ΟΙΚΟΝΟΜΙΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ

ATHENS UNIVERSITY
OF ECONOMICS
AND BUSINESS

SCHOOL OF INFORMATION SCIENCES & TECHNOLOGY

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

Testing for Non-Stationary Stochastic Seasonality with an application to the Greek Inflation

By

Theonymfi I. Boura

A THESIS
Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfilment of the requirements for
the degree of Master of Science in Statistics

Athens, Greece
Month Year

ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ & ΤΕΧΝΟΛΟΓΙΑΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ

Έλεγχος για Μη-Στάσιμη Στοχαστική Εποχικότητα με εφαρμογή στον Ελληνικό Πληθωρισμό

Θεονύμφη Ι. Μπούρα

ΔΙΑΤΡΙΒΗ

Που υποβλήθηκε στο Τμήμα Στατιστικής

του Οικονομικού Πανεπιστημίου Αθηνών

ως μέρος των απαιτήσεων για την απόκτηση

Μεταπτυχιακού Διπλώματος Ειδίκευσης στη Στατιστική

Αθήνα

Μήνας Έτος

DEDICATION

To my mother

ACKNOWLEDGEMENTS

I would first like to thank my thesis supervisor Evangelos Ioannidis, assistant professor of the Department of Statistics at Athens University of Economics and Business. He consistently supported me and gave me the right directions, comments and remarks through the learning process of this master thesis.

I would also like to thank my loved ones and express my profound gratitude to my mother and my sister for providing me with support and encouragement through my years of study and especially through the process of writing this thesis.

This accomplishment would not have been possible without them. Thank you.

VITA

My name is Theonymfi I. Boura and I was born in Heraklion, Crete on June, 24, 1992. I finished high-school in 2010 and between 2010 and 2015 I studied Mathematics at the National and Kapodistrian University of Athens. On October of 2015, I started the Full-Time MSc of Statistics at the Athens University of Economics and Business which I am finishing this coming September.

ABSTRACT

Theonymfi Boura

Testing for Non-Stationary Stochastic Seasonality with an application to the Greek Inflation

September 2017

In Time Series Analysis, many processes apart from trend may display seasonality. Although, the most famous and commonly used is the deterministic, there are two other types of seasonality that differ significantly from this, the so-called non-stationary and stationary stochastic seasonality.

With regard to the stochastic seasonality, we detect and differentiate the non-stationary from the stationary stochastic seasonality by conducting seasonal Unit Root Tests. Seasonal Unit Root Tests constitute the extension to seasonal models of the well-known Unit Roots test for the null of a series being integrated (e.g. a random walk) versus it being stationary. The main focus of this thesis is to present and discuss two such tests. The first one is the seasonal Augmented Dickey Fuller test and the second one is the so-called HEGY unit root test. Both of them test the null hypothesis of non-stationary stochastic seasonality versus the alternative of stationarity stochastic seasonality. They do however make different assumptions on the structure of the null and the alternative and focus on somehow different aspects of it.

The use as well as the main characteristics of these tests are illustrated with an application using the dataset of the Greek Inflation.

ΠΕΡΙΛΗΨΗ

Θεονύμφη Μπούρα

Έλεγχος για μη στάσιμη στοχαστική εποχικότητα με εφαρμογή στον Ελληνικό Πληθωρισμό

Σεπτέμβρης 2017

Στην Ανάλυση Χρονοσειρών, πολλές διαδικασίες πέρα από τάση παρουσιάζουν και εποχικότητα. Η ντετερμινιστική εποχικότητα αποτελεί την πιο γνωστή μορφή εποχικότητας και χρησιμοποείται στον μεγαλύτερο βαθμό. Εν τούτοις, υπάρχουν και άλλες δύο μορφές εποχικότητας. Η μη-στάσιμη και η στάσιμη στοχαστική εποχικότητα.

Αναφορικά με την στοχαστική εποχικότητα, την διακρίνουμε και ταυτόχρονα διαχωρίζουμε τη μη-στάσιμη από τη στάσιμη, πραγματοποιώντας του εποχικούς ελέγχους Μοναδιαίας Ρίζας. Οι εποχικοί έλεγχοι Μοναδιαίας Ρίζας, αποτελούν την επέκταση των ελέγχων Μοναδιαίας ρίζας οι οποίοι έχουν ως μηδενική υπόθεση τη χρονοσειρά να είναι ολοκληρωμένη (π.χ τυχαίος περίπατος) έναντι της εναλλακτικής τη χρονοσειρά να είναι στάσιμη, στα μοντέλα που παρουσιάζουν εποχικότητα. Η παρακάτω διπλωματική εργασία παρουσιάζει και πραγματεύεται δύο τέτοιους ελέγχους.

Ο πρώτος είναι ο εποχικός αυξημένος Dickey-Fuller έλεγχος και ο δεύτερος είναι ο HEGY έλεγχος μοναδιαίας ρίζας. Και οι δύο έχουν ως μηδενική υπόθεση την ύπαρξη μη-στάσιμη στοχαστικής εποχικότητας και ως εναλλακτική την ύπαρξη στάσιμη στοχαστικής εποχικότητας. Παρόλ’ αυτά έχουν διαφορετικές υποθέσεις σχετικά με τη δομή της μηδενικής και της εναλλακτικής και εστιάζουν σε διαφορετικές πλευρές του προβλήματος.

Η χρήση καθώς και τα κύρια χαρακτηριστικά των ελέγχων αυτών παρουσιάζονται σε μία εφαρμογή με τα δεδομένα του Ελληνικού Πληθωρισμού.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

A common assumption in many time series techniques as well as traditional econometric methods require the data to fulfil stationarity. Most of the macroeconomics time series though, do not satisfy stationarity conditions as they display trend, seasonality or both.

Many stationarity tests exist in the literature. Pagan and Schewert (1990) proposed several non-parametric tests such us the Cumulative Sum test (CUSUM) or Modified scaled range test. Subsequently, Ahamada and Butahar (2002) proposed two other non-parametric tests to examine covariance stationarity. See also Carlo De Michele and Harry Pavlopoulos (2007) for an application to rainfall data and Priestley (1965) for an approach based on evolutionary spectra.

In the present thesis though, we will concerned with testing for Non-stationarity Stochastic Seasonality in the same sense as traditional testing for unit-roots: in these approaches the null hypothesis is formulated the presence of a specific form of non-stationarity, namely the presence of a root on the unit circle, which ios tested against the alternative of stationarity. This is different from the approaches of testing for stationarity in the previously mentioned papers in that in these papers the null hypothesis is the one of stationarity which is tested against the alternative of non-stationarity.

In the present thesis though, we will concerned with the so-called Non-stationarity Stochastic Seasonality rather than the covariance Non-stationarity.

With regard to seasonal pattern, there are three different types of seasonality and there are various models that display these types. The first one is the deterministic seasonality which describes behavior in which the periodic pattern is due to the unconditional mean of the time series.

Except for the deterministic seasonality that maintains a constant seasonal pattern and is the most familiar in use in Time Series Analysis there are also the so-called stochastic stationary and stochastic non-stationary seasonality that display a seasonal pattern which will randomly vary from one cycle to the next.

Stochastic seasonality is often described by the mixed seasonal ARMA$(P,Q)_S$ which is defined by the equation:

$$\boldsymbol{\varphi}(B)\boldsymbol{\Phi}(B^S)Y_t = \boldsymbol{\theta}(B)\boldsymbol{\Theta}(B^S)Z_t\ ,\ Z_t \sim WN(0,\sigma^2)$$

where $\boldsymbol{\varphi}(z)$, $\boldsymbol{\theta}(z)$, $\boldsymbol{\Phi}(z)$ and $\boldsymbol{\Theta}(z)$ are the seasonal and non-seasonal AR and MA polynomials and P and Q are the orders of the non-seasonal polynomials and S is the period of the seasonal pattern.

The stochastic stationary seasonality refers to the roots of $\boldsymbol{\varphi}(z)$ and $\boldsymbol{\Phi}(z)$, which are all assumed to lie outside the unit circle. Specifically, it is more pronounced when the roots of the polynomial $\boldsymbol{\Phi}(z^S)$ are close to the unit circle, but also when the roots of $\boldsymbol{\varphi}(z)$ are close to the unit circle too. The only difference between the roots of these polynomials is that the roots of $\boldsymbol{\Phi}(z^S)$ come in groups of S members with a specific structure and the same modulus, whereas the roots of $\boldsymbol{\varphi}(z)$ have their own flexibility.

Allowing for differencing the mixed seasonal ARMA$(P,Q)_S$ models lead us to the models for nonstationary stochastic seasonality which are often described by the Seasonal ARIMA$(p,d,q)x(P,D,Q)_S$ models:

If d and D are non-negative integers then Y_t is said to be a seasonal ARIMA$(p,d,q)x(P,D,Q)_S$ process with period S if the differenced series $X_t=(1-B)^d(1-B^S)^DY_t$ is a casual ARMA process. Thus we assume that satisfies:

$$\boldsymbol{\varphi}(B)\boldsymbol{\Phi}(B^S)\ (1-B)^d(1-B^S)^DY_t=\boldsymbol{\theta}(B)\boldsymbol{\Theta}(B^S)Z_t,\ Z_t\sim WN(0,\sigma^2)$$

where $\boldsymbol{\varphi}(z)$, $\boldsymbol{\theta}(z)$, $\boldsymbol{\Phi}(z)$ and $\boldsymbol{\Theta}(z)$ are the seasonal and non-seasonal AR and MA polynomials. The letters p,q and P and Q are the orders of the non-seasonal and the seasonal polynomials respectively and S is the period of the seasonal pattern.

Generalizing these models, let us consider the ARMA(p,q) model that satisfies the equation below:

$$\widetilde{\boldsymbol{\Phi}(\ddot{B})}Y_t = \widetilde{\boldsymbol{\Theta}(\ddot{B})}Z_t\ ,\ Zt\sim WN(0,\sigma^2)$$

with $\widetilde{\ddot{\boldsymbol{\Phi}}(z)}$ and $\widetilde{\ddot{\boldsymbol{\Theta}}(\mathrm{B})}$ some other polynomials. In this general case of an ARMA(p,q) model, any root of $\widetilde{\ddot{\boldsymbol{\Phi}}(z)}$on the unit circle will result a non-stationary Y_t.

We detect and differentiate the non-stationary from the stationary stochastic seasonality by conducting seasonal Unit Root Tests. Main focus of this thesis is to present and discuss two such tests.

The first one is the seasonal Dickey Fuller (seasonal DF test) and Augmented Dickey Fuller (seasonal ADF test) Unit Root Testing of Dickey, Hasza and Fuller (1984). These tests examine the null hypothesis that all the roots of $\boldsymbol{\Phi}(B^S)$ are on the unit circle versus the alternative that the roots have the same modulus. They are the straight forward extension of the simple Dickey Fuller and Augmented Dickey Fuller Unit Root test proposed by Dickey and Fuller (1976).

Specifically, consider SARIMA model:

$$Y_t = \alpha_S Y_{t-S} + \varepsilon_t \ ,\varepsilon_t \sim WN(0,\sigma^2)$$

The seasonal ADF test will test the null hypothesis of α_s being equal to unity (the null of stochastic non-stationary seasonality) against the alternative hypothesis of α_s being smaller than unity.

However, seasonal unit roots may be present at some, but not at all the frequencies. Therefore, a joint test for all the seasonal frequencies simultaneously, such as the one proposed by seasonal ADF test, will not provide the appropriate result. Therefore, we demonstrate and describe the HEGY Unit Root test proposed by Hylleberg, Engle, Granger and Yoo (1990) for quarterly data (S=4) and Beaulieu and Miron (1993) for monthly data (S=12).

HEGY Unit Root test allows testing for individual roots as it has the benefit to look for unit roots at any single seasonal frequency (as well as the zero frequency) without imposing roots at other frequencies. The procedure of transforming the provided data in order to present the final estimated equation as well as the total derivation of the HEGY test are also explained in this thesis.

To sum up, in Chapter 1 is described the classical Unit root Dickey- Fuller and Augmented Dickey Fuller test for trending models that do not display seasonality. In Chapter 2 are defined the different types of seasonality as well as the models that we use in order to describe them. Subsequently in the Chapter 3 is discussed in detail the seasonal DF and the seasonal ADF Unit Root tests while the fourth contains the description and the derivation of the HEGY Unit Root test. Finally, in Chapter 5 we apply the seasonal ADF and HEGY Unit Root test to a dataset of the Greek Inflation using the statistical package R. The purpose of this application is to detect seasonal and non-seasonal unit roots and therefore examine the presence of stationary and non-stationary stochastic seasonality in the Greek Inflation

CHAPTER 2

Unit Root and Stationary Processes

A common assumption in many time series techniques as well as traditional econometric methods require the data to fulfil stationarity. Most of the macroeconomics time series though, do not carry out stationary conditions as they display trend, heteroscedasticity or both. A covariance-stationary process has the property that the mean, variance and autocorrelation structure do not change over time. More specifically, according to Wold's Theorem all the covariance-stationary processes, can be written in the form of:

$$\mathbf{Y_t = \mu + \sum_{j=0}^{\infty} \boldsymbol{\psi_j}\, \boldsymbol{\varepsilon_{t-j}}}$$

Where ε_t is the white noise error one would make forecasting Y_t as a linear function of lagged Y_t where $\sum_{j=0}^{\infty} \psi j^2 < \infty$ and $\psi_0=1$.

2.1 <u>Trending Time Series Models</u>

In the case of a trending time series, some form of trend removal is required. There are two popular "detrending" procedures, the time-trend removal and the first differencing. The first one is suitable for trend-stationary or I(0) processes and the second one is suitable for unit root or I(1) processes. The I(0) processes are stationary after the trend removal, while the I(1) processes are non-stationary and the stationarity is achieved by applying first differences. To determine which "detrending" method is the appropriate, we apply a Unit Root Test. The null hypothesis of this test indicates a process Y_t to be non-stationary (while the first difference ΔY_t is stationary) and the alternative hypothesis indicates the opposite (Y_t to be stationary).
Therefore, a Unit Root Test examines the hypothesis of the following form:

H_0: I(1) process- ΔY_t is stationary
H_1: I(0) process- $Y_t - E(Y_t)$ is stationary

This form can be used in order to determine if trending data should be differenced or regressed on deterministic functions of time to render the data stationary.

The equations below, which we will discuss in this chapter, are two popular processes in which due to a trend in the data (upward or downward over time) the mean is not constant. Although they both have a trend, the nature of their non-deterministic part is different:

$$\mathbf{Y_t=\delta+Y_{t-1} + \varepsilon_t} \quad \textit{(1.1.1)}$$
$$\mathbf{Y_t= \alpha+\beta t +\varepsilon_t} \quad \textit{(1.1.2)}$$

where ε_t is white noise, *iid*(0,1). The process in equation (*1.1.1*), is known as a random walk with drift δ or a process with unit root which has a so-called *stochastic trend (beyond the linear)*. The generated process in equation (*1.1.2*) has an intercept α as well as a *deterministic (linear) time trend* with slope equals to β. We will refer to the first process as a Unit Root process with drift and to the second one as a Trend Stationary Process. Trend Stationary and Unit Root processes are both trending over time, have the same mean but conduct a different stochastic behavior. For the general definition see below.

In the Trend Stationary Process, the mean is replaced by a linear function of the date t and if one subtracts the deterministic trend $\alpha+\beta t$, the result is a stationary process. On the other hand, in the Unit Root Process the mean is a linear trend and its variance is not constant. Thus, the Unit Root Process is nonstationary even after removing the trend.

It is useful at this point to report the properties as well as the differentiation among these processes. In order to define those properly, we will express the above models using the form described in Chapter 3 and 15 of J.D.Hamilton's book - Time Series Analysis.

The general definition of these two models described above, is as follows.

<u>Trend Stationary Model</u>

The processes that include a deterministic time trend are defined as:

$$\mathbf{Y_t= \alpha+\beta t +\psi(L)\varepsilon_t = \alpha+\beta t + \varepsilon_t+ \psi_1\varepsilon_{t-1}+\psi_2\varepsilon_{t-2}+\ldots}$$

where $\psi(z)$ is the polynomial $1+\psi_1 z+\psi_2 z^2+\psi_3 z^3+\ldots$ and $\varepsilon_t \sim WN(0,\sigma^2)$.

Unit Root Process

The unit root processes that shape a stochastic trend are defined as:

$$\mathbf{(1-L)Y_t=\delta + \psi(L)\varepsilon_t = \delta + \varepsilon_t + \psi_1\varepsilon_{t-1}+\psi_2\varepsilon_{t-2}+\ldots}$$

where $\psi(z)$ is the polynomial $1+\psi_1 z+\psi_2 z^2+\psi_3 z^3+\ldots$, $\varepsilon_t \sim WN(0,\sigma^2)$and $\psi(1)\neq 0$.

2.2 Trend Stationarity and Unit Root Processes

In this part we will point the main and very crucial differences between the aforementioned models.

A. Forecasts

Trend-Stationary

To forecast a trend-stationary process we add the deterministic component **α+βt** to the forecast of the stationary stochastic component. Therefore, the proper forecast is $\hat{Y}_{t+s|t} = \alpha+\beta(t+s)+ \psi_s\varepsilon_t+ \psi_{s+1}\varepsilon_{t-1}+ \psi_{s+2}\varepsilon_{t-2}+\ldots$. Furthermore, it is proven that as s reaches the infinity the forecast tends to reach the initial line:

$$\hat{Y}_{t+s|t} \rightarrow \alpha+\beta t \quad \text{as } s\rightarrow +\infty$$

and the Mean Squared Error converges to zero:

$$E[\hat{Y}_{t+s|t}-\alpha-\delta(t+s)]^2 \rightarrow \sigma^2\sum_{j=s}^{\infty}\psi_j^2 \rightarrow 0 \quad \text{as} \quad s\rightarrow+\infty$$

Therefore, for large s the information we have until time T, is being lost and we forecast the stationary process by using its expected value.

Unit Root

The forecast of a Unit Root process is:

$$\hat{Y}_{t+s|t} = s\delta + Y_t + (\psi_s+\psi_{s-1}+..+\psi_1)\varepsilon_t + (\psi_{s+1}+\psi_s+..+\psi_2)\varepsilon_{t-1} + ..$$

For the special case of the random walk with drift δ, where $\psi_1=\psi_2=..=0$, the forecast is:

$$\hat{Y}_{t+s|t} = s\delta + Y_t$$

In this case, the forecast moves parallel to the initial line. As a result, once a stochastic disturbance extracts the time series from the initial line, the time series is not predicted to reach the line again.

The exact proof of the forecast is being explained in the appendix-1a.

B. Comparison of forecasts errors and Variance

Trend-Stationary

The s-period-ahead forecast error for a trend stationary process is:

$$Y_{t+s}-\hat{Y}_{t+s|t}= \{\alpha+\delta(t+s)+\varepsilon_{t+s}+\psi_1\varepsilon_{t+s-1}+..+\psi_{s-1}\varepsilon_{t+1}+\psi_s\varepsilon_t+\psi_{s+1}\varepsilon_{t-1}+..\} - \{\alpha+\delta(t+s)+\psi_s\varepsilon_t+\psi_{s+1}\varepsilon_{t-1}+ \psi_{s+2}\varepsilon_{t-2}+\ldots\}= \varepsilon_{t+s}+\psi_1\varepsilon_{t+s-1}+ \psi_2\varepsilon_{t+s-2}..+\psi_{s-1}\varepsilon_{t+1}$$

The Mean Squared Error is:

$$E[Y_{t+s} -\hat{Y}_{t+s|t}]^2=\{1+\psi_1^2+\psi_2^2+\psi_3^2+..+\psi_{s-1}^2\}\sigma^2$$

and as ***s*** reaches the infinity (***s***$\rightarrow\infty$) it converges to the unconditional variance of the stationary component $\psi(L)\varepsilon_t$.

As we can see, as the length of the forecast horizon becomes large, the MSE of a stationary process reaches a finite bound.

