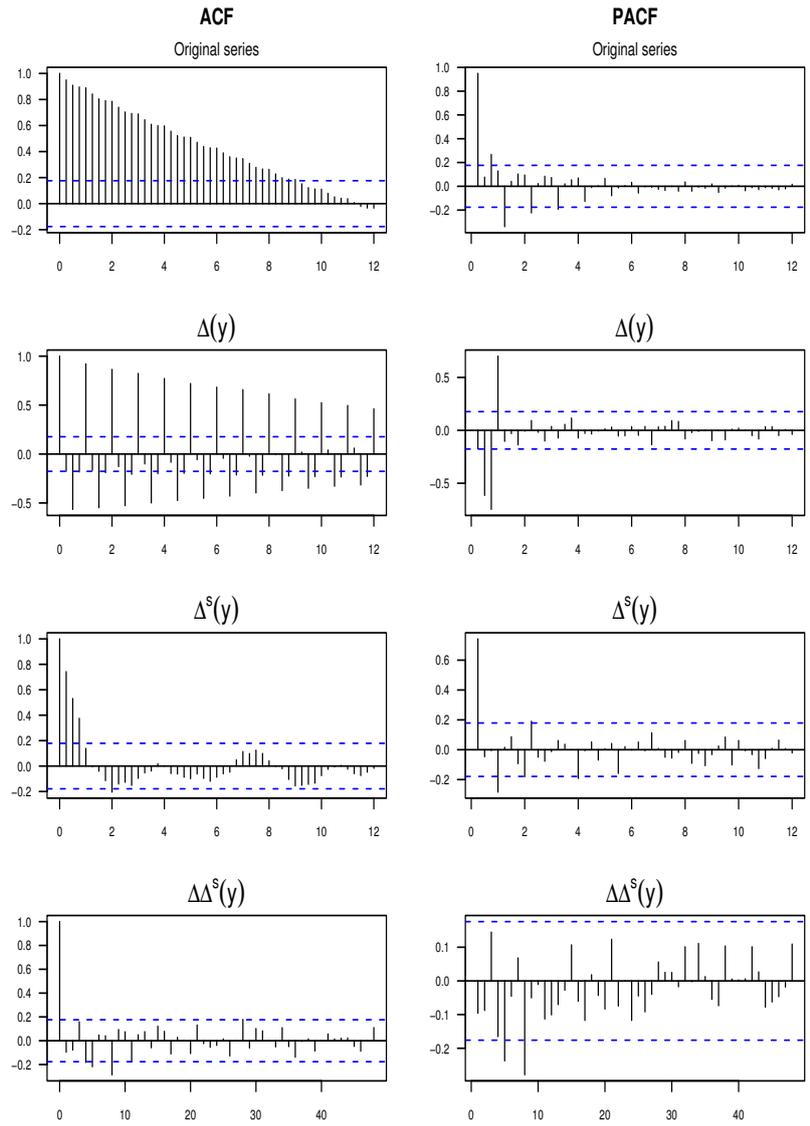


Figure 4: Real GNP in Germany. Seasonal cycles

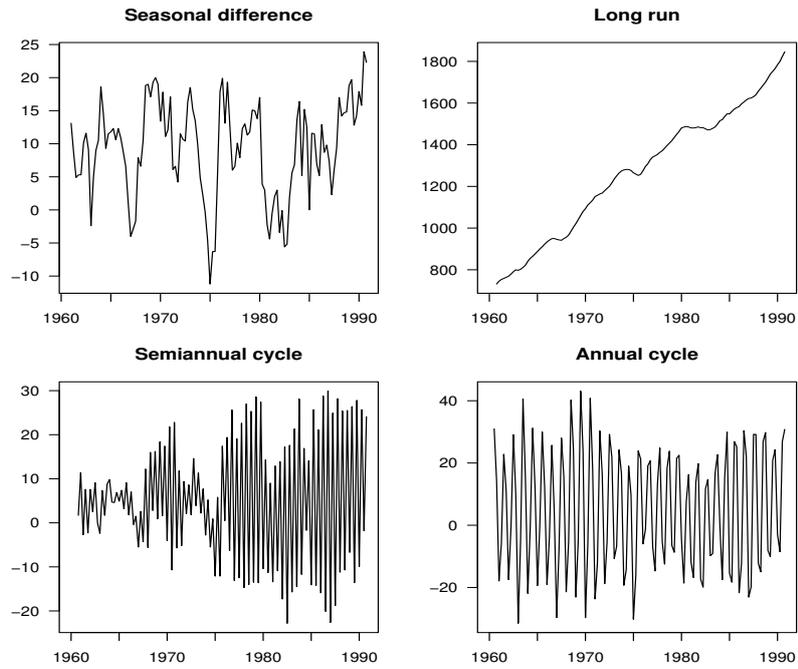
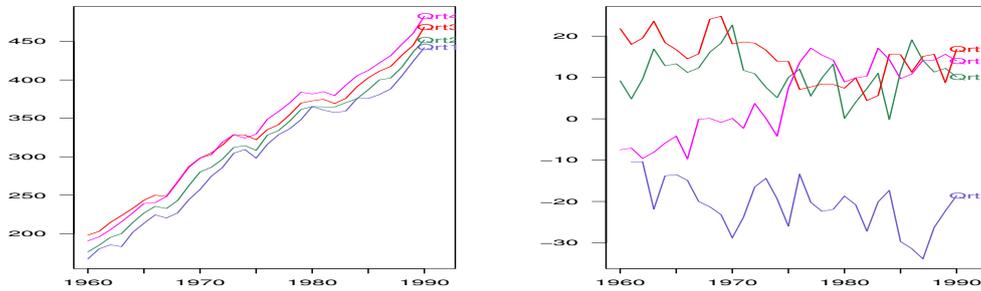