Unit Root

The s-period-ahead forecast error for a unit root process is:

$$Y_{t+s}-\hat{Y}_{t+s|t}= \{\Delta Y_{t+s}+\Delta Y_{t+s-1}+..+\Delta Y_{t+1}+Y_t\}- \{\Delta\hat{Y}_{t+s|t}+ \Delta\hat{Y}_{t+s-1|t}+\ldots+\Delta\hat{Y}_{t+1|t}+ Y_t \}= \varepsilon_{t+s}- \{1+\psi_1\}+\{1+\psi_1+\psi_2\}\varepsilon_{t+s-2}+..+\{1+\psi_1+\psi_2+\psi_3+..+\psi_{s-1}\}\varepsilon_{t+1}$$

The Mean Squared Error is:

$$E[Y_{t+s} -\hat{Y}_{t+s|t}]^2= \{1+ (1+\psi_1)^2+(1+\psi_1+\psi_2)^2+...+ (1+\psi_1+\psi_2+\psi_3+..+\psi_{s-1})^2\}\sigma^2$$

As we can see the MSE of a unit root process, grows linearly as the forecast horizon becomes large.

These results are illustrated in the figure below.

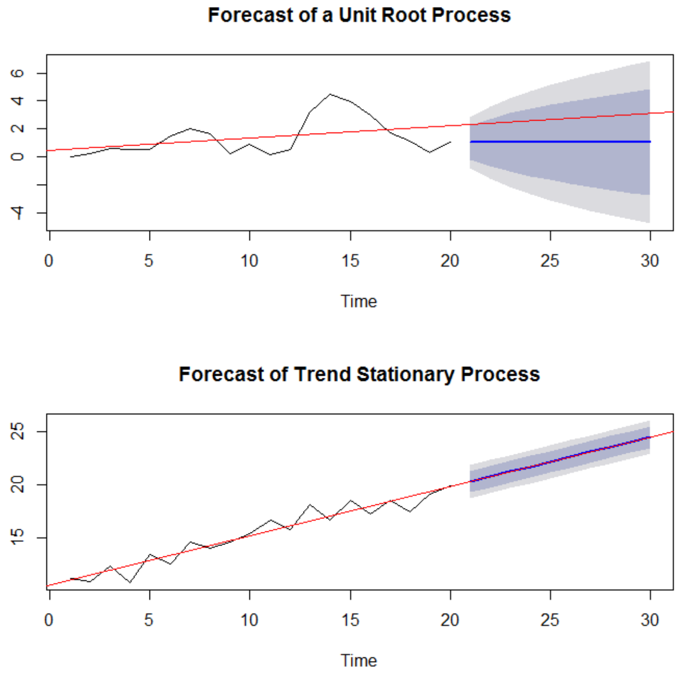


Figure 2.1*- Forecasts and 95% confidence intervals*

C. Dynamic Multipliers

In the case of dynamic multiplier, we will assume the consequences on Y_{t+s} if ε_t were to increase by one unit, with ε's for all other dates unaffected.

Trend-Stationary

$$\frac{\partial Y_{t+s}}{\partial \varepsilon_t} = \psi_s$$

For a trend-stationary process, the limiting dynamic multiplier as $s \to +\infty$ is zero. Therefore, the impact of any stochastic disturbance (shock), eventually disappear.

Unit Root

$$\frac{\partial Y_{t+s}}{\partial \varepsilon_t} = \frac{\partial \Delta Y_{t+s}}{\partial \varepsilon_t} + .. + \frac{\partial \Delta Y_{t+1}}{\partial \varepsilon_t} + \frac{\partial Y_t}{\partial \varepsilon_t} = \psi_s + \psi_{s-1} + .. + \psi_1 + 1$$

For a unit-root process, the limiting dynamic multiplier as $s \to +\infty$ is $\psi(1)$. Therefore, the effect of a great shock will retain.

D. <u>Transformations to achieve stationarity</u>

<u>Trend-Stationary</u>

In order to produce a stationary process, we simply subtract δt from the time series equation.

<u>Unit Root</u>

The proper procedure to achieve stationarity is to difference the time series.

2.3 The Dickey-Fuller Unit Root Test

One of the most popular Unit Root tests, is the Dickey-Fuller test, which will be described in this part. First, we will define the models listed below, with which this unit root test will be presented:

- $Y_t = \boldsymbol{\rho} Y_{t-1} + \varepsilon_t$(the AR(1) model) $\Leftrightarrow \Delta Y_t = (\boldsymbol{\rho}-1) Y_{t-1} + \varepsilon_t$ *(1.3.1)*
- $Y_t = \mu + \boldsymbol{\rho} Y_{t-1} + \varepsilon_t \Leftrightarrow \Delta Y_t = \mu + (\boldsymbol{\rho}-1) Y_{t-1} + \varepsilon_t$ *(1.3.2)*
- $Y_t = \alpha + \beta t + \boldsymbol{\rho} Y_{t-1} + \varepsilon_t \Leftrightarrow \Delta Y_t = \alpha + \beta t + (\boldsymbol{\rho}-1) Y_{t-1} + \varepsilon_t$ *(1.3.3)*

where $\varepsilon_t \sim WN(0,\sigma^2)$.

According to David A. Dickey and Wayne A. Fuller (1979), in order to perform the Dickey Fuller test, we need to compare the hypothesis presented below[1]:

H_0: Y_t is random walk (pure, with drift or with linear term)
H_1: Yt-E(Yt) is stationary AR(1) (simple, with intercept or with linear term)

These hypotheses in the framework of equations (1.3.1)-(1.3.3), can be written equivalently as:

H_0: **$\rho=1$,** $\{\Delta Y_t\}$ is a stationary process

[1] For a different approach to Unit root testing, see e.g. Phillips (1987)

H_1: $\boldsymbol{\rho}$<1, $Y_t - E(Y_t)$ is a stationary process AR(1)

Under the assumption of $\boldsymbol{\rho}=\pm1$, Dickey and Fuller observed and computed the limiting distribution of the OLS estimator of $T(\boldsymbol{\hat{\rho}-1})$[2] as well as the regression $\mathbf{t}=\frac{\hat{\rho}-1}{s.e(\hat{\rho})}$ statistic. Specifically, for each one of the models *(1.3.1)*, *(1.3.2)* and *(1.3.3)*, where $Y_0=0$, representations for the limiting distributions of the OLS estimator $\boldsymbol{\hat{\rho}}$ and the regression statistic $\boldsymbol{\hat{t}}$ were derived. These representations were used for the construction of tables of percentage points for these statistics. Having these tables at our disposal, we are able to test the above hypothesis and perform the Dickey-Fuller test where we use the **$T(\hat{\rho}$-1)** statistic, as well as the Dickey-Fuller **t** test where we use the $\hat{t}=\frac{\hat{\rho}-1}{s.e(\hat{\rho})}$ statistic.

In the table below, are being displayed the 5% percentage points of the estimators of the three models for $T=+\infty$.

Estimated Equation	5% Percentage Points of $\hat{\rho}$	5% Percentage Points of $\hat{t}$
$Y_t=\boldsymbol{\rho}Y_{t-1}+\varepsilon_t$	-8.1	-1.95
$Y_t=\mu+\boldsymbol{\rho}Y_{t-1}+\varepsilon_t$	-14.1	-2.86
$Y_t=\mu+\beta t+\boldsymbol{\rho}Y_{t-1}+\varepsilon_t$	-21.8	-3.41

Table 2.1*-The 5% percentage points of the estimators for T=+∞*

As far as the expected value of the aforementioned models is concerned, we conclude to the results below:

- If the data generating process is the $Y_t=\boldsymbol{\rho}Y_{t-1}+\varepsilon_t$ then $\begin{cases} E(Y_t)=0 \ \ if\ \rho\neq 1 \\ E(Y_t)=constant\ if\ \rho=1 \end{cases}$

- If the data generating process is the $Y_t=\mu+\boldsymbol{\rho}Y_{t-1}+\varepsilon_t$ then $\begin{cases} E(Y_t)=constant\ \ if\ \rho\neq 1 \\ E(Y_t)=linear\ if\ \rho=1 \end{cases}$

- If the data generating process is the $Y_t=\alpha+\beta t+\boldsymbol{\rho}Y_{t-1}+\varepsilon_t$ then $\begin{cases} E(Y_t)=linear\ \ if\ \rho\neq 1 \\ E(Y_t)=quadratic\ if\ \rho=1 \end{cases}$

[2] In the case of the AR(1) model, the OLS estimator is the $\hat{\rho}=\frac{\sum_{t=1}^{n}Y_tY_{t-1}}{\sum_{t=1}^{n}Y_{t-1}^2}$.

Therefore, if we detect that the expected value $E(Y_t)$ of the process Y_t is constant then we estimate the model $Y_t=\mu+\boldsymbol{\rho}Y_{t-1}+\varepsilon_t$ in order to describe the trending behavior under the null as well as the alternative hypothesis.(H_0: $\rho=1$ vs H_1: $\rho\neq1$).If $E(Y_t)$ appears as linear, we estimate the model $Y_t=\alpha+\beta t+\boldsymbol{\rho}Y_{t-1}+\varepsilon_t$ so that the trending behavior is described under the two hypothesis. Correspondingly, if $E(Y_t)$ appears as quadratic we estimate the model $Y_t=\alpha+\beta t+\gamma t^2+\boldsymbol{\rho}Y_{t-1}+\varepsilon_t$ and so on.

Finally, the powers of these statistics were computed and compared with that of the Box-Pierce Q[3] Statistic. It was concluded that the statistics proposed are uniformly more powerful than the Q statistics

These conclusions as well as the Dickey Fuller test, concern the three aforementioned models under the assumption of ε_t's to be White Noise. This hypothesis though, raise the forthcoming question:

"What is the respective Unit Root Test and what are the arising results for corresponding models in the presence of serial correlation of the errors?"

In other words, what is the extension of the Dickey Fuller Unit Root test and what are the limiting distributions for the corresponding statistics for models which shape the problem of serial correlation of the errors ($\varepsilon_t$is not White Noise)?

2.4 <u>The Augmented Dickey-Fuller Unit Root Test</u>

In this part, we set $u_t= Y_t -\mathbf{r}Y_{t-1}$ (thus $u_t= \Delta Y_t$ under r=1) and we generalizing by allowing $u_t\sim AR(p)$ while in 2.3 the corresponding u_t is assumed as White Noise. Therefore, we apply the Augmented Dickey Fuller test or ADF test.

The ADF test is the extension of the simple Dickey Fuller Test. Once the problem of autocorrelation occurs (the errors ε_t's are not White Noise), there are included extra lagged terms in the dependent variable in order to achieve the error term to be white noise.

Specifically, consider the model where $u_t$is an AR(p), p>1 stationary process:

[3] The Box and Pierce test uses the statistic $Q_k=n\sum_{k=1}^{K} r_k^2$ where $r_k=\frac{\sum_{t=k+1}^{n}\hat{e}_t\hat{e}_{t-k}}{\sum_{t=1}^{n}\hat{e}_t^2}$ and the $\widehat{e_t}$'s are the residuals from the fitted model. Under the null hypothesis, the Q_k statistic is approximately distributed as a chi-squared random variable with K-p degrees of freedom, where p is the number of parameters estimated. If Y_t includes a unit root, then $\rho=0$ under the null hypothesis and $\widehat{e_t}= Y_t-Y_{t-1}$.

$Y_t = \mathbf{r}Y_{t-1}+u_t$, where $u_t =\rho_1u_{t-1} + \rho_2u_{t-2}+\rho_3u_{t-3}+\ldots+ \rho_pu_{t-p}+\varepsilon_t$ where $\varepsilon_t\sim NID(0,\sigma^2)$ *(1.4.1)*

It is easy to observe that this equation is equivalent to $(1-\mathbf{r}L)Y_t = u_t$ and $(1-\rho_1L- \rho_2L^2- \rho_3L^3 -\ldots- \rho_pL^p)u_t=\varepsilon_t$ *(1.4.2)*, which may be written as $(1-\mathbf{r}L)(1-\rho_1L- \rho_2L^2- \rho_3L^3 -\ldots- \rho_pL^p)Y_t = \varepsilon_t$ where $\varepsilon_t\sim NID(0,\sigma^2)$ *(1.4.3).*

We will now manipulate these equations, in order to bring them in a more convenient form.

Specifically, setting the equations $\boldsymbol{\rho}=\rho_1+ \rho_2+\ldots+\rho_p$ *(1.4.3)* and $J_j = -[\rho_{j+1} + \rho_{j+2} +\ldots \rho_p]$ *(1.4.5)* for j=1,2,..,p-1, it is concluded that:

$$(1-\boldsymbol{\rho}L)-(J_1L+J_2L^2+..+J_{p-1}L^{p-1})(1-L)= 1-\rho_1L- \rho_2L^2- \rho_3L^3 -\ldots- \rho_pL^p$$

where L is the backshift operator.

Thus assuming that u_t is an AR(p), the process can be written equivalently as:

$$(1-\boldsymbol{\rho}L)-(J_1L+J_2L^2+..+J_{p-1}L^{p-1})(1-L)Y_t=\varepsilon_t \quad \textit{(1.4.6)}$$

or

$$Y_t = \boldsymbol{\rho}Y_{t-1}+J_1\Delta Y_{t-1}+ J_2\Delta Y_{t-2} +\ldots+ J_{p-1}\Delta Y_{t-p+1} + \varepsilon_t \quad \textit{(1.4.7)}$$

which can easily be transformed to the following equation:

$$\boxed{\Delta Y_t = (\rho-1)Y_{t-1}+J_1\Delta Y_{t-1}+ J_2\Delta Y_{t-2} +\ldots+ J_{p-1}\Delta Y_{t-p+1} + \varepsilon_t \quad \textit{(1.4.8)}}$$

where $\rho_1=\boldsymbol{\rho}+J_1$, $\rho_i=J_i-J_{i-1}$, i=2,..,p and $\rho_{p+1}= -J_p$

The form *(1.4.8)* is the most suitable in order to perform unit root testing and we will refer to it as the ADF equation.

Deterministic terms as intercept and linear trend can be also added to the ADF equation. This will give the opportunity to perform the ADF test for models with $E(Y_t)$ other than zero. Therefore we define:

$$\Delta Y_t=\mu+ (\rho-1)Y_{t-1}+J_1\Delta Y_{t-1}+ J_2\Delta Y_{t-2} +\ldots+ J_{p-1}\Delta Y_{t-p+1} + \varepsilon_t \quad \textit{(1.4.9)}$$

$$\Delta Y_t=\alpha+\beta t+ (\rho-1)Y_{t-1}+J_1\Delta Y_{t-1}+ J_2\Delta Y_{t-2} +\ldots+ J_{p-1}\Delta Y_{t-p+1} + \varepsilon_t \quad \textit{(1.4.10)}$$

Taking into consideration the forms *(1.4.8), (1.4.9)* and *(1.4.10)*, in order to test the hypothesis that our process Y_t is a random walk or stationary AR(p) and by extension perform a unit root test, it is sufficiently equivalent to test the hypothesis below:

H_0: ρ=1, $\{\Delta Y_t\}$ is an AR(p) process (simple, with intercept or linear trend)
H_1: ρ<1, Y_t is an AR(p+1) process (simple with intercept or linear trend)

The parameters ρ, $J_{1,\dots}$,J_{p-1} of the ADF equation estimated with Ordinary Least Squares and the distribution of the estimator $\boldsymbol{\hat{\rho}}$ is not dependent on the Ji's under the null hypothesis. For the case of the simple ADF equation *(1.4.8)*, we test for unit root by using either of the two following statistics:

The statistic $Z=\frac{T(\hat{\rho}-1)}{(1-\widehat{\zeta_1}-..\widehat{\zeta_p})}$ and the t statistic $\frac{\hat{\rho}-1}{\widehat{\sigma_{\hat{\rho}}}}$ which both have the same limiting distribution as the corresponding statistics $\hat{\rho}$ and $\hat{t}$ in the case of the first order models (models where ε_t's is white noise), discussed in the previous part. The same conclusions also apply for the equations *(1.4.9)* and *(1.4.10)*.

Correspondingly, the expected value of the aforementioned models arises as above.

Finally, as we notice, the equation that we estimate in order to perform the ADF test has the same form as the estimated equation in the case of the simple DF test *(1.3.1)*. Their significant difference, is that extra lagged differences have been added in the ADF equation to achieve the elimination of the autocorrelation in the u_t terms.

CHAPTER 3

The Different Types of Seasonality

Seasonality in time series is the presence of variations that occur at specific regular intervals such as weekly monthly or quarterly periods. Generally, any pattern in a time series that recurs or repeats over a certain period can be said to be seasonal. Moreover, in the ACF plot there is a repeating pattern which decays slowly. Seasonality, may be caused by various factors such as weather, vacation and holidays consisting periodic repetitive and generally regular and predictable patterns in the levels of a time series.

3.1 <u>The Detection of Seasonality</u>

It is important to consider and describe the effects of seasonality, in order to understand the impact of this component upon a given series. For example, a business that presents higher sales in certain seasons appear to be having significant profit during peak seasons and significant losings during off-peak seasons. After detecting the seasonality and establishing the seasonal pattern, specific techniques can be applied in order to eliminate it from the time-series. This procedure is known as "de-seasonalizing". Furthermore, the past patterns of seasonal variations can be used to the forecasting and the prediction of the future trends.
There are various graphical methods that can be used to detect seasonality. Some of them are listed below:

1. <u>The Run-Sequence Plot</u>: It is an easy way to plot a univariate dataset. In the Run Sequence plot, all the possible shifts in location and scale, as well as the outliers are sufficiently obvious. The vertical axis contains the response variable Y_t while the horizontal axis, contains the index i (i=1,2,3…).

2. <u>The Autocorrelation Plot</u>: It is the most commonly-used plots in time series analysis. In the vertical axis is placed the autocorrelation coefficient r_h[4]and in the horizontal axis it is placed the time lag h.

[4]The autocorrelation coefficient is defined as $r_h=\frac{\gamma_h}{\gamma_0}$ where γ_h is the autocovariance function and γ_0 is the variance function.

3. The Seasonal Subseries Plot: It is very popular tool for detecting seasonality in time series. Although, it is only useful if the period of seasonality is already known. The vertical axis, contains the response variable Y_t while the horizontal axis is contains the time ordered by season.

At this point it is important to realize how we use these plots in order to detect seasonality. For this purpose, we chose the dataset 'nottem' from the statistical package R which is known as times series that displays seasonality. The dataset, contains the *Average Air Temperature* at Nottingham Castle in degrees Fahrenheit for 20 years (1920-1939). Subsequently, we applied the plots described above and we tried to detect seasonal patterns.

It is enough evident from Figure 3.1 below that this time series dataset display seasonality. The Run Sequence plot although it shows periodic behavior, we can hardly say that seasonality is displayed, but observing the other two we conclude with certainty that it really is. Specifically, the Autocorrelation plot depicts a strong seasonality pattern while the Seasonal Subseries plot reveals that during the summer months the average air temperature is remarkably higher than the one during autumn or spring.

If Seasonality is present and by extension, it must be incorporated in the time series model.

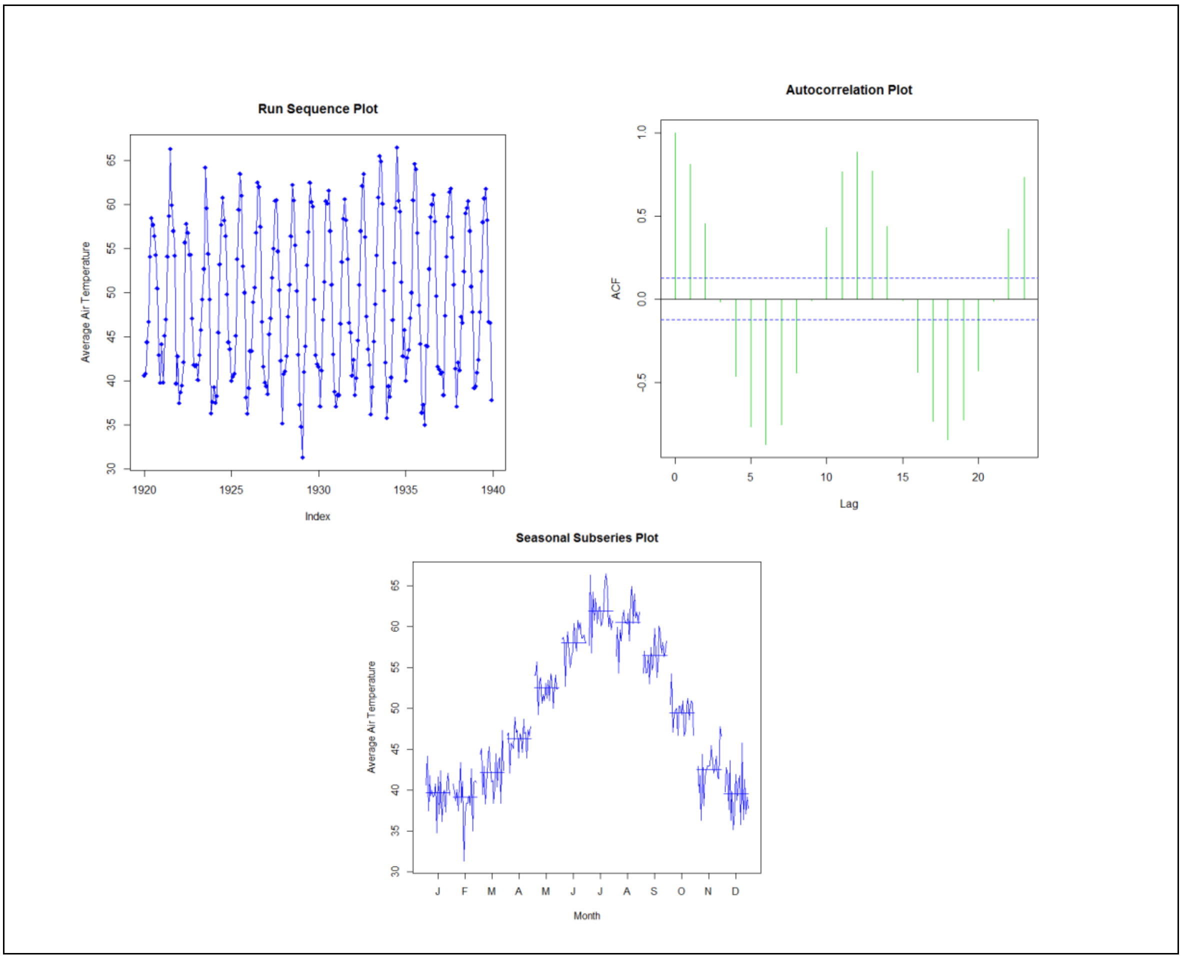


Figure 3.1*-Plots for Seasonality Detection*

3.2 <u>The Different Types of Seasonality</u>

As it was defined in the previous parts, by seasonality we mean the periodic patterns that exist at regular intervals. However, there are different seasonal models that describe seasonality and are divided in three different classes. These classes are:

- The Deterministic Seasonality
- The Stochastic Stationary Seasonality
- The Stochastic Non-Stationary Seasonality

3.2.1 <u>The Deterministic Seasonality</u>

<u>Deterministic Seasonality</u>, is the first type of seasonality. It describes behavior in which the periodic pattern is due to the unconditional mean of the time series, for example this concept is applied to time constant seasonal mean that differ across quarters or months. Deterministic seasonality, can be expressed by two alternative ways. The dummy variable representation where means of seasonal dummy variables that are 1 in specific quarters and 0 otherwise, are applied in the model, as well as the trigonometric representation.
We will list these representations as they were defined in Eric Ghysel's book, "The Econometric Analysis of Seasonal Time Series".

3.2.1.a <u>The Dummy Variable Representation</u>

The conventional dummy variable representation of seasonality can be written as:

$$Y_t=\sum_{s=1}^{S}\gamma_s\delta_{st} + z_t \text{ , } t=1,2,..,T \quad \textit{(2.3.1.a)}$$

where z_t is a stationary stochastic process with zero mean and $\delta_{st}=\begin{cases}1, t = s(modS)\\0, otherwise\end{cases}$ s=1,2,..S are seasonal dummy variables.
Therefore, for season s of year τthe expected value is $E(Y_t)=\gamma_s$, s=1,2,…S which implies that the process has sifting mean and that's the reason that Y_t is not stationary. However, by subtracting the mean of each season $\mu=\frac{1}{s}\sum_{s=1}^{S}\gamma_s$the deterministic seasonal effect is $m_s=\gamma_s-\mu$ and stationarity is being achieved. Hence the deviations $Y_t-E(Y_t)=z_t$ are stationary.

The above definition though, implies that $\sum_{s=1}^{S}m_s=0$ with the interpretation that there is no deterministic seasonality when observations are summed over a year. When the level of the series is separated from the seasonal component, then:

$$Y_t=\mu+ \sum_{s=1}^{S}m_s\delta_{st} + z_t \text{ , } t=1,2,..,T \quad \textit{(2.3.1.b)}$$

Finally, when μ is replaced by $\mu_0+\mu_1 t$, the equation *(2.3.1.b)*, can be generalized to include a trend component.

3.2.1.b The Trigonometric Representation

The deterministic seasonal process can be equivalently be written in terms of trigonometric functions:

$$Yt=\mu+\sum_{k=1}^{S/2}[\alpha_k \cos(\frac{2\pi kt}{S})+\beta_k \sin(\frac{2\pi kt}{S})]+ z_t, \quad t=1,2,..T \quad \textit{(2.3.1.c)}$$

This representation is equivalent to *(2.3.1.b)* and it is obvious that μ is the overall mean. For quarterly data where S=4, there are the following trigonometric components:

- $\cos(\frac{2\pi t}{4})=\cos(\frac{\pi t}{2})$= 0,-1,0,1…
- $\cos(\frac{4\pi t}{4})=\cos(\pi t)$= -1,+1,-1,…
- $\sin(\frac{2\pi t}{4})=\sin(\frac{\pi t}{2})$= 1,0,-1,0…
- $\sin(\frac{4\pi t}{4})=0$

According to the two representations described above, the seasonal dummy variable coefficients are related to the deterministic components of the trigonometric representation by the following equations:

- $\gamma_1= \mu+\beta_1-\alpha_2$
- $\gamma_2=\mu-\alpha_1+\alpha_2$
- $\gamma_3=\mu-\beta_1-\alpha_2$
- $\gamma_4=\mu+\alpha_1+\alpha_2$

The terms α_1 and β_1 denote the annual wave while α_2 gives the half year component.

3.2.2 The Stochastic Stationary Seasonality

The Stochastic Stationary Seasonality, refers to the roots at seasonal frequencies and it constitutes a whole different issue from deterministic seasonality. The deterministic seasonality never changes its shape, as it maintains a constant seasonal pattern. On the other hand, the stochastic stationary and (as we will explain in the next part of this chapter) the stochastic non-stationary seasonality display a random seasonal pattern from one cycle to the next.

At this point it is necessary to define the models that display this type of seasonality,as they were described in Peter J. Brockwell's and Richard A. Davis's book, " Time Series Theory and Methods", in order to understand the definition of the stochastic stationary seasonality in depth.

3.2.2.a The Seasonal ARMA and the mixed Seasonal ARMA model

Suppose we have r years of monthly data (S=12) tabulated below:

	Month			
Year	1	2	. . .	12
1	Y_1	Y_2	. . .	Y_{12}
2	Y_{13}	Y_{14}	. . .	Y_{24}
3	Y_{25}	Y_{26}	. . .	Y_{36}
.	.	.		.
.	.	.		.
.	.	.		.
r	$Y_{1+12(r-1)}$	$Y_{2+12(r-1)}$	. . .	$Y_{12+12(r-1)}$

Table 3.1- Monthly Data

Each column of this table is considered as a realization of time series. Suppose that each one of these time series is generated by the same ARMA(P,Q) models. The corresponding to the j^{th}month series Y_{j+12t} , t=0,…,r-1 satisfies a difference equation of the form:

$$Y_{j+12t} = \Phi_1 Y_{j+12(t-1)} + \ldots + \Phi_P Y_{j+12(t-P)} + U_{j+12t} + \Theta_1 U_{j+12(t-1)} + \ldots + \Theta_Q U_{j+12(t-Q)} , \quad (2.3.2.a.i)$$

where $\{U_{j+12t} , t = \ldots,-1,0,1,\ldots\} \sim WN(0, \sigma_U^2)$.

Since the same ARMA(P,Q) models is assumed to apply each month, the equation*(2.3.2.a.i)*can be written equivalently for all t as:

$$\boxed{\begin{gathered} Y_t = \Phi_1 Y_{t-12} + \ldots + \Phi_P Y_{t-12P} + U_t + \Theta_1 U_{t-12} + \ldots + \Theta_Q U_{t-12Q}, \quad (2.3.2.a.ii) \\ \text{for each } j=1,\ldots 12 \end{gathered}}$$

Equation *(2.3.2.a.ii)*can be rewritten in the form of:

$$\boxed{\boldsymbol{\Phi}(B^S)Y_t = \boldsymbol{\Theta}(B^S)U_t , \quad (2.3.2.a.iii)}$$

where $\mathbf{\Phi}(z) = 1- \Phi_1 z-\ldots- \Phi_P z^P$, $\mathbf{\Theta}(z) = 1 + \Theta_1 z+\ldots+ \Theta_Q z^Q$ and $\{U_{j+12t}, t = \ldots,-1,0,1,\ldots\}$ ~ WN(0, σ_U^2) for each j. The model *(2.3.2.a.iii)* is the *between-year model* or **the seasonal ARMA(P,Q)$_s$** model and P and Q denote the seasonal AR and MA orders.

If the White Noise sequences $\{U_{j+12t}, t = \ldots,-1,0,1,\ldots\}$ ~ WN(0, σ_U^2) for different months are uncorrelated with each other, then the columns itself are uncorrelated. However, it is unlikely that the 12 series corresponding to the different months are uncorrelated. To incorporate dependence between these series, we assume now that $\{U_t\}$~ARMA(p,q) model:

$$\boldsymbol{\varphi}(B)U_t = \boldsymbol{\theta}(B)Z_t, \ Z_t \sim WN(0,\sigma^2) \quad \textit{(2.3.2.a.iv)}$$

This non-zero correlation between the consecutive values of U_t implies a non-zero correlation within the twelve sequences of $\{U_{j+12t}, t = \ldots,-1,0,1,\ldots\}$.

Combining the two models *(2.3.2.a.iii)* and *(2.3.2.a.iv)*, lead us to the definition of the general multiplicative **mixed seasonal ARMA (p,q)x(P,Q)$_s$** process:

$$\boxed{\boldsymbol{\varphi}(B)\mathbf{\Phi}(B^S)Y_t = \boldsymbol{\theta}(B)\mathbf{\Theta}(B^S)Z_t, \ Z_t \sim WN(0,\sigma^2) \quad \textit{(2.3.2.a.v)}}$$

where $\boldsymbol{\varphi}(z)=1-\varphi_1 z-\ldots- \varphi_p z^p$, $\boldsymbol{\theta}(z)=1+\theta_1 z+\ldots+\theta_q z^q$, $\mathbf{\Phi}(z)=1-\Phi_1 z-\ldots-\Phi_P z^P$ and $\mathbf{\Theta}(z)=1+\Theta_1 z+\ldots+\Theta_Q z^Q$.

The seasonal ARMA(P,Q)$_S$ and mixed seasonal ARMA(p,q)x(P,Q)$_S$ models can be also extended in quarterly data (S=4).

<u>Examples</u>

(1) A very common example of a seasonal ARMA(P,Q)$_S$ model is the first-order *seasonal* Autoregressive. It is defined as:

$$Y_t = \Phi Y_{t-s} + Z_t \quad \textit{(2.3.2.a.vi)}$$

where Z_t~ WN(0,σ_Z^2). Using the lag operator $B^k Y_t=Y_{t-k}$, the equation *(2.3.2.a.vi)* can be written equivalently as:

$$(1-\Phi B^s)Y_t = Z_t \quad \textit{(2.3.2.a.vii)}$$

If $\Phi<1$ and therefore the roots of the polynomial $1-\Phi z^S$ lie outside the unit circle, the process *(2.3.2.a.vi)* is stationary.

The unconditional mean of the process is equal to zero. On the other hand, the conditional mean on past Y_t displays seasonal pattern if φis close to unity:

$$E(Y_t|Y_{t-1},\ldots)=\varphi Y_{t-s} \quad \textit{(2.3.2.a.viii)}$$

and the autocorrelation is different from zero at lags that are multiples of S only and this non-zero autocorrelation decays over time.

(2) A very common example of mixed seasonal ARMA(p,q)x$(P,Q)_s$ model is the ARMA(0,1)x$(1,0)_s$. Such model has the following form:

$$\mathbf{\Phi}(B^S)Y_t = \boldsymbol{\theta}(B)Z_t \Leftrightarrow (1-\Phi^s)Y_t = (1+\theta)\, Z_t \Leftrightarrow Y_t - \Phi Y_{t-S} = Z_t + \theta Z_{t-1}, \quad \textit{(2.3.2.a.ix)}$$

and $Z_t \sim WN(0,\sigma^2)$.

3.2.2.b –<u>The Roots of the polynomials φ(z) and Φ(z) and the relation with stochastic seasonality.</u>

In the mixed seasonal ARMA$(P,Q)_S$ model the stochastic stationary seasonality is more pronounced when the roots of the polynomial $\mathbf{\Phi}(z^S)$ are close to the unit circle. The roots of $\mathbf{\Phi}(z^S)$ however follow forcefully a certain structure:

Specifically, let z_0 be a root of the seasonal polynomial $\mathbf{\Phi}(z)$of a seasonal ARMA$(P,Q)_S$ model. Every z_0 root of $\mathbf{\Phi}(z)$induces S roots of $\mathbf{\Phi}(z^S)$.

If z_0 so that $\mathbf{\Phi}(z_0) = 0$ =>there exists $z_{0,1},z_{0,2},\ldots,z_{0,S}$, such that $(z_{0,k}^S) = z_0$ for k = 1,2,…,S.

Thus, for each root of $\mathbf{\Phi}(z)$ there is a group of S roots of $\mathbf{\Phi}(z^S)$ which all have the same modulus and their angles differ by $\frac{2\pi}{S}$.

<u>Example</u>

(1) Assume the model $Y_t = \Phi Y_{t-1}+ \varepsilon_t \Leftrightarrow (1-\Phi B^s)\, Y_t = \varepsilon_t$, $\varepsilon_t \sim WN(0,\sigma^2)$ and $\mathbf{\Phi}(z)=1-\Phi z^S$ is the polynomial of the seasonal part. Let $z_0 = \Phi^{-1}$be the root of $\mathbf{\Phi}(z)$ then setting $z_{0,k} = \Phi^{-\frac{1}{S}} e^{ik\frac{2\pi}{S}}$ we conclude that $(z_{0,k}^S) = z_0$. Therefore,$\mathbf{\Phi}(z_{0,k}{}^S) = \mathbf{\Phi}(z_0)$ and thus $z_{0,k}$, k=1,…,S are all roots of $\mathbf{\Phi}(z^S)$.

The table below shows the frequencies as well as the corresponding seasonal roots for monthly and quarterly series.

Frequency	**Roots**
Monthly Series	
0	$(\Phi^{-\frac{1}{12}})\cdot 1$
$\frac{\pi}{6}, \frac{11\pi}{6}$	$(\Phi^{-\frac{1}{12}})\cdot[\frac{1}{2}(\sqrt{3} \pm i)]$
$\frac{\pi}{3}, \frac{5\pi}{3}$	$(\Phi^{-\frac{1}{12}})\cdot[\frac{1}{2}(1 \pm \sqrt{3}i)]$
$\frac{\pi}{2}, \frac{3\pi}{2}$	$(\Phi^{-\frac{1}{12}})\cdot(\pm i)$
$\frac{2\pi}{3}, \frac{4\pi}{3}$	$(\Phi^{-\frac{1}{12}})\cdot[-\frac{1}{2}(1 \pm \sqrt{3}i)]$
$\frac{5\pi}{6}, \frac{7\pi}{6}$	$(\Phi^{-\frac{1}{12}})\cdot[-\frac{1}{2}(\sqrt{3} \pm i)]$
π	$(\Phi^{-\frac{1}{12}})\cdot(-1)$
Quarterly Series	
0	$(\Phi^{-\frac{1}{4}})\cdot 1$
$\frac{\pi}{2}, \frac{3\pi}{2}$	$(\Phi^{-\frac{1}{4}})\cdot(\pm i)$
π	$(\Phi^{-\frac{1}{4}})\cdot(-1)$

Table 3.2- *Frequencies and Roots*

At this point it must be emphasized that **not** only the roots of the seasonal polynomial $\mathbf{\Phi}(z)$ are responsible for the stochastic stationary seasonality but also those of the non-seasonal polynomial $\boldsymbol{\varphi}(z)$. If a root $z_0=\rho e^{i\lambda}$ of $\boldsymbol{\varphi}(z)$ is close to the unit circle ($\rho\approx1$) then the series will exhibit stochastic seasonality of period $S=\frac{2\pi}{\lambda}$.The only difference to the roots of $\mathbf{\Phi}(z^S)$ is that the latter come in groups of S members with a specific structure. Therefore, an ARMA(p,q) or a simple AR(p) model can also produce stochastic stationary seasonality.

Examples

(1) Consider the AR(2) model:

$$Y_t = \mu + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varepsilon_t\ ,\ \varepsilon_t \sim WN(0,\sigma^2)$$
$$\Leftrightarrow \varphi(B) Y_t = \mu + \varepsilon_t$$

where $\boldsymbol{\varphi}(z) = 1- \varphi_1 z - \varphi_2 z$ is the corresponding characteristic polynomial.

Let z_0^{-1} the inverse root of the characteristic polynomial and $z_0^{-1}= \rho^{-1}(\cos(\lambda)+ i\sin(\lambda))$ its polar form. As ρ is denoted the radial coordinate and as λ the angular coordinate. Furthermore, we choose:

$$\lambda= \frac{2\pi}{d}, \text{where d is the period and } |z_0|=\rho\approx 1$$

The characteristic polynomial then can be written equivalently as:

$$\varphi(z) = (1- z_0^{-1}z)(1-\overline{z_0^{-1}}z) = 1- (z_0^{-1}+ \overline{z_0^{-1}})z + |z_0^{-1}|^2z^2$$

where $\varphi_1 = z_0^{-1}+ \overline{z_0^{-1}} = 2\mathrm{Re}(z_0^{-1}) = 2\rho^{-1}\cos(\lambda)$ and $\varphi_2 = - |z_0^{-1}|^2$.

Setting z_0^{-1} close to unity, the AR(2) model, produces stochastic stationary seasonality. In order to illustrate that, we simulated data of an AR(2) model with z_0^{-1} close to unity and we created the figures below.

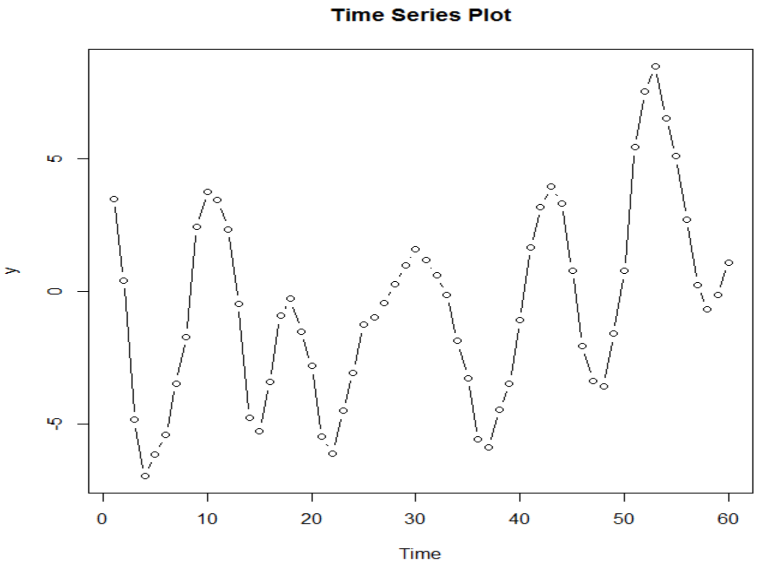


Figure 3.2*-Time series Plot*

In the Figure 3.2 it is enough evident that the series produces stochastic seasonality since its seasonal pattern is not deterministic and its changes over time.