Figure 5: Real GNP in Germany. Buys-Ballot plot



References

- Banerjee, A., Dolado, J., Galbraith, J. and Hendry, D. (1993), *Co-Integration, Error Correction, and the Econometric Analysis of Non-Stationary data*, Advanced Texts in Econometrics. Oxford University Press.
- Beaulieu, J. and Miron, J. (1993), ‘Seasonal Unit Roots in Aggregate U.S. Data’, *Journal of Econometrics* **54**, 305–28.
- Burridge, P., Gjorstrup, F. and Taylor, A. (2003), Robust inference on seasonal unit roots via bootstrap applied to OECD macroeconomic series, City Economics Discussion Paper 0408.
- Burridge, P. and Taylor, A. (2001), ‘On the Properties of Regression-Based Tests for Seasonal Unit Roots in the Presence of Higher Order Serial Correlation’, *Journal of Business and Economic Statistics* **19**, 374–379.
- Burridge, P. and Taylor, A. (2004), ‘Bootstrapping the HEGY Seasonal Unit Root Tests’, *Journal of Econometrics* **123**(1), 67–87.
- Canova, F. and Hansen, B. E. (1995), ‘Are seasonal patterns constant over time? a test for seasonal stability’, *Journal of Business and Economic Statistics* **13**, 237–252.
- Dickey, D. and Fuller, W. (1981), ‘Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root’, *Econometrica* **49**, 1057–1071.
- Dickey, D., Hasza, D. and Fuller, W. (1984), ‘Testing for Unit Roots in Seasonal Time Series’, *Journal of American Statistical Association* **79**, 355–67.
- Durlauf, S. and Phillips, P. (1988), ‘Trends versus Random Walks in Time Series Analysis’, *Econometrica* **56**, 1333–54.
- Díaz-Emparanza, I. and López-de Lacalle, J. (2004), Estacionalidad Determinista y Estocástica en Series Temporales Macroeconómicas, BILTOKI DT 200402, EAII, EAIII, FAE e IEP: Universidad del País Vasco.
- Franses, P. (1990), Testing for Seasonal Unit Roots in Monthly Data, Technical Report 9032, Econometric Institute.
- Franses, P. (1996), *Periodicity and Stochastic Trends in Economic Time Series*, Advanced Texts in Econometrics. Oxford University Press.

- Franses, P. and Hobijn, B. (1997), ‘Critical values for unit root tests in seasonal time series’, *Journal of Applied Statistics* **24**, 25–47.
- Fuller, W. (1976), *Introduction to Statistical Time Series*, John Wiley.
- Hylleberg, S. (1992), *Modeling Seasonality*, Advanced Texts in Econometrics. Oxford University Press.
- Hylleberg, S. (1995), ‘Tests for Seasonal Unit Roots. General to Specific or Specific to General?’, *Journal of Econometrics* **69**, 5–25.
- Hylleberg, S., Engle, R., Granger, C. and Yoo, B. (1990), ‘Seasonal Integration and Cointegration’, *Journal of Econometrics* **44**, 215–38.
- Kwiatkowski, D., Phillips, P., Schmidt, P. and Shin, Y. (1992), ‘Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?’, *Journal of Econometrics* **54**, 159–178.
- Nelson, C. and Plosser, C. (1982), ‘Trends Versus Random Walks in Macroeconomic Time Series’, *Journal of Monetary Economics* **10**, 139–62.
- Newey, W. and West, K. (1987), ‘A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix’, *Econometrica* **55**, 703–8.
- Phillips, P. (1987), ‘Time Series Regression with a Unit Root’, *Econometrica* **55**(2), 277–301.
- R Development Core Team (2005), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL: <http://www.R-project.org>