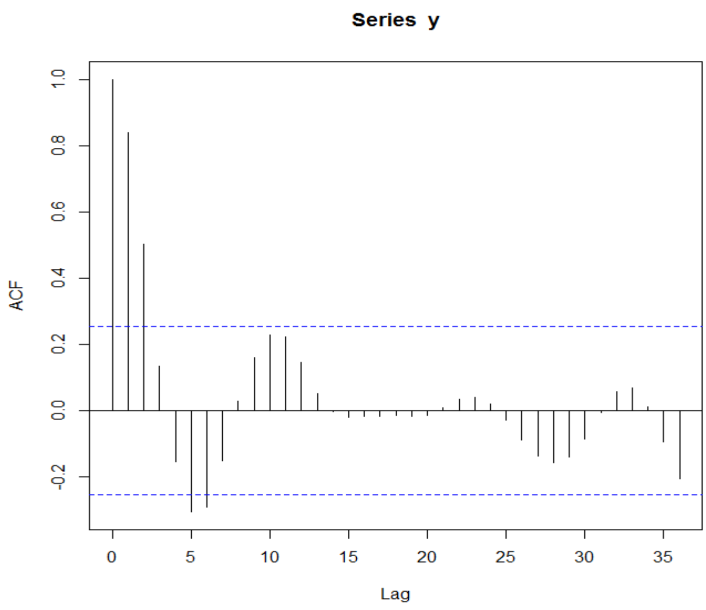


Figure 3.3-*Plot of the Autocorrelation Function*

Seasonality can be also be detected from the diagram of its autocorrelation function.

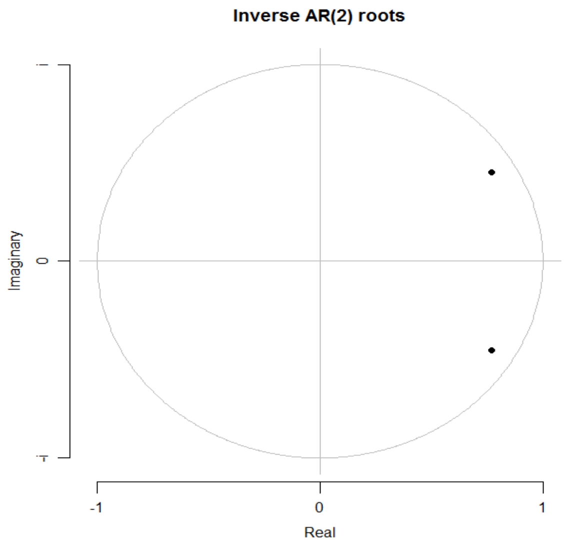


Figure 3.4-*The Inverse Roots*

The Figure 3.4 shows the inverse roots of the characteristic polynomial of the AR(2) that we simulated. We observe that both of them are on very close to the unit circle which proves the presence of stochastic stationary seasonality.

(2) Below are also depicted the inverse roots of the $\mathbf{\Phi}(z^S)$ and $\boldsymbol{\varphi}(z)$ polynomials of the mixed Seasonal ARMA$(1,0)\mathrm{x}(1,0)_{12}$ model:

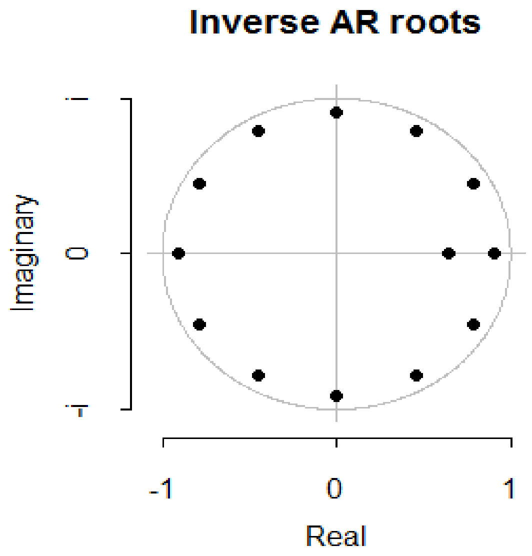


Figure 3.5-*The Inverse Roots*

We can see that all the seasonal roots are close to the unit circle.

3.2.3 The Stochastic Non-Stationary Seasonality

According to Ghysels, Osborn and Rodrigues (1999), the Nonstationary Stochastic Process Y_t, observed at S equally lengths per year, is said to be **Seasonally Integrated of order D**, denoted $Y_t \sim SI(D)$, if $\Delta_S^D Y_t = (1-B^S)^D Y_t$ is a stationary invertible[5] ARMA process.

The definition seasonal integration refers to the seasonal differencing of the process in order to induce stationarity when unit roots occur. Consequently, if a first order seasonal differencing makes Y_t a stationary and invertible process, then $Y_t \sim SI(1)$. The simplest case of such a process is the seasonal random walk.

We remark already at this point that **Seasonal Integrated** is not the only model producing non-stationary stochastic seasonality and this issue will be discussed further in 3.2.3.b.

[5]A linear process $\{X_t\}$ is invertible (an invertible function of $\{W_t\}$) if there is a $\pi(B)=\pi_0+\pi_1 B+\pi_2 B^2+\ldots$ with $\sum_{j=0}^{+\infty}|\pi_j|<+\infty$ and $W_t=\pi(B)X_t$.

The Seasonal Differencing

In order to achieve stationarity, seasonal differencing is applied to the process. Differencing the series $\{Y_t\}$ at lag S is a convenient way to eliminate a seasonal component of period S. Specifically, a seasonal operator of order 1 for S=12 acts on Y_t as: $(1-B^{12})Y_t = Y_t - Y_{t-12}$ and for S=4 as $(1-B^4)Y_t = Y_t - Y_{t-4}$. Moreover, a seasonal operator of order D is defined as: $(1-B^S)^D Y_t$. It is common though that D=1 is sufficient to obtain seasonal stationarity.

3.2.3.a The Seasonal ARIMA(p,d,q)x(P,D,Q)$_S$ models

At this point it is necessary to define the models that display this type of seasonality,(see Peter J. Brockwell's and Richard A. Davis's book, " Time Series Theory and Methods") in order to understand the definition of the stochastic non-stationary seasonality in depth. Combining the two models (2.3.2.α.iii) and (2.3.2.α.iv) from the previous part and allowing for differencing lead us to the definition of the general seasonal multiplicative SARIMA process.

Definition of SARIMA(p,d,q)x(P,D,Q)$_S$

If d and D are non-negative integers then X_t is said to be a seasonal ARIMA(p,d,q)x(P,D,Q)$_S$ process with period S if the differenced series $Y_t=(1-B)^d(1-B^S)^D X_t$ is a casual [6] ARMA process defined by:

$$\boldsymbol{\varphi}(B)\boldsymbol{\Phi}(B^s)Y_t=\boldsymbol{\theta}(B)\boldsymbol{\Theta}(B^s)Z_t,\ Z_t \sim WN(0,\sigma^2) \quad \textit{(2.3.3.a)}$$

where $\varphi(z)=1-\varphi_1 z-\ldots-\varphi_p z^p$, $\theta(z)=1+\theta_1 z+\ldots+\theta_q z^q$, $\Phi(z)=1-\Phi_1 z-\ldots-\Phi_P z^P$ and $\Theta(z)=1+\Theta_1 z+\ldots+\Theta_Q z^Q$..

Model Identification

Because of its nature, identifying a SARIMA models can be quite complicated. However, there is a general guideline for this identification. First of all, d and D should be found:

[6]An ARMA (p,q) process defined by the equation $\varphi(B)X_t=\theta(B)Z_t$ is said to be casual if there exists a sequence of constants $\{\psi_j\}$ such that $\sum_{j=0}^{+\infty} |\psi_j| < +\infty$ and $X_t = \sum_{j=0}^{+\infty} \psi_j Z_{t-j}$ t=0,±1,…

- If there is seasonality and no trend in the data, take a difference of lag S (seasonal differencing).
- If there is linear trend and no obvious seasonality, take a first difference (non-seasonal differencing).
- If there is both trend and seasonality apply both seasonal and non-seasonal difference to the data.

Obviously, if d=D=0 our process displays stochastic stationary seasonality. On the other hand if d, D $\neq$ 0 we have to deal with non-stationary stochastic seasonality and we have to specify these parameters so as to make the differenced observations $Y_t=(1-B)^d(1-B^S)^D$ stationary.

After that, it follows the examination of the sample autocorrelation and partial autocorrelation functions of $\{Y_t\}$ at lags which are multiples of S in order to find the orders P and Q. The orders P and Q should be chosen so that the autocorrelation function $\hat{p}$(ks), k=1,2,.. is compatible with the autocorrelation function of an ARMA(P,Q) process (identification of the seasonal terms), while the orders p and q are chosen so that $\hat{p}(1),\ldots,\hat{p}(s-1)$ match with the autocorrelation function of an ARMA(p,q) process (identification of the non-seasonal terms).

3.2.3.b –The Roots of the polynomials φ(z) and Φ(z) and the relation with stochastic non-stationary seasonality.

For a SARIMA model of the form *(2.3.3.a)*,once a seasonal root of $\mathbf{\Phi}(z^S)$ lies very close or on the unit circle, the rest S-1 roots lie very close/on the unit circle as well (stationary/non-stationary stochastic seasonality). Correspondingly, if a root lies outside the unit circle the rest S-1 will lie there too (stationarity). This is because all these roots have the same modulus and their angles differ by multiples a$t\,\frac{2\pi}{S}$. Therefore, for SARIMA models when stochastic non-stationary seasonality is detected, **a whole group of roots** of the seasonal characteristic polynomial $\mathbf{\Phi}(z^S)$ are considered to be on the unit circle.

However, stochastic non-stationary seasonality can be produced even a single root of the **φ**(z) polynomial lie on the unit circle. Specifically, any root of **φ**(z) equal to unity in absolute value (zero frequency) “contributes” to the production of stochastic non-stationary seasonality.

Therefore, a **SARIMA(p,d,q)x(P,D,Q)$_S$** model:

- With all groups of roots of its seasonal polynomial $\mathbf{\Phi}(z^S)$ lie outside the unit circle and all roots of the non-seasonal polynomial $\boldsymbol{\varphi}(z)$ lie outside the unit circle, is considered stationary.
- With any roots of its seasonal polynomial $\mathbf{\Phi}(z^S)$ lie on the unit circle or any root of the non-seasonal polynomial $\boldsymbol{\varphi}(z)$ lies on the unit circle is considered to produce stochastic non-stationary seasonality.

Summarizing the above details, in the seasonal ARMA(P,Q)$_S$ and the mixed seasonal ARMA(p,q)x(P,Q)$_S$ the stochastic non-stationary seasonality occurs when either groups of roots of the polynomial $\mathbf{\Phi}(z^S)$ are on the unit circle or any of the independent roots of $\boldsymbol{\varphi}(z)$ are on the unit circle.

Consider now the general Seasonal ARIMA(p,d,q)x(P,D,Q)$_S$ model in equation *(2.3.3.a)*:

$$\boldsymbol{\varphi}(B)\mathbf{\Phi}(B^s)(1-B)^d(1-B^S)^D X_t=\boldsymbol{\theta}(B)\mathbf{\Theta}(B^s)z_t,\ z_t\sim WN(0,\sigma^2)$$

This model can be written alternatively as:

$$\boxed{\ddot{\boldsymbol{\Phi}}(B)X_t=\ddot{\boldsymbol{\Theta}}(B)z_t,\ z_t\sim WN(0,\sigma^2)\quad(2.3.3.b)}$$

with $\ddot{\boldsymbol{\Phi}}(z)=\boldsymbol{\varphi}(z)\mathbf{\Phi}(z^s)(1-z)^d(1-z^S)^D$ and $\ddot{\boldsymbol{\Theta}}(B)=\boldsymbol{\theta}(z)\mathbf{\Theta}(z^s)$.

Seasonal Integration occurs when $\ddot{\boldsymbol{\Phi}}(z)$ has a group of roots of multiplicity D $z_{0,k}$ = $e^{\frac{i2\pi}{S}}$, k=1,…,S at the seasonal frequencies. However, any (other) root of $\ddot{\boldsymbol{\Phi}}(z)$ on the unit circle will result in a non-stationary X_t and will produce seasonal behavior of X_t with period S. We will speak in this case as well as non-stationary seasonality in the next chapters.

Generallizing the aforementioned inference, let the ARMA(p,q) modelthat satisfies the equation below:

$$\boxed{\widetilde{\ddot{\boldsymbol{\Phi}}}(B)X_t=\widetilde{\ddot{\boldsymbol{\Theta}}(B)}z_t,\ z_t\sim WN(0,\sigma^2)\quad(2.3.3.c)}$$

with $\widetilde{\ddot{\boldsymbol{\Phi}}}(z)$ and $\widetilde{\ddot{\boldsymbol{\Theta}}(z)}$some other polynomials.

In this general case of an ARMA(p,q) model, any root of $\widetilde{\ddot{\boldsymbol{\Phi}}}(z)$ on the unit circle will also result in a non-stationary X_t.

3.2.3.c –Examples of SARIMA models

(1) The Seasonal Random Walk

The seasonal random walk of order 1 is defined by the following equation:

$$Y_t=Y_{t-S} + \varepsilon_t \text{ , } t=1,2,\dots,T \quad \mathit{(2.3.3.b)}$$

with $\varepsilon_t \sim WN(0,\sigma^2)$.

It is obvious that the equation *(2.3.3.a)* is the generalization of the conventional nonseasonal random walk. Setting s_t the season in which observation t falls as $s_t= 1+(t-1)\text{mod}S$, backward substitution for lagged Y_t in the process above, implies that

$$Y_t = Y_{s_{t-S}}+\sum_{j=0}^{n_t -1} \varepsilon_{t-S_j} \quad \mathit{(2.3.3.c)}$$

where $n_t= 1+ [(t-1)/S]$.

There random walk described above, contains the disturbances for the season $s_t$with the summation over the current disturbance ε_t plus the disturbance for this season in the n_t-1 previous years of the observation period. Also, the equation *(2.3.3.c)* implies that $E(Y_t)=E(Y_{s_{t-S}})$, so when $E(Y_{s_{t-S}})$ is nonzero and varies over s_t=1,…S, deterministic seasonal effects are included in the equation *(2.3.3.b).*

In time series analysis, the common notation for a process is Y_t, where Y_t is the value of the variable we interested in at the date t. However, for time series that display seasonality the double subscript notation $Y_{s\tau}$ is being used. The subscript s denotes the season of the year, s=1,2,..,S and S is the number of season per year (S=4 for quarterly data and S=12 for monthly data). The subscript τ is obvious that it refers to the year. Hence, if s=4 then $Y_{s-4,\tau}=Y_{s,\tau-4}$.

Moreover, using the notation of the two subscripts described in the start of this chapter the equation *(2.3.3.b)* can be written equivalently as:

$$Y_{sn}=Y_{s,0}\_ \sum_{j=1}^{n} \varepsilon_{sj}, \text{ s=1,…S and n=1,…,N} \quad \mathit{(2.3.3.d)}$$

with the assumption that observations are available for precisely N=T/S complete years.

(2) The model SARIMA(0,0,1)x(0,0,1)$_{12}$ includes non-seasonal MA(1) and a seasonal MA(1) term, no differencing terms, no AR terms and its span seasonality is S=12. The non-seasonal MA(1) polynomial is the $\theta(B)=1+\theta_1 B$ while the seasonal MA(1) polynomial is $\Theta(B^{12})=1+\Theta_1 B^{12}$. Since there is no differencing term the model will have intercept and the equation will be:

$$(X_t-\mu)= \Theta(B^{12})\ \theta(B)z_t\ ,\ z_t\sim WN(0,\sigma^2)$$

which is equivalent to $X_t-\mu=(1+\theta_1 B)(1+\Theta_1 B^{12})z_t \Leftrightarrow X_t-\mu=(1+\Theta_1 B^{12} + \theta_1 B + \theta_1\Theta_1 B^{13})z_t$. Thus the true model has MA terms at lags 1,12 and 13.

(3) The model SARIMA(1,0,0)x(1,0,0)$_{12}$ includes non-seasonal AR(1) and a seasonal AR(1) term, no differencing and MA terms and S=12. The model is $(1-\Phi_1 B^{12})(1-\varphi_1 B)(X_t-\mu)=z_t$, $z_t\sim WN(0,\sigma^2)$. Let $w_t= X_t-\mu$ then $w_t= \varphi_1 w_{t-1}+\Phi_1 w_{t-12} - \varphi_1\Phi_1\ w_{t-13} + z_t$. This is an AR model with predictors at lags 1,12 and 13.

(4) The model SARIMA(0,1,1)x(0,1,1)$_{12}$ includes non-seasonal MA(1),seasonal MA(1), the differencing terms d=1 and D=1, no AR terms and its span seasonality is S=12. Since there are differencing terms, the model will not have an intercept and the equation will be:

$$(1-B)(1-B^{12})X_t = (1+\theta_1 B)(1+\Theta_1 B^{12})z_t,\ z_t\sim WN(0,\sigma^2)$$
$$\Leftrightarrow X_t= X_{t-12} - X_{t-1} - X_{t-13} + (1+\theta_1 B)(1+\Theta_1 B^{12})z_t..$$

(5) Below are illustrated the inverse roots of the polynomials $\boldsymbol{\Phi}(z^{S)}$ and $\boldsymbol{\varphi}(z)$ of the SARIMA(1,0,1)x(1,1,1)$_{12}$. We can see the seasonal inverse roots lie on the unit circle.

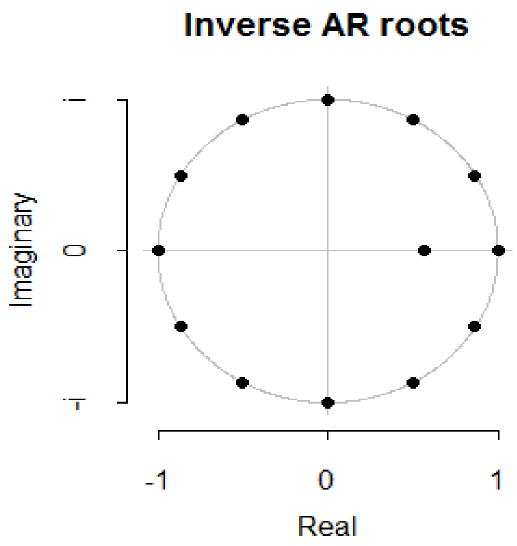


Figure 3.6-*The Inverse Roots*

3.2.3 "Assuming the correct vs the wrong type of seasonality": An illustration example

In this part, we illustrate the results of modelling three different types of data generating processes, fitting on the one side the dummy variable representation and on the other side the Seasonal ARIMA models. Specifically, we simulated data that follow the models:

- $Y_t=Y_{t-S}+\varepsilon_t$
- $Y_t=\mu+\sum_{s=1}^{S} m_s\delta_{st}+\varepsilon_t$
- $Y_t=\mu+Y_{t-S}+\sum_{s=1}^{S} m_s\delta_{st}+\varepsilon_t$

Where $\varepsilon_t \sim WN(0,\sigma^2)$.

Then, we estimated each one of them with the models:

- $Y_t=\mu+\alpha_S Y_{t-S}+\varepsilon_t$ *(2.3.4.a)*
- $Y_t=\mu+\sum_{s=1}^{S} m_s\delta_{st}+\varepsilon_t$*(2.3.4.b)*
- $Y_t=\mu+\alpha_S Y_{t-S}+\sum_{s=1}^{S} m_s\delta_{st}+\varepsilon_t$ *(2.3.4.c)*

The purpose of this test is to note all the possible consequences of modelling a process that displays deterministic seasonality with SARIMA models and vice versa, the consequences of modelling a process that displays stochastic (stationary of non-stationary) seasonality with the dummy variable representation (deterministic seasonality). The table below represents the results in the estimations, the residuals as well as the ACF and PACF plots of the residuals of modelling these three data generating processes with these three different types of models.

True Model			
	$Y_t = Y_{t-S} + \varepsilon_t$	$Y_t = \mu + \sum_{s=1}^{S} m_s \delta_{st} + \varepsilon_t$	$Y_t = \mu + Y_{t-S} + \sum_{s=1}^{S} m_s \delta_{st} + \varepsilon_t$
$Y_t = \mu + \alpha_S Y_{t-S} + \varepsilon_t$			
$Y_t = \mu + \sum_{s=1}^{S} m_s \delta_{st} + \varepsilon_t$	**The residuals are correlated.** <u>ACF plot</u>: We can see peaks at lags that are multiples of 6. <u>PACF plot</u>: Peaks at lags 1,2,3,6,7,8,9,10,11 and 12. After lag 12, the residuals are within the limits and they finally wear off.	The ACF and PACF plot shows that the residuals are almost white noise.	The residuals are correlated. <u>ACF plot</u>: We can see peaks at various lags. <u>PACF plot</u>: Peaks at lags 1,2,3,6,7,8,9,10,11 and 12. After lag 12, the residuals are within the limits and they finally wear off.
$Y_t = \mu + \alpha_S Y_{t-S} + \sum_{s=1}^{S} m_s \delta_{st} + \varepsilon_t$	The estimation of α is very close to unity and the ACF and PACF plots indicate that $\varepsilon_t \sim WN(0,\sigma^2)$. The estimation of dummy coefficients are significant.	The ACF and PACF plot shows that the residuals are almost white noise.	The estimation of α is very close to unity and the ACF and PACF plots indicate that $\varepsilon_t \sim WN(0,\sigma^2)$

***Table 3.3**- Results of the Modelling*

We can see from the table above that modelling a seasonal random walk $Y_t = Y_{t-S} + \varepsilon_t$ (non-stationary stochastic seasonality) as a process that displays deterministic seasonality $Y_t = \mu + \sum_{s=1}^{S} m_s \delta_{st} + \varepsilon_t$, are unable to obtain White Noise residuals. Specifically, the corresponding plots shown in figure 3.7 illustrate peaks at lags multiples of 6 in ACF plot and various peaks in PACF plot and therefore the residuals are uncorrelated.

Furthermore, modelling the process $Y_t=\mu+Y_{t-S}+\sum_{s=1}^{S} m_s\delta_{st}+\varepsilon_t$ as a model that displays deterministic seasonality we are also unable to obtain White Noise residuals. The corresponding results of the ACF and PACF plots are almost the same as those of figure 3.4

The code and the corresponding output in R, as well as the ACF and PACF plots are listed in the Appendix (B. Code - **Code, plotsand output in R of the Example)**

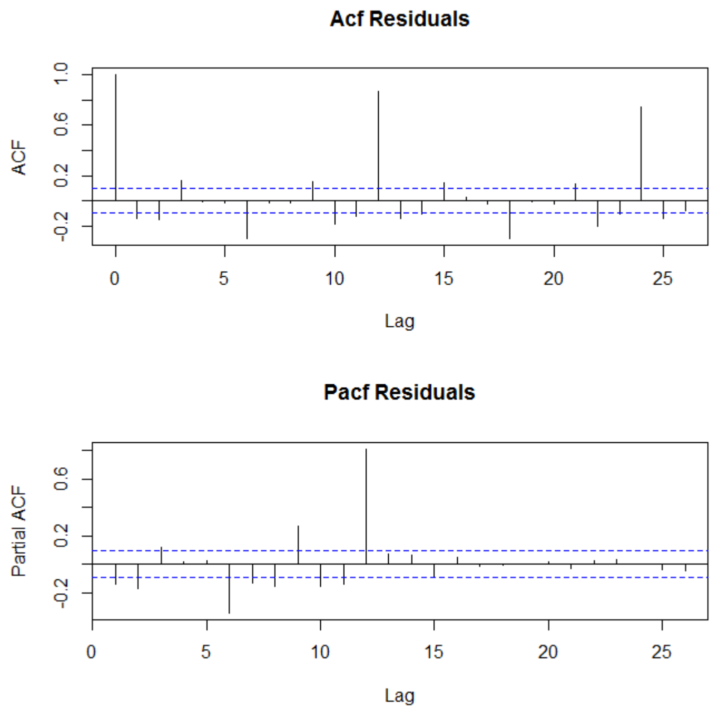


Figure 3.7*-The ACF and PACF plots of the Residuals*

CHAPTER 4

The Seasonal Dickey-Fuller and Augmented Dickey-Fuller Unit Root test

Some of the most popular seasonal Unit Root tests, are the seasonal Dickey-Fuller test (seasonal DF test) and the seasonal Augmented Dickey-Fuller test. Both of them will be described in this part.

4.1 The seasonal Dickey-Fuller Unit Root test

First of all, assume the **Seasonally Integrated model of order D** model (SI(D) model) described in part 3.2.3 in Chapter 3. In the framework of this model the hypotheses of the seasonal DF test are formed as follows:

H_0: $Y_t \sim SI(D) \Leftrightarrow \Delta_S^D Y_t = (1-B^S)^D \cdot Y_t$ is a stationary process
H_1: Y_t is stationary

More specifically, let the model $Y_t = \alpha_S \cdot Y_{t-S} + \varepsilon_t$, $\varepsilon_t \sim WN(0,\sigma^2)$ *(3.1.1.α)* which is the model *(2.3.4.a)* of the previous part with mean equal to zero. The corresponding hypothesis of the seasonal DF test for the model *(3.1.1.α)* are the following:

H_0: $\alpha_S = 1$
H_1: $\alpha_S < 1$

Therefore, in general under the null hypothesis of the seasonal DF test the Autoregressive polynomials of a model contains all the roots 1, ±i for quarterly data and all the roots ±1, ±i, $\frac{1}{2}(\sqrt{3} \pm i)$, $\frac{1}{2}(1 \pm \sqrt{3}i)$, $-\frac{1}{2}(1 \pm \sqrt{3}i)$, $-\frac{1}{2}(\sqrt{3} \pm i)$ for monthly data. In other words under the null hypothesis all roots of the of the Autoregressive polynomials are on the unit circle and under the alternative all roots of polynomial have the same modulus smaller than unity.
In order to describe the seasonal DF test, we will use the zero mean and the seasonal means model, as well as the single mean model, defined by the following equations:

- $Y_t = \alpha_S \cdot Y_{t-S} + \varepsilon_t$, $\varepsilon_t \sim WN(0,\sigma^2)$, the Zero Mean Model *(3.1.1.α)*

- $Y_t = \alpha_S Y_{t-S} + \sum_{s=1}^{S} \gamma_s \delta_{st} + \varepsilon_t$, $\varepsilon_t \sim WN(0,\sigma^2)$, the Seasonal Means Model *(3.1.1.b)*

- $Y_t = \mu + \alpha_S Y_{t-S} + \varepsilon_t$, $\varepsilon_t \sim WN(0,\sigma^2)$, the Single Mean Model *(3.1.1.c)*

For these three models Dickey, Hasza and Fuller (1984), described the original least squares estimators of their coefficients as well as the corresponding Studentized regression statistics. For the *Zero Mean Model* the ordinary least square estimator of $\alpha_S$is the $\hat{\alpha}_S = \frac{\sum_{i=1}^{n} Y_{t-S} \cdot Y_t}{\sum_{i=1}^{n} Y_{t-S}^2}$ and the Studentized statistic is the $\hat{\tau}_S = [(\sum_{t=1}^{n} Y_{t-S}^2)^{-1} \cdot \widehat{\sigma_r}^2]^{-\frac{1}{2}} \cdot (\hat{\alpha}_S - 1)$ where $\widehat{\sigma_r}^2 = (n-1)^{-1} \cdot \sum_{t=1}^{n} (Y_t - \hat{\alpha}_S \cdot Y_{t-S})^2$. For the *Seasonal Means Model*, regressing Y_t on $\delta_{1t}, \delta_{2t}, \ldots, \delta_{St}$, Y_{t-S} for t=1,2,…,n, yields the coefficients $\widehat{\gamma_1}, \widehat{\gamma_2}, \ldots, \widehat{\gamma_S}$, and $\widehat{\alpha_{\mu S}}$, as well as the Studentized statistic $\hat{\tau}_{\mu S}$. Finally, for the *Single Mean Model* the results are similar to the previous and μ and $\widehat{\alpha_S^*}$ denote the estimated coefficients of the model and $\widehat{\tau_S^*}$ is the corresponding Studentized statistic.

Extending the approach of Dickey (1976) who computed the percentiles of the limit distributions of $n \cdot (\hat{\alpha}_1 - 1)$, $\hat{\tau}_1$ and $\hat{\tau}_{\mu 1}$, to the models with S>1, Dickey, Hasza and Fuller obtained the limit percentiles of the limit distributions of the statistics described above. (see Theorem 1- Dickey, Hasza and Fuller, 1984). Monte Carlo integration and other techniques were used for the computation of the percentiles of distributions for time series that all of its roots are on the unit circle (under the null hypothesis). The provided tabled distributions are used to test the hypothesis listed above.

Estimated Equation	5% Percentage points of the OLS estimator		5% Percentage points of the Studentized Statistic	
		S=4 -9.16		S=4 -1.90
		S=12 -11.58		S=12 -1.80
		S=4 -27.88		S=4 -4.04
		S=12 -59.45		S=12 -5.82
		S=4 -12.62		S=4 -2.38
		S=12 -13.65		S=12 -2.06

Table 4.1*-The 5% percentage points of the estimators for* $T=+\infty$

These conclusions as well as the Dickey Fuller test, concern the three aforementioned models under the assumption of ε_t's to be White Noise.

4.2 <u>The seasonal Augmented Dickey-Fuller Unit Root test</u>

In the presence of the serial correlation of the residuals, the extension of the seasonal Dickey-Fuller Unit root test is the seasonal Augmented Dickey-Fuller Unit root test (seasonal ADF test).

Let the multiplicative model:

$$(1-\alpha_S B^S)\cdot(1-\theta_1 B-\ldots-\theta_p B^P)\cdot Y_t = \varepsilon_t \quad \textit{(3.2.1.α)}$$

where ε_t is a sequence of iid $(0,\sigma^2)$ random variables. The equation *(3.2.1.α)* is the SARIMA$(p,0,0)x(1,0,0)_S$ model with mean equal to zero and it defines the errors ε_t as a nonlinear function of $(\alpha_S,\boldsymbol{\theta})$, where $\boldsymbol{\theta}'=(\theta_1,\theta_2,\ldots,\theta_p)$.

The estimator $\hat{\alpha}_{S}$**-1** will arise from a procedure named the two-step regression and it can be used to test the hypothesis H_0: $\alpha_S=1$ versus H_1:$\alpha_S\neq1$.

The two-step regression procedure

Note that in the framework of equation *(3.2.1.α)* and satisfy $\dot{Y_t}$= Y_t-Y_{t-S} corresponds to the initial estimate that α_S=1. This suggests the following procedure:

(1) Regress $\dot{Y_t}$ on $\dot{Y}_{t-1}$,$\dot{Y}_{t-2}$,…,$\dot{Y}_{t-p}$to obtain an initial estimator of $\boldsymbol{\theta}$ **=** **(**θ_1,θ_2,…,θ_p) that is consistent for$\boldsymbol{\theta}$ under the null hypothesis that α_S=1.

(2) Compute the residuals $\varepsilon_t(1,\widehat{\boldsymbol{\theta}})$ and regress $\varepsilon_t(1,\widehat{\boldsymbol{\theta}})$ on [(1-$\widehat{\theta 1}$B-…..-$\widehat{\theta p}$$B^P$)·$Y_{t-S}$, $\dot{Y}_{t-1}$,$\dot{Y}_{t-2}$,…,$\dot{Y}_{t-p}$] to obtain the estimators (α_S-1,$\boldsymbol{\theta}$**-**$\widehat{\boldsymbol{\theta}}$).

As it is subsequently explained in a Theorem 5 (see Dickey, Hasza, Fuller-1984) the limit percentiles obtained for the first order models can be extended to the multiplicative model. Specifically, if α_S=1 in model *(3.2.1.α)*, the two-step regression procedure results in an estimator $\widehat{\boldsymbol{\alpha}}_s$ and a corresponding Studentized statistic with the same limit distribution as that of the statistic one would obtain by regressing Z_t-Z_{t-S}=Z_t on Z_{t-S} where Z_t=Y_t-$\theta_1 Y_{t-1}$-….- $\theta_p Y_{t-p}$. The estimators θ_i, obtained by adding the estimates of $\boldsymbol{\theta_i}$**-**$\widehat{\boldsymbol{\theta_\iota}}$ to$\widehat{\boldsymbol{\theta_\iota}}$ have the same asymptotic distribution as the coefficients in a regression of $\dot{Y_t}$ on $\dot{Y}_{t-1}$,$\dot{Y}_{t-2}$,…,$\dot{Y}_{t-p}$.

The Theorem 5 implies that the tabulated limit percentiles of estimators in the Zero Mean Model are also applicable in the multiplicative model for large sample sizes. Therefore, the estimator (α_S-1) and the corresponding Studentized statistic will have the percentiles with those of the Zero Mean Model.

As far as the seasonal means and the single mean models are concerned, the extension of the theorem is immediate. Specifically, let

$$y_t = Y_t - \sum_{i=1}^{S} \delta_{it}\, \widetilde{\mu_\iota} \quad (3.2.1.b)$$

Replacing Y_t by y_t in the two-step regression procedure results in the regression of the errors $\varepsilon_t(1,\widehat{\boldsymbol{\theta}})$ on [(1-$\widehat{\theta 1}$B-…..-$\widehat{\theta p}$$B^P$)·$y_{t-S}$, $\dot{Y}_{t-1}$,$\dot{Y}_{t-2}$,…,$\dot{Y}_{t-p}$].

Using these arguments, it follows that the first coefficient $\widehat{a_{\mu S}}$and its Studentized statistic converge to the limit distribution of the corresponding estimators of the Zero Mean model.

CHAPTER 5

The HEGY Unit Root test

In the previous part, it was described the seasonal unit root test by Dickey, Hasza and Fuller (1984). Specifically, it was listed the asymptotic distribution of the least-square estimators for three different models and these results were extended to the case of high-orders stationary dynamics. A crucial disadvantage of the seasonal ADF unit root test is that the null hypothesis implies that all roots of $\boldsymbol{\Phi}(B^S)$ are on the unit circle while the alternative restricts the roots to have the same modulus.

5.1 <u>The HEGY Unit Root test for quarterly data (S=4)</u>

To begin with, we will assume the general autoregressive model

$$\boldsymbol{\varphi}(B)(Y_t-\mu_t) = \varepsilon_t \; \textit{(4.1.1.a)}$$

where $\boldsymbol{\varphi}(B)$ is the autoregressive polynomial, $\mu_t = \mu + \sum_{s=1}^{4} m_s \delta_{st}$ and $\varepsilon_t \sim WN(0,\sigma^2)$.

The equation *(4.1.1.a)* can be written equivalently as:

$$\mathbf{d}(B)\mathbf{a}(B)(Y_t-\mu_t) = \varepsilon_t \; \textit{(4.1.1.b)}$$

where all the roots of $\mathbf{d}(z) = 0$ lie on the unit circle and therefore its roots $\theta_k \in \{+1,-1,+i,-i\}$ for $k \leq 4$ since we consider quarterly seasonality only. All the roots of $\mathbf{a}(z) = 0$ are assumed to lie outside the unit circle and therefore all the stationary components are absorbed into $\mathbf{a}(B)$ and finally μ_t describes the deterministic seasonality when there are no seasonal unit roots in $\mathbf{d}(B)$.

In practice, unit roots may be present at some, but not at all the frequencies. Therefore, a joint test at all seasonal frequencies (seasonal ADF test) simultaneously will not provide the appropriate result. The benefit of HEGY unit root test, is that it can look for unit roots at any single seasonal frequency (as well as the zero frequency) without imposing roots at other frequencies.

Therefore, HEGY unit root test allows to test for individual roots. Specifically, they are tested the following separate hypotheses:

- H_0: θ_1 =1 is root of the polynomial **d**(z) =0
  H_1: θ_1 =1 is not root of the polynomial **d**(z) =0

- H_0: θ_2 =-1 is root of the polynomial **d**(z) =0
  H_1: θ_2 =-1 is not root of the polynomial **d**(z) =0

- H_0: θ_3=i and θ_4 =-i are both roots of the polynomial **d**(z) =0
  H_1: θ_3=i and θ_4 =-i are not roots of the polynomial **d**(z) =0

We will show subsequently that the equation *(4.1.1.a)* can be also transformed to the form

$$\boldsymbol{\varphi}^*(\boldsymbol{B})y_{4t} = \pi_1 y_{1t-1} + \pi_2 y_{2t-1} + \pi_3 y_{3t-2} + \pi_4 y_{3t-1} + \varepsilon_t \quad (4.1.1.c)$$

where,

$$\boldsymbol{\varphi}^*(\boldsymbol{B}) = 1-\varphi_1 B-\varphi_2 B^2-\ldots-\varphi_p B^p \text{ and } \varepsilon_t \text{ is white noise}$$

and

$$\begin{aligned} y_{1t} &= (1+B+B^2+B^3)Y_t \\ y_{2t} &= -(1-B+B^2-B^3)Y_t \qquad (4.1.1.d) \\ y_{3t} &= -(1-B^2)Y_t \\ y_{4t} &= (1-B^4)Y_t \end{aligned}$$

and these y's are asymptotically uncorrelated.

Moreover, it will be shown that $\pi_k = 0$ describes exactly the hypothesis that θ_k is a root of **φ**(B). Thus, in order to conduct the test one will compute $y_{4t} = Y_t - Y_{t-4}$ and will then estimate the equation:

$$y_{4t} = \varphi_1 y_{4t-1} + \cdots + \varphi_p y_{4t-p} + \pi_1 y_{1t-1} + \pi_2 y_{2t-1} + \pi_3 y_{3t-2} + \pi_4 y_{3t-1} + \varepsilon_t$$

with OLS to obtain the estimates $\widehat{\varphi_1}$, $\widehat{\varphi_2}$,…,$\widehat{\varphi_p}$, $\widehat{\pi_1}$, $\widehat{\pi_2}$, $\widehat{\pi_3}$,$\widehat{\pi_4}$. Here p has been selected by criteria such as AIC, BIC e.t.c.

In order to test the hypothesis that **φ(**θ_k**)** = 0 where θ_k is either 1,-1, ±i, it is simply tested that π_k is zero. Specifically, a test that 1 is a root of *(4.1.1.c)* is a test for $\pi_1 = 0$ and correspondingly, a test for -1 is a test for $\pi_2 = 0$. For the complex roots iand -i it is suggested a joint test that π_3 and $\pi_4$are equal to zero. It is evident that a series has no unit roots if each one of the π's is different from zero.

Therefore, in the framework of the equation *(4.1.1.c)*, in order to look for unit roots at the zero frequency as well as the others seasonal frequencies, the following hypothesis are being tested:

1) H_0: $\pi_1 = 0$, 1 is a root of the polynomial
H_1: $\pi_1<0$, 1 is not a root of the polynomial

2) H_0: $\pi_2 = 0$, -1 is a root of the polynomial
H_1: $\pi_2<0$, -1 is not a root of the polynomial

3) H_0: $\pi_3 = \pi_4 = 0$, $\pm i$ are roots of the polynomial
H_1: $\pi_3 \neq 0$ and $\pi_4 \neq 0$, $\pm i$ are not roots of the polynomial

The π's as well as the φ_i's are estimated with ordinary least squares. Hylleberg, Engle, Granger and Yoo (1990) studied the asymptotic distribution of the appropriate t and F statistics and computed by Monte Carlo integration the critical values for the one-sided 't' tests on π_1 and π_2 as well as the critical values for the 'F' test on $\pi_3 \cap \pi_4 = 0$. The first ones are very close to the Monte Carlo values from Dickey-Fuller and Dickey-Hasza-Fuller for the situations in which they tabulated the statistics. Below are tabulated the 5% percentage points of the π_1, π_2 and π_3 and the 95% percentage points for $\pi_3 \cap \pi_4$ for T=200 of five different models.

Estimated Equation	**'t': π_1**	**'t': π_2**
No intercept *No seas. Dummies* *No trend*	-1.94	-1.95
Intercept *No seas. Dummies* *No trend*	-2.87	-1.92
Intercept *Seas. Dummies* *No trend*	-2.91	-2.89
Intercept *No seas. Dummies* *Trend*	-3.44	-1.95
Intercept *Seas. Dummies* *Trend*	-3.49	-2.91

Table 5.1*-The 5% percentage points of the estimators for T=200*

Estimated Equation	**'F': $\pi_3 \cap \pi_4$**
No intercept *No seas. Dummies* *No trend*	3.16
Intercept *No seas. Dummies* *No trend*	3.12
Intercept *Seas. Dummies* *No trend*	6.61
Intercept *No seas. Dummies* *Trend*	3.07
Intercept *Seas. Dummies* *Trend*	6.57

Table 5.2*-The 95% percentage points of the estimator for T=200*

Some notes about the aforementioned critical values is that:

1) If the π's are truly different from zero then the models has no unit roots at theses frequencies and the corresponding y's are stationary. As a result, the regression is equivalent to a standard augmented unit-root test. If some of the π's are zero,

the distribution of the others test statistics will not be affected since the y's are asymptotically uncorrelated. For example, the test for $\pi_1=0$ will have the same limiting distribution regardless of the presence of y_2 in the regression.

2) If deterministic components are added in the regression the distributions of the π's change. The intercept and a trend term affect only the distribution of π_1 while the intercept in combination with the three seasonal dummies influences the distribution of π_2, π_3 and π_4.

5.2 <u>The derivation of HEGY Unit Root test for quarterly data</u>

First of all we will list a Lagrange's proposition that is useful in the description of the HEGY test:

"Any (possibly infinite or rational) polynomial $\varphi(B)$, which is finite-valued at the distinct, non-zero, possibly complex points $\theta_1, \theta_2, \ldots, \theta_p$, can be expressed in terms of elementary polynomials and a remainder as follows:

$$\boldsymbol{\varphi}(\mathrm{B}) = \sum_{k=1}^{K} \lambda_k \Delta(B)/\delta_k(B) + \Delta(\mathrm{B})\boldsymbol{\varphi}^{**}(\mathrm{B}), \quad (4.2.1.\alpha)$$

where the λ_k are a set of constants, $\varphi^{**}(\mathrm{B})$ is a (possible infinite or rational) polynomial and

$$\delta_k(B) = 1\text{-}\frac{1}{\theta_k}\mathrm{B}, \quad \Delta(\mathrm{B}) = \prod_{k=1}^{K} \delta_k(B) \text{ "}$$

In the above, λ_k are defined as:

$$\lambda_k = \varphi(\theta_k) / \prod_{j \neq k} \delta_j(\theta_k), \text{ which always exists.}$$

By adding and subtracting $\Delta(\mathrm{B})\sum \lambda_k$ to *(4.2.1.α)* we get the following form:

$$\boldsymbol{\varphi}(\mathrm{B}) = \sum_{k=1}^{K} \lambda_k \Delta(B)(1-\delta_k(B))/\delta_k(B) + \Delta(\mathrm{B})\boldsymbol{\varphi}^{*}(\mathrm{B}), \quad (4.2.1.b)$$

where$\boldsymbol{\varphi}^*(B) = \boldsymbol{\varphi}^{**}(B) + \sum \lambda_k$.

Representation *(4.2.1.b)* indicates two major notes:

- The polynomial $\boldsymbol{\varphi}(B)$ will have a root at θ_k if and only if $\lambda_k = 0$ and therefore,
- testing for unit roots can be performed equivalently by testing for parameters $\lambda_k=0$.

In order to test for seasonal unit roots in quarterly data Hylleberg, Engle, Granger and Yoo, (1990) expanded a polynomial $\varphi(B)$ about the roots +1, -1 , i and –i as θ_k, k=1,…,4. Then, from *(4.1.1.b)*:

$$\begin{aligned}\boldsymbol{\varphi}(B) = {} & \lambda_1 B(1+B)(1+B^2) + \lambda_2(-B)(1-B)(1+B^2) \\ & + \lambda_3(-iB)(1-B)(1+B)(1-iB) \\ & + \lambda_4(iB)(1-B)(1+B)(1+iB) \\ & + \boldsymbol{\varphi}^*(B)(1-B^4). \qquad \textit{(4.2.1.c)}\end{aligned}$$

Notice that since $\boldsymbol{\varphi}(B)$ is real, λ_3 and λ_4 should be complex conjugates. More details of this derivation are given in the appendix- 1b.

Replacing the equations $\pi_1 = -\lambda_1$, $\pi_2 = -\lambda_2$, $2\lambda_3 = -\pi_3 + i\pi_4$ and $2\lambda_4 = -\pi_3 - i\pi_4$, in the *(4.2.1.c)*, yields the following form:

$$\boldsymbol{\varphi}(\mathbf{B}) = -\boldsymbol{\pi_1 B}(\mathbf{1}+\boldsymbol{B}+\boldsymbol{B}^{\mathbf{2}}+\boldsymbol{B}^{\mathbf{3}}) - \boldsymbol{\pi_2}(-\boldsymbol{B})(\mathbf{1}-\boldsymbol{B}+\boldsymbol{B}^{\mathbf{2}}-\boldsymbol{B}^{\mathbf{3}}) - (\boldsymbol{\pi_4}+\boldsymbol{\pi_3 B})(-\boldsymbol{B})(\mathbf{1}-\boldsymbol{B}^{\mathbf{2}}) + \boldsymbol{\varphi}^*(\boldsymbol{B})(\mathbf{1}-\boldsymbol{B}^{\mathbf{4}}) \quad \textit{(4.2.1.d)}$$

More details of this derivation are given in the appendix- 1c

Let $\boldsymbol{\varphi}(B)Y_t = \varepsilon_t$ be the data generating process. Replacing the equation *(4.2.1.d)* to $\boldsymbol{\varphi}(B)$, gives,

$$\boldsymbol{\varphi}^*(\boldsymbol{B})y_{4t} = \pi_1 y_{1t-1} + \pi_2 y_{2t-1} + \pi_3 y_{3t-2} + \pi_4 y_{3t-1} + \varepsilon_t$$

where

$y_{1t} = (1+B+B^2+B^3)Y_t$

$y_{2t} = -(1-B+B^2-B^3)Y_t$

$y_{3t} = -(1-B^2)Y_t$

$y_{4t} = (1-B^4)Y_t$

5.3 The HEGY Unit Root test for monthly data (S=12)

Correspondingly with the previous, let $\boldsymbol{\varphi}(B)Y_t = \varepsilon_t$ be the data generating process. We want to know if the polynomial $\boldsymbol{\varphi}(B)$ has roots equal to 1 in absolute value at zero frequency or seasonal frequencies. In order to test for non-seasonal and seasonal unit roots in monthly data, J.J. Beaulieu and J. A. Miron(1993) expanded the polynomial $\boldsymbol{\varphi}(B)$ about the roots of $z^{12}-1=0$ which are ± 1, $\pm i$, $\frac{1}{2}(\sqrt{3}\pm i)$, $\frac{1}{2}(1\pm\sqrt{3}i)$, $-\frac{1}{2}(1\pm\sqrt{3}i)$, $-\frac{1}{2}(\sqrt{3}\pm i)$ according to the Lagrange's proposition we listed above. The aim is to test the presence of a particular unit root without taking into consideration whether other seasonal unit roots are present.

For monthly data, replacing *(4.2.1.b)* into $\boldsymbol{\varphi}(B)Y_t = \varepsilon_t$ yields to the following equation:

$$\boldsymbol{\varphi}^*(B)y_{13t} = \sum_{k=1}^{12} \pi_k \, y_{k,t-1} + \varepsilon_t, \quad (4.3.1.a)$$

where,

$\boldsymbol{\varphi}^*(\boldsymbol{B}) = 1-\varphi_1 B-\varphi_2 B^2-\ldots-\varphi_p B^p$, and p can be chosen with criteria such as AIC, BIC e.t.c,

ε_t is white noise,

and

$y_{1t} = (1+B+B^2+B^3+B^4+B^5+B^6+B^7+B^8+B^9+B^{10}+B^{11})Y_t$

$y_{2t} = -(1-B+B^2-B^3+B^4-B^5+B^6-B^7+B^8-B^9+B^{10}-B^{11})Y_t$

$y_{3t} = -(B-B^3+B^5-B^7-B^9-B^{11})Y_t$

$y_{4t} = -(1-B^2+B^4-B^6+B^8-B^{10})Y_t$

$y_{5t} = -\frac{1}{2}(1+B-2B^2+B^3+B^4-2B^5+B^6+B^7-2B^8+B^9+B^{10}-2B^{11})Y_t$

$y_{6t} = \frac{\sqrt{3}}{2}(1-B+B^3-B^4+B^6-B^7+B^9-B^{10})Y_t$

$y_{7t} = \frac{1}{2}(1-B-2B^2-B^3+B^4+2B^5+B^6-B^7-2B^8-B^9+B^{10}+2B^{11})Y_t$

$y_{8t} = -\frac{\sqrt{3}}{2}(1+B-B^3-B^4+B^6+B^7-B^9-B^{10})Y_t$

$y_{9t} = -\frac{1}{2}(\sqrt{3}-B+B^3-\sqrt{3}B^4+2B^5-\sqrt{3}B^6+B^7-B^9+\sqrt{3}B^{10}-2B^{11})Y_t$

$y_{10t} = \frac{1}{2}(1-\sqrt{3}B+2B^2-\sqrt{3}B^3+B^4-B^6+\sqrt{3}B^7-2B^8+\sqrt{3}B^9-B^{10})Y_t$

$y_{11t} = \frac{1}{2}(\sqrt{3}+B-B^3-\sqrt{3}B^4-2B^5-\sqrt{3}B^6-B^7+B^9+\sqrt{3}B^{10}+2B^{11})Y_t$

$y_{12t} = -\frac{1}{2}(1+\sqrt{3}B+2B^2+\sqrt{3}B^3+B^4-B^6-\sqrt{3}B^7-2B^8-\sqrt{3}B^9-B^{10})Y_t$

$y_{13t} = (1-B^{12})Y_t$

For the frequencies 0 and π, in order to test the hypothesis that $\boldsymbol{\varphi(\theta_k) = 0}$ where θ_k is either 1,-1, k=1,2 it is simply tested that $\pi_k = 0$ against $\pi_k < 0$. For the seasonal frequencies we test if $\pi_k = \pi_{k-1} = 0$ versus $\pi_k \neq 0$ or $\pi_{k-1} \neq 0$ with a joint test or simply $\pi_{odd} = \pi_{even} = 0$ versus $\pi_{odd} \neq 0$ or $\pi_{even} \neq 0$.

Therefore, in the framework of the equation, *(4.3.1.a)* the hypothesis for the detection of unit roots at the zero as well as all the others seasonal frequencies are the following:

1. H_0: $\pi_1 = 0$, 1 is a root of the polynomial

H_1: $\pi_1 < 0$, 1 is not a root of the polynomial

2. H_0: $\pi_2 = 0$, -1 is a root of the polynomial

H_1: $\pi_2 < 0$, -1 is not a root of the polynomial

3. H_0: $\pi_3 = \pi_4 = 0$, $\pm i$ are roots of the polynomial

H_1: $\pi_3 \neq 0$ or $\pi_4 \neq 0$, $\pm i$ are not roots of the polynomial

4. H_0: $\pi_5 = \pi_6 = 0$, $-\frac{1}{2}(1 \pm \sqrt{3}i)$ are roots of the polynomial

H_1: $\pi_5 \neq 0$ or $\pi_6 \neq 0$, $-\frac{1}{2}(1 \pm \sqrt{3}i)$ are not roots of the polynomial

5. H_0: $\pi_7 = \pi_8 = 0$, $\frac{1}{2}(1 \pm \sqrt{3}i)$ are roots of the polynomial

H_1: $\pi_7 \neq 0$ or $\pi_8 \neq 0$, $\frac{1}{2}(1 \pm \sqrt{3}i)$ are not roots of the polynomial

6. H_0: $\pi_9 = \pi_{10} = 0$, $-\frac{1}{2}(\sqrt{3} \pm i)$ are roots of the polynomial

H_1: $\pi_9 \neq 0$ or $\pi_{10} \neq 0$, $-\frac{1}{2}(\sqrt{3} \pm i)$ are not roots of the polynomial

7. H_0: $\pi_{11} = \pi_{12} = 0$, $\frac{1}{2}(\sqrt{3} \pm i)$ are roots of the polynomial

H_1: $\pi_{11} \neq 0$ or $\pi_{12} \neq 0$, $\frac{1}{2}(\sqrt{3} \pm i)$ are not roots of the polynomial

The π's as well as the φ_i's are estimated with ordinary least squares. Beaulieu and Miron computed by Monte Carlo integration the critical values for the one-sided 't' tests on π_1 and π_2 as well as the critical values for the 'F' test on $\pi_k \cap \pi_{k-1}=0$ for k=4,…12 . Below are tabulated the 5% percentage points of the π_1 and $\pi_2$and the 95% percentage points for $\pi_k \cap \pi_{k-1}$ for $T=+\infty$ of five different models.

Estimated Equation	't': π_1	't': π_2
No intercept No seas. Dummies No trend	-1.95	-1.95
Intercept No seas. Dummies No trend	-2.86	-1.95
Intercept Seas. Dummies No trend	-2.86	-2.86
Intercept No seas. Dummies Trend	-3.40	-1.95
Intercept Seas. Dummies Trend	-3.40	-2.86

Table 5.3-*The 5% percentage points of the estimators for* $T=+\infty$

Estimated Equation	**'F': $\pi_{odd} \cap \pi_{even}$**
No intercept *No seas. Dummies* *No trend*	3.10
Intercept *No seas. Dummies* *No trend*	3.10
Intercept *Seas. Dummies* *No trend*	6.67
Intercept *No seas. Dummies* *Trend*	3.10
Intercept *Seas. Dummies* *Trend*	6.67

Table 5.4*-The 95% percentage points of the estimator for T=200*

CHAPTER 6

Application of the Seasonal Unit Root tests to the Greek Inflation

In this part, we conduct the seasonal ADF and the HEGY unit root test in order to detect seasonal and non-seasonal unit roots as well as to compare these two methods, using a dataset in the statistical package R. The time series analyzed in this example is the Greek inflation, labeled as infl91. Data are collected monthly from 1977.1 to 1991.12. Since 1991, where the adjustment of the drachma-inflation to the Euro-inflation began, the average inflation started to drop to an important extent. This fact results in a break in the expected value of the time series. In order to avoid the presence of this break in the data we analyzed[7], the data collected until December 1991. The data we had at our disposal is the Consumer Price Index in Greece, labeled as cpi. Taking the first logarithmic differences we created the time series of the Greek inflation at the corresponding period.

First, we will note the characteristics of our data.

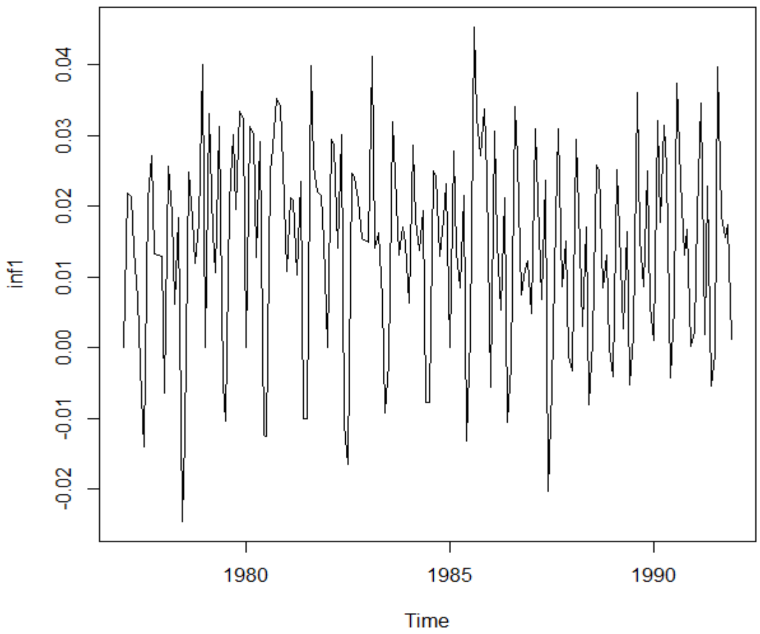


Figure 6.1*-Time Series Plot*

[7] The inclusion of a dummy to capture this break might alter the distribution of the statistics considered. This is a well-known fact for the classical Dickey-Fuller test.

The above time series plot seems to have a seasonal path while the ACF (which does not decrease exponentially). Moreover, the PACF plot indicates that an AR(12) would be necessary to describe the data.

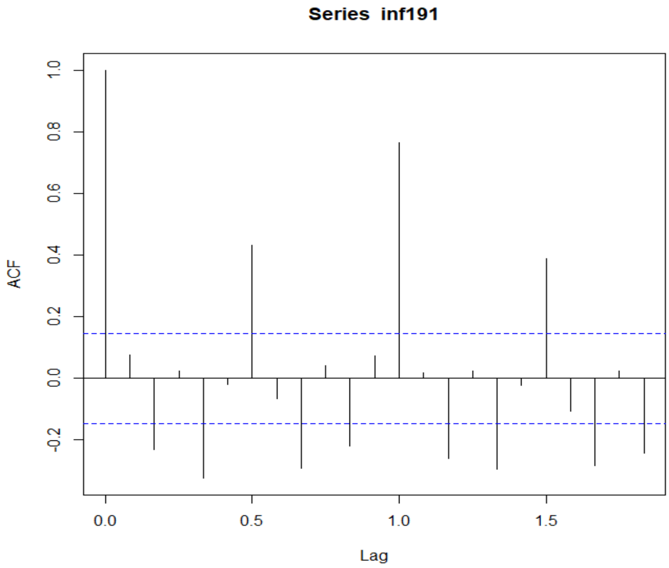


Figure 6.2*-Autocorrelation Function Plot*

In the ACF and PACF plots above, in the x-axis, the number 1.0 corresponds to the twelfth lag.

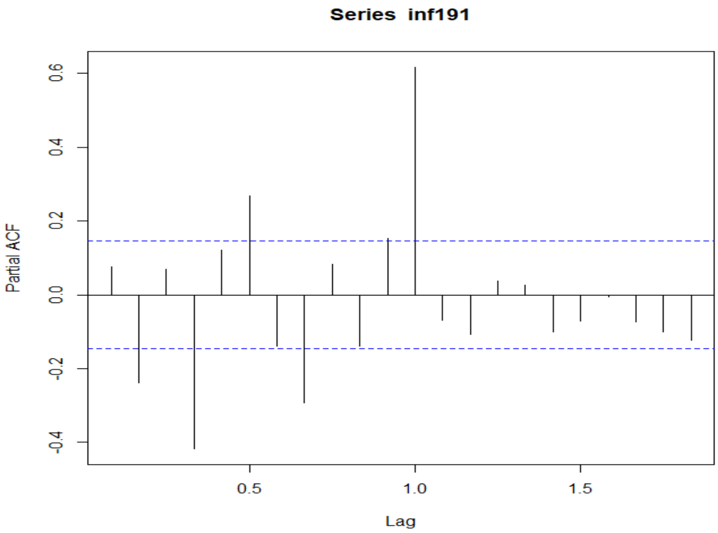


Figure 6.3*-Partial Autocorrelation Function Plot*

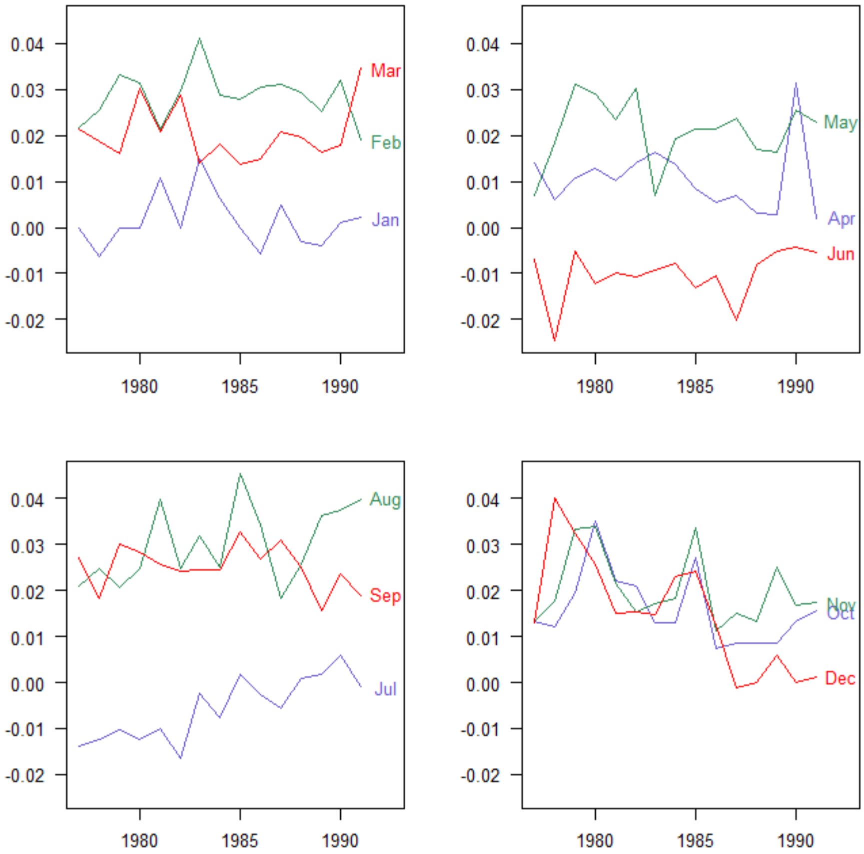


Figure 6.4*-The bbPlot*

The graphic above is a bbplot and it is provided in the 'uroot' package in R. It illustrates the 12 monthly paths of the series. The fact that these monthly paths are not parallel suggests that the seasonal patterns are not constant and hence some stochastic seasonality (stationary or non-stationary) will be present

7.1 The seasonal ADF unit root test: Application to the Greek Inflation

In this part, we performed the seasonal ADF using our data in the statistical package R. In order to do that, we applied the procedure of the **two-step regression** described in chapter 3. We remind that this procedure contains two regressions that finally give the ols estimators (α_S-1,$\boldsymbol{\theta}$-$\hat{\boldsymbol{\theta}}$) of the multiplicative model $(1-\alpha_S B^S)\cdot(1-\theta_1 B-\ldots-\theta_p B^P)\cdot(Y_t-\mu)=\varepsilon_t$.

The order of p in the first step of the two-step procedure is chosen by AIC and is equal to 19. The estimates of the coefficients with their standard errors and the t-statistics of the second stage regression are given in table 6.1.

	Estimates	**Std. Error**	**t-Statistic**
Intercept (μ)	0.001	0.001	0.91
$\boldsymbol{\alpha_S}$**- 1**	-0.055	0.027	-2.05
θ_1	0.022	0.087	0.26
θ_2	0.003	0.086	0.04
θ_3	-0.008	0.085	-0.10
θ_4	0.005	0.084	0.07
θ_5	-0.007	0.083	-0.09
θ_6	$-5.7\cdot10^{-5}$	0.084	-0.00
θ_7	0.014	0.084	0.18
θ_8	0.003	0.077	0.03
θ_9	-0.009	0.079	-0.12
θ_{10}	0.001	0.080	0.01
θ_{11}	-0.006	0.079	-0.07
θ_{12}	0.005	0.079	0.06
θ_{13}	0.014	0.088	0.16
θ_{14}	-0.003	0.087	-0.04
θ_{15}	-0.002	0.084	-0.03
θ_{16}	0.017	0.085	0.21
θ_{17}	-0.018	0.086	-0.21
θ_{18}	-0.008	0.087	-0.09
θ_{19}	0.018	0.086	0.21

Table 6.1-*Seasonal ADF Statistics*

We can see that the estimation of $\boldsymbol{\alpha_S}$**-1** is not close to zero and from Table 4 we see in the paper of Dickey, Hasza and Fuller (1984), the since our t-statistic is -2.05 the p-value is smaller than 0.05. Therefore, the null hypothesis of the seasonal ADF test is rejected in favor of stationarity of the series.

6.3 The HEGY unit root test: Application to the Greek Inflation

In this subsection, we performed the HEGY unit root and computed the corresponding statistics using our data in the statistical package R. The purpose is to detect seasonal unit roots in the period 1977.1-1991.12 of the Greek inflation at any single seasonal frequency. To conduct this test and obtain the statistics we used the 'uroot' package and specifically the function **HEGY.test**. As far as the deterministic components of the model are concerned, we added the intercept as well as the eleven seasonal dummies, while the order p was selected by the criterion of the significance of the φ_i's.

In the table 6.2 are illustrated the indexes of the seasonal unit roots, the seasonal unit roots as well as the corresponding frequencies.

Index	Unit Roots	Frequencies
1	1	0
2	-1	π
3,4	$\pm i$	$\frac{\pi}{2}, \frac{3\pi}{2}$
5,6	$-\frac{1}{2}(1\pm\sqrt{3}i)$	$\frac{2\pi}{3}, \frac{4\pi}{3}$
7,8	$\frac{1}{2}(1\pm\sqrt{3}i)$	$\frac{\pi}{3}, \frac{5\pi}{3}$
9,10	$-\frac{1}{2}(\sqrt{3}\pm i)$	$\frac{5\pi}{6}, \frac{7\pi}{6}$
11,12	$\frac{1}{2}(\sqrt{3}\pm i)$	$\frac{\pi}{6}, \frac{11\pi}{6}$

Table 6.2*-Seasonal Unit Roots and Frequencies*

In the table 6.3 are presented the Hegy statistics with their corresponding p-values. We remind that the null hypothesis of each test is the presence of unit root while the alternative is stationarity.

Test	Statistics	P-Value
tpi_1	-2.321	**0.100**
tpi_2	-3.200	0.014
Fpi_3:4	12.751	0.010
Fpi_5:6	12.364	0.010
Fpi_7:8	9.362	0.010
Fpi_9:10	6.771	0.036
Fpi_11:12	5.109	**0.100**

Table 6.3*-Hegy Statistics*

According to the table 6.3 the null hypothesis is not rejected in the zero frequency (t:-2.32) as well as the seasonal frequencies $\frac{\pi}{6}$ and $\frac{11\pi}{6}$ (F: 5.109) at the 5% level of significance. Therefore, we consider the presence of the unit root 1 and the seasonal unit roots $\frac{1}{2}(\sqrt{3}+i)$ and $\frac{1}{2}(\sqrt{3}-i)$.

	Estimates	Std.Error	P-Value
Intercept	0.008	0.005	0.118
SeasDummy.1	-0.019	0.006	0.004
SeasDummy.2	0.011	0.007	0.115
SeasDummy.3	0.006	0.005	0.268
SeasDummy.4	-0.007	0.007	0.326
SeasDummy.5	0.008	0.007	0.209
SeasDummy.6	-0.021	0.004	0.000
SeasDummy.7	-0.020	0.006	0.002
SeasDummy.8	0.019	0.007	0.006
SeasDummy.9	0.006	0.005	0.251
SeasDummy.10	-0.009	0.007	0.185
SeasDummy.11	0.006	0.006	0.383

Table 6.4*-Deterministic Regressors Estimates*

In the table 6.4 are tabulated the estimates as well as the standard errors and the corresponding p-values of the t-statistics of the deterministic regressors.

	Estimates	Std.Error
π_1	0.008	0.005
π_2	-0.019	0.006
π_3	0.011	0.007
π_4	0.006	0.005
π_5	-0.007	0.007
π_6	0.008	0.007
π_7	-0.021	0.004
π_8	-0.020	0.006
π_9	0.019	0.007
π_{10}	0.006	0.005
π_{11}	-0.009	0.007
π_{12}	0.006	0.006

Table 6.5*-Hegy Regressors Estimates*

The Table 6.5 illustrates the estimates and the standard errors of the π_i'sof the equation (4.3.1.a) in the framework of the time series analyzed.

If the seasonal ADF test did not reject the null hypothesis and thus all the seasonal frequencies are considered non-stationary, we should apply twelve differences in order to achieve stationarity. Therefore, for $\boldsymbol{\varphi}(B) = (1-B^{12})$, the series

$$\boldsymbol{\varphi}(B)Y_t \text{ would be stationary.}$$

On the contrary, if the HEGY test indicates the presence of unit roots in the seasonal frequencies $\lambda_0 = \pm\frac{\pi}{6}$ and at $\lambda=0$, the form of the filter changes.

Let $z_0 = e^{i\lambda_0}$ and $z_1 = 1$ be the roots of the polynomial $\boldsymbol{\varphi}(B)$, then:

$$\begin{aligned}\boldsymbol{\varphi}(B) &= (1-B)(1- z_0^{-1}B)(1-\overline{z_0^{-1}}B) = \\ &(1-B)(1- 2\mathrm{Re}(z_0)B + |z_0^{-1}|^2B^2) = \\ &(1-B)(1- \cos(\lambda_0)B + B^2)\end{aligned}$$

Therefore, for $\boldsymbol{\varphi}(B) = (1-B)(1- \cos(\lambda_0)B + B^2)$, the series

$$\boldsymbol{\varphi}(B)Y_t \text{ would be stationary.}$$

6.5 Conclusions

From the aforementioned facts it appears that the seasonal ADF and the HEGY unit root tests in the framework of the Greek inflation 1977.1-1991.12 conclude to different results. On the one hand, the seasonal ADF test rejects the null that the roots of the corresponding autoregressive polynomial $\boldsymbol{\varphi}(B)$ are on the unit circle. Thus it rejects non-stationary seasonality in favor of stationary seasonality. On the other hand, the HEGY test rejects all seasonal unit roots except those at the frequencies $\frac{\pi}{6}$ and $\frac{11\pi}{6}$ and 0 Thus HEGY does not reject the presence of non-stationary seasonality. It does, however, indicate a filter to achieve stationarity which is different from the usually employed twelfth-differences-filter.

CHAPTER 7

Appendix

7.1 Proofs

1a. The change ΔY_τ is a stationary process that can be forecasting using the standard method:

$$\Delta\hat{Y}_{t+s|t} \equiv \hat{E}[(Y_{t+s}-Y_{t+s-1}|Y_t, Y_{t-1},..)] = \delta + \psi_s\varepsilon_t + \psi_{s+1}\varepsilon_{t-1} + \psi_{s+2}\varepsilon_{t-2} + \ldots \quad (1)$$

Making the proper transformation, we note that the variable at the level t+s, is the sum of the changes between the time t and t+s:

$$Yt+s = (Y_{t+s}-Y_{t+s-1}) + (Y_{t+s-1}-Y_{t+s-2}) + .. + (Y_{t+1}-Y_t) + Y_t = \Delta Y_{t+s} + \Delta Y_{t+s-1} + .. + \Delta Y_{t+1} + Y_t \quad (2)$$

Therefore, taking into consideration the equation *(2)* and the forecast of the change ΔY_t in equation *(1)*, it is concluded that the forecast of a unit root process has the following form:

$$\hat{Y}_{t+s|t} = \Delta\hat{Y}_{t+s|t} + \Delta\hat{Y}_{t+s-1|t} + \ldots + \Delta\hat{Y}_{t+1|t} + Y_t$$

$$= \{\delta + \psi_s\varepsilon_t + \psi_{s+1}\varepsilon_{t-1} + \psi_{s+2}\varepsilon_{t-2} + ..\} + \{\delta + \psi_{s-1}\varepsilon_t + \psi_s\varepsilon_{t-1} + \psi_{s+1}\varepsilon_{t-2} + ..\} + \{\delta + \psi_{s-2}\varepsilon_t + \psi_{s-1}\varepsilon_{t-1} + \psi_s\varepsilon_{t-2} + ..\} + \ldots + \{\delta + \psi_1\varepsilon_t + \psi_2\varepsilon_{t-1} + \psi_3\varepsilon_{t-2} + ..\} + Y_t\,.$$

$$\hat{Y}_{t+s|t} = s\delta + Y_t + (\psi_s + \psi_{s-1} + .. + \psi_1)\varepsilon_t + (\psi_{s+1} + \psi_s + .. + \psi_2)\varepsilon_{t-1} + ..$$

1b. For $\theta_1 = +1$, $\theta_2 = -1$, $\theta_3 = i$ and $\theta_4 = -i$;

- $\delta_k(B) = 1-\frac{1}{\theta_k}B$: $\delta_1(B) = 1-B$

  $\delta_2(B) = 1+B$

  $\delta_3(B) = 1-\frac{1}{i}B = 1 + iB$

  $\delta_4(B) = 1+\frac{1}{i}B = 1-iB$

- $\Delta(B) = \prod_{k=1}^{4}\delta_k(B) = (1-B)(1+B)(1+iB)(1-iB) = (1-B^2)(1+B^2) = (1-B^4)$

- $\lambda_k \Delta(B)(1-\delta_k(B))/\delta_k(B)$ **:**

  For k=1, $\lambda_1(1-B^4)(1-1+B)/(1-B) = \lambda_1(1+B)(1+B^2)B$

  For k=2, $\lambda_2(1-B^4)(1-1-B)/(1+B) = \lambda_1(1+B)(1+B^2)(-B)$

  For k=3, $\lambda_3(1-B^4)(1-1-iB)/(1+iB) = \lambda_3(1+B)(1-B)(-iB)(1-iB)$

  For k=4, $\lambda_4(1-B^4)(1-1+iB)/(1-iB) = \lambda_4(1+B)(1-B)(iB)(1+iB)$

- $\Delta(B)\varphi^*(B) = (1-B^4)\varphi^*(B)$

Substituting the equations above in the equation *(4.1.1.b)* gives the equation *(4.1.1.c)***.**

1c. $\lambda_1 B(1+B)(1+B^2) + \lambda_2(-B)(1-B)(1+B^2)$

$+\lambda_3(-iB)(1-B)(1+B)(1-iB)$

$+ \lambda_4(iB)(1-B)(1+B)(1+iB) =$

$= - \pi_1 B(1+B+B^2+B^3) - \pi_2(-B)(1-B+B^2-B^3)$

$+ (\frac{-\pi_3 + i\pi_4}{2})(-iB)(1-B)(1+B)(1-iB) + + (\frac{-\pi_3 - i\pi_4}{2})(iB)(1-$

$B)(1+B)(1+iB)=$

$= - \pi_1 B(1+B+B^2+B^3) - \pi_2(-B)(1-B+B^2-B^3) +$

$-\frac{\pi_3}{2}(1-B)(1+B)[(iB)(1+iB)+(-iB)(1-iB)] - \frac{\pi_4 i}{2}(1-B)(1+B)[(iB)(1-$

$iB)+iB(1+iB)]=$

$= \ - \pi_1 B(1+B+B^2+B^3) - \pi_2(-B)(1-B+B^2-B^3) +$

$-\frac{\pi_3}{2}(1-B^2)[iB - B^2 - iB - B^2] \ - \frac{\pi_4 i}{2}(1-B^2)[iB + B^2 + iB - B^2]=$

$$= -\pi_1 B(1+B+B^2+B^3) - \pi_2(-B)(1-B+B^2-B^3) +$$

$$-\frac{\pi_3}{2}(1-B^2)(-2B^2) - \frac{\pi_4 i}{2}(1-B^2)(2iB) =$$

$$= -\pi_1 B(1+B+B^2+B^3) - \pi_2(-B)(1-B+B^2-B^3) + \pi_3(1-B^2)B^2 + \pi_4(1-B^2)B =$$

$$= -\pi_1 B(1+B+B^2+B^3) - \pi_2(-B)(1-B+B^2-B^3) - (\pi_3 B + \pi_4)(-B)(1-B^2).$$

So,

$$\varphi(B) = \lambda_1 B(1+B)(1+B^2) + \lambda_2(-B)(1-B)(1+B^2) + \lambda_3(-iB)(1-B)(1+B)(1-iB)$$
$$+ \lambda_4(iB)(1-B)(1+B)(1+iB) + \varphi^*(B)(1-B^4) =$$

$$= -\pi_1 B(1+B+B^2+B^3) - \pi_2(-B)(1-B+B^2-B^3) - (\pi_3 B + \pi_4)(-B)(1-B^2) + \varphi^*(B)(1-B^4).$$

7.3 Code

Code for the Figure 2.1

Simulate a random walk

```
n <- 20
eps<- rnorm(n)
x0<- rep(0, n)
d<-0.2
for (i in seq.int(2, n))
x0[i] <- d+x0[i-1] + eps[i]
```

Simulate a trend Stationary process

```
innovs<- rnorm(20,0,1)
x<-1:20    #time variable
mu<-10+0.5*x+innovs   #linear trend


library(forecast)
```

```
par(mfrow=c(2,1))
plot.forecast(forecast(x0),xlab="Time",main="Forecast of a Unit Root Process")
abline(lsfit(1:20,ts(x0), intercept =TRUE ),col=2)

plot.forecast(forecast(mu),xlab="Time",main="Forecast of Trend Stationary Process"
)
abline(lsfit(1:20,ts(mu), intercept =TRUE),col=2)
```

Code for Figures 3.2-3.4

Simulation of AR(2) model with inverse root close to unity

```
pre.ss<-100
period<- 12
mod.inv<- 0.9
ss<- period *5


lamda<-2*pi/period
p<-2
fi<-c(2*mod.inv*cos(lamda ),-mod.inv^2)


e <- rnorm(ss+pre.ss)


y.init<- rep(NA,ss+pre.ss)
for (i in 1:p)
        {
        y.init[i]<-e[i]
        }

for (i in (p+1):(ss+pre.ss))
        {
        y.init[i]<- fi[1]*y.init[i-1]+fi[2]*y.init[i-2]+e[i]
        }

y <- rep(NA,ss)
y[1:ss] <- y.init[(pre.ss+1):(pre.ss+ss)]
```

Figure 3.2

```
ts.plot(y,type="b",main="Time Series Plot")
```

Figure 3.3

```
ts.plot(acf(y,lag.max=3*period))
```

Figure 3.4

```
fit<- Arima(y,order=c(2,0,0),seasonal=c(0,0,0))
plot(arroots(fit),main="Inverse AR(2) roots")
```

Code for Figure 3.7

```
plot(nottem,type="o",col=4,pch=18,main="Run Sequence Plot",ylab="Average Air Temperature",xlab="Index")
plot(acf(nottem),col=3,main="Autocorrelation Plot")
monthplot(nottem,col=4,main="Seasonal Subseries Plot",ylab="Average Air Temperature",xlab="Month")
```

Code for Figure 3.4

Finding AR roots

```
arroots<- function(object)
   {
if(!("Arima" %in% class(object)) & !("ar" %in% class(object)))
stop("object must be of class Arima or ar")
if("Arima" %in% class(object))
parvec<- object$model$phi
else
parvec<- object$ar
if(length(parvec) > 0)
     {
last.nonzero<- max(which(abs(parvec) > 1e-08))
if (last.nonzero> 0)
return(structure(list(roots=polyroot(c(1,-parvec[1:last.nonzero])),
          type="AR"), class='armaroots'))
     }
return(structure(list(roots=numeric(0),type="AR"),class='armaroots'))
   }
```

Plot Inverse Roots

```
plot.armaroots<- function(x, xlab="Real",ylab="Imaginary",
main=paste("Inverse roots of",x$type,"characteristic polynomial"),
        ...)
   {
oldpar<- par(pty='s')
on.exit(par(oldpar))
plot(c(-1,1),c(-1,1),xlab=xlab,ylab=ylab,
type="n",bty="n",xaxt="n",yaxt="n", main=main, ...)
axis(1,at=c(-1,0,1),line=0.5,tck=-0.025)
      axis(2,at=c(-1,0,1),label=c("-i","0","i"),line=0.5,tck=-0.025)
circx<- seq(-1,1,l=501)
circy<- sqrt(1-circx^2)
lines(c(circx,circx),c(circy,-circy),col='gray')
```

```
lines(c(-2,2),c(0,0),col='gray')
lines(c(0,0),c(-2,2),col='gray')
if(length(x$roots) > 0) {
inside<- abs(x$roots) > 1
points(1/x$roots[inside],pch=19,col='black')
if(sum(!inside) > 0)
points(1/x$roots[!inside],pch=19,col='red')
    }
  }
```

Generating the SARIMA models

```
model <- Arima(ts(rnorm(100),freq=12), order=c(1,1,1), seasonal=c(1,1,1),
fixed=c(phi=0.5, theta=-0.4, Phi=0.99, Theta=-0.2))
foo<- simulate(model, nsim=1000)
fit1 <- Arima(foo, order=c(1,1,1), seasonal=c(1,1,1))



model <- Arima(ts(rnorm(100),freq=12), order=c(1,1,0), seasonal=c(1,1,0),
fixed=c(phi=0.6, Phi=0.3))
foo<- simulate(model, nsim=1000)
fit2 <- Arima(foo, order=c(1,1,0), seasonal=c(1,1,0))
```

Plotting the Inverse Roots

```
par(mfrow=c(1,2))
plot(arroots(fit1),main="Inverse AR roots")
plot(arroots(fit2),main="Inverse AR roots")
```

Code, plots and output in R of the Example:

3.2.3 “Assuming the correct vs the wrong type of seasonality”: An illustration example

1) Modelling $Y_t=Y_{t-d}+et$ as $Y_t =\mu + \alpha Y_{t-d} + \varepsilon_t$

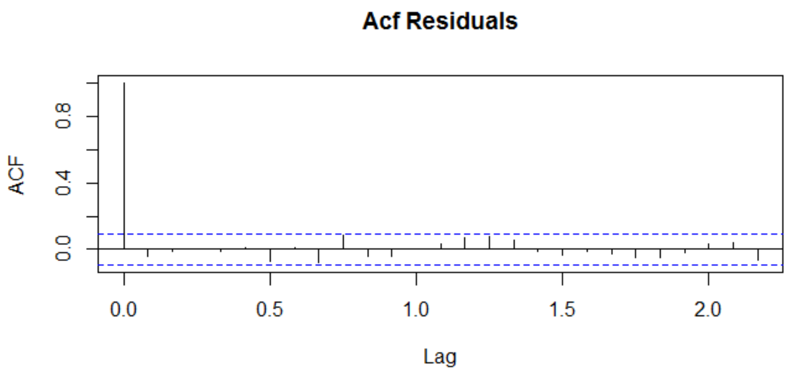
model1_data1<- Arima(data1, order=c(0,0,0), seasonal=c(1,0,0))
summary(model1_data1)
Series: data1
ARIMA(0,0,0)(1,0,0)[12] with non-zero mean

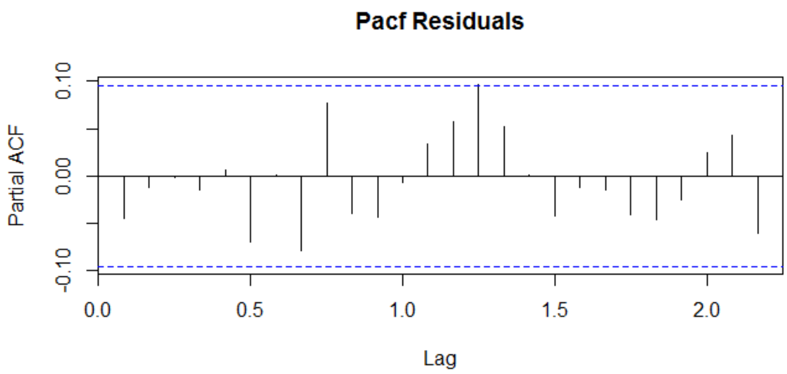
Coefficients:

	sar1	intercept
	0.9675	**0.8922**
s.e.	**0.0090**	**0.8790**

Plots

```
par(mfrow=c(2,1))
acf(model1_data1$residuals, main="Acf Residuals")
pacf(model1_data1$residuals, main="Pacf Residuals")
```





2) <u>Modelling $Y_t=Y_{t-d}+et$ as $Y_t = \mu + \sum_{s=1}^{S} m_s \delta_{st} + \varepsilon_t$</u>

model2_data1<-lm(data1~ dummies)

summary(model2_data1)

Residuals:

Min	1Q	Median	3Q	Max
-7.8729	-1.6145	0.0396	1.5569	6.9854

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	**-2.9513**	0.3942	-7.487	4.35e-13	***
dummiesJan	2.3545	0.5536	4.253	2.61e-05	***
dummiesFeb	4.1176	0.5536	7.438	6.02e-13	***
dummiesMar	4.8632	0.5574	8.724	< 2e-16	***
dummiesApr	0.5062	0.5574	0.908	0.364339	
dummiesMay	2.1116	0.5574	3.788	0.000175	***
dummiesJun	-0.2775	0.5574	-0.498	0.618896	
dummiesJul	6.4631	0.5574	11.594	< 2e-16	***

dummiesAug7.7878   0.5574 13.970 < 2e-16 ***
dummiesSep0.6601   0.5574   1.184   0.237071
dummiesOct9.4955   0.5574 17.034 < 2e-16 ***
dummiesNov7.4477   0.5574 13.360 < 2e-16 ***

Residual standard error: 2.332 on 410 degrees of freedom
Multiple R-squared: 0.6655,   Adjusted R-squared: 0.6565
F-statistic: 74.16 on 11 and 410 DF, p-value: < 2.2e-16

3)Modelling$Y_t = Y_{t-d} + et$ as $Y_t = \mu + \alpha Y_{t-d} + \sum_{s=1}^{S} m_s \delta_{st} + \varepsilon_t$

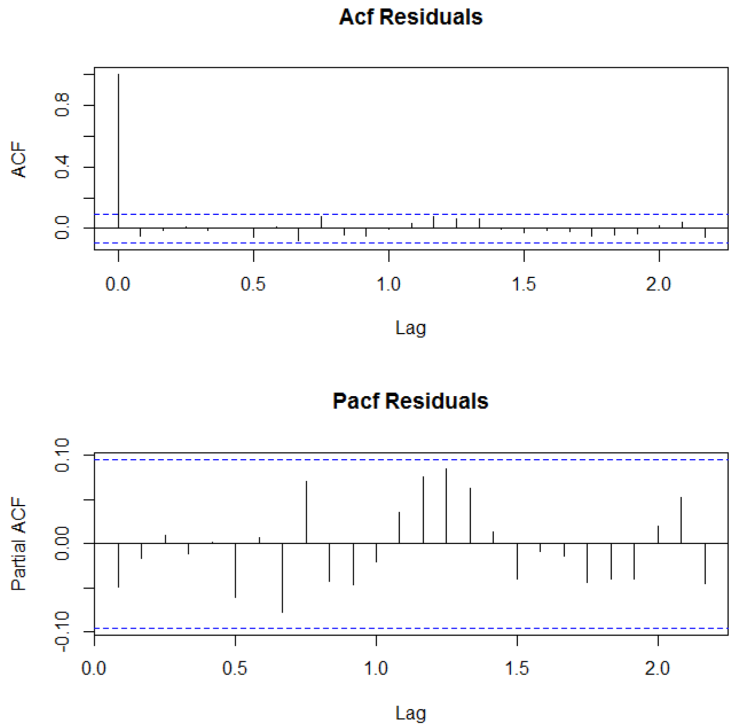
model3_data1<- Arima(data1, order=c(0,0,0), seasonal=c(1,0,0),xreg=dummies)

summary(model3_data1)
ARIMA(0,0,0)(1,0,0)[12] with non-zero mean

	Sar1	Inter	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
estimate	0.9220	-2.12	2.1	3.74	2.92	0.470	1.586	-1.078	5.63	6.606	-0.080	8.26	5.72

sigma^2 estimated as 0.9403: log likelihood=-590.59
AIC=1209.19   AICc=1210.22   BIC=1265.82

par(mfrow=c(2,1))
acf(model3_data1$residuals, main="Acf Residuals")
pacf(model3_data1$residuals, main="Pacf Residuals")



4) Modelling $Y_t=\mu+\sum_{s=1}^{S} m_s \delta_{st} + \varepsilon_t$ as $Y_t = Y_{t-d}$+et

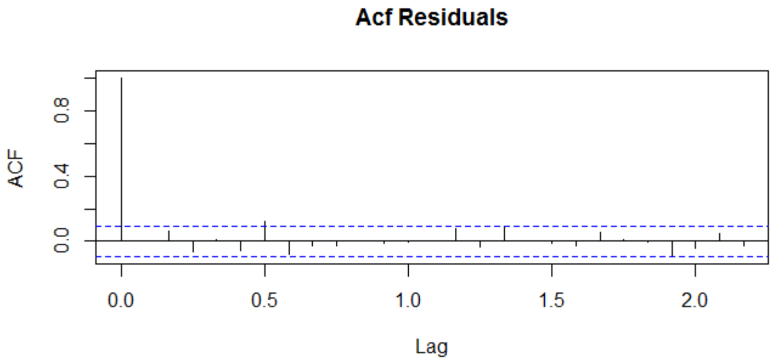
model1_data2<- Arima(data2, order=c(0,0,0), seasonal=c(1,0,0))

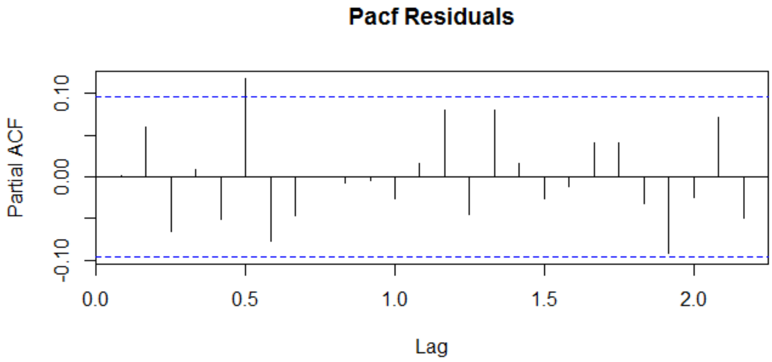
summary(model1_data2)

ARIMA(0,0,0)(1,0,0)[12] with non-zero mean

```
Coefficients:
sar1  intercept
     -0.0254    0.0969
s.e.   0.0491    0.0496

sigma^2 estimated as 1.094:  log likelihood=-616.7
AIC=1239.4   AICc=1239.45   BIC=1251.53
```

```
par(mfrow=c(2,1))
acf(model1_data2$residuals, main="Acf Residuals")
pacf(model1_data2$residuals, main="Pacf Residuals")
```





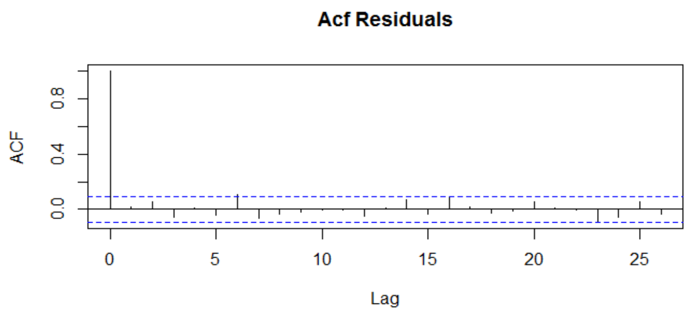
5)<u>Modelling $Y_t=\mu+\sum_{s=1}^{S} m_s\delta_{st}+\varepsilon_t$ as $Y_t=\mu+\sum_{s=1}^{S} m_s\delta_{st}+\varepsilon_t$</u>

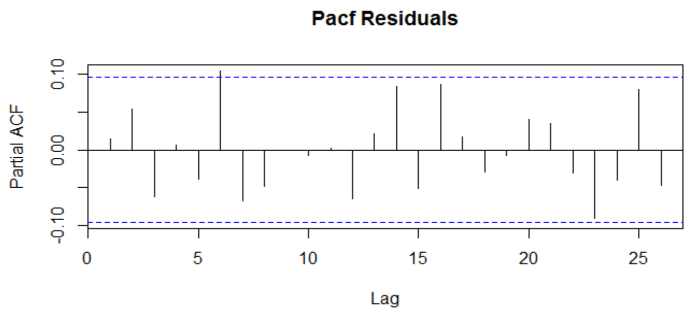
model2_data2<-lm(data2~ dummies)
summary(model2_data2)

```
lm(formula = data2 ~ dummies)

Residuals:
     Min       1Q   Median       3Q      Max
-2.81844 -0.72723 -0.02632  0.72704  2.91865

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.049362   0.176883  -0.279   0.7803
dummiesJan   0.004981   0.248408   0.020   0.9840
dummiesFeb   0.044578   0.248408   0.179   0.8577
dummiesMar   0.419974   0.250151   1.679   0.0939 .
dummiesApr   0.057479   0.250151   0.230   0.8184
dummiesMay   0.327103   0.250151   1.308   0.1917
dummiesJun   0.133369   0.250151   0.533   0.5942
dummiesJul   0.118638   0.250151   0.474   0.6356
dummiesAug   0.110486   0.250151   0.442   0.6590
dummiesSep   0.317813   0.250151   1.270   0.2046
dummiesOct  -0.113989   0.250151  -0.456   0.6489
dummiesNov   0.340994   0.250151   1.363   0.1736

Residual standard error: 1.046 on 410 degrees of freedom
```

Multiple R-squared: 0.02325,   Adjusted R-squared: -0.002955
F-statistic: 0.8872 on 11 and 410 DF, p-value: 0.553

Par(mfrow=c(2,1))
acf(model2_data2$residuals, main="Acf Residuals")
pacf(model2_data2$residuals, main="Pacf Residuals")





6) Modelling $Y_t=\mu+\sum_{s=1}^{S} m_s \delta_{st}+\varepsilon_t$ as $Y_t=\mu+\alpha Y_{t-d}+\sum_{s=1}^{S} m_s \delta_{st}+\varepsilon_t$

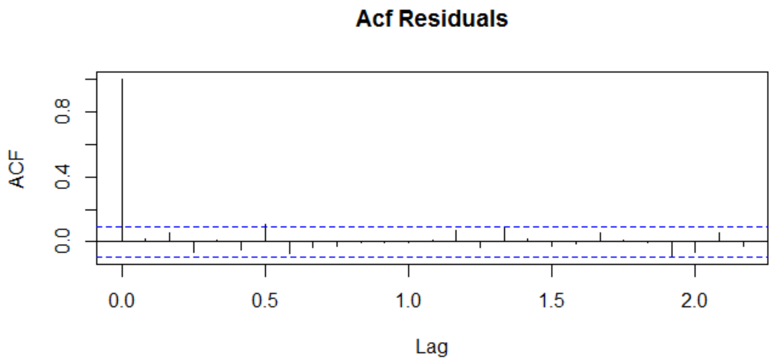
model3_data2<- Arima(data2, order=c(0,0,0), seasonal=c(1,0,0),xreg=dummies)
summary(model3_data2)

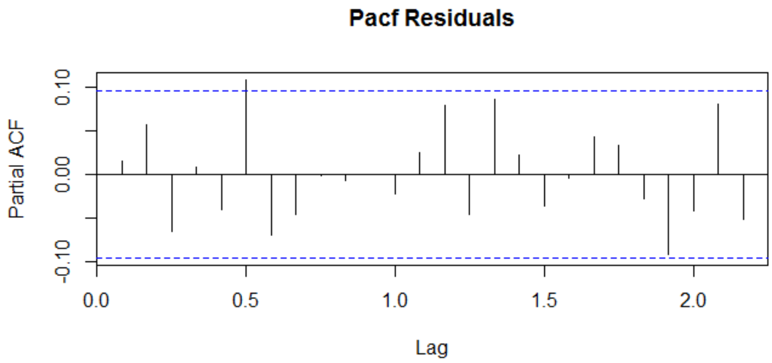
ARIMA(0,0,0)(1,0,0)[12] with non-zero mean

Coefficients:
sar1   intercept   Jan   Feb   Mar   Apr   May   Jun   Jul   Aug
      -0.0464   -0.0511   0.0068   0.0476 0.4205 0.0621 0.3294 0.1368 0.1198
0.1143
s.e.   0.0490   0.1667   0.2340 0.2340 0.2357 0.2357 0.2357 0.2357 0.2357
0.2357
        Sep   Oct   Nov
0.3163 -0.1112 0.3409
s.e. 0.2357   0.2357 0.2357

sigma^2 estimated as 1.095: log likelihood=-611.42
AIC=1250.84   AICc=1251.87   BIC=1307.47

```
par(mfrow=c(2,1))
acf(model3_data2$residuals, main="Acf Residuals")
pacf(model3_data2$residuals, main="Pacf Residuals")
```



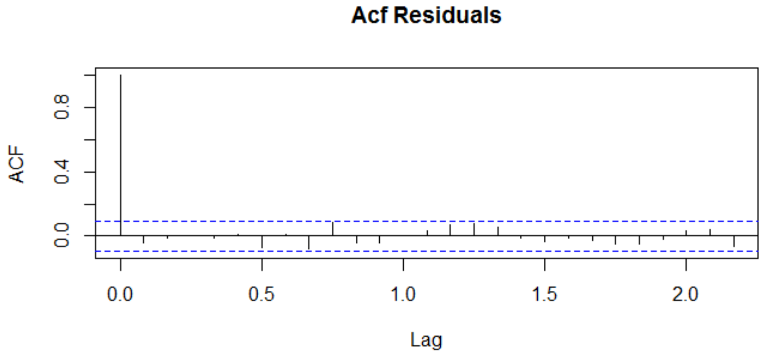


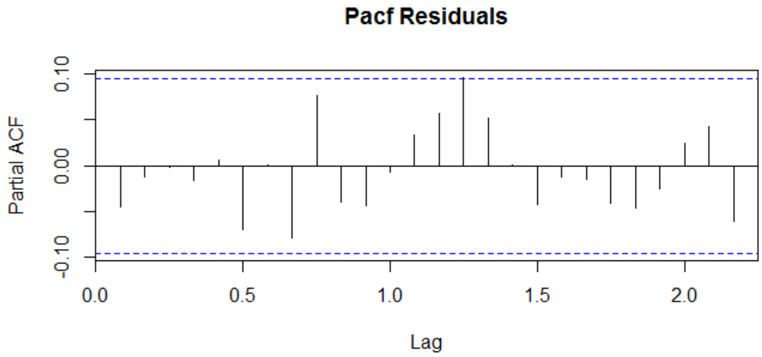
7) Modelling $Y_t=\mu+\alpha Y_{t-d}+\sum_{s=1}^{S} m_s \delta_{st}+\varepsilon_t$ as $Y_t=Y_{t-d}+et$

model1_data3<- Arima(data3, order=c(0,0,0), seasonal=c(1,0,0))
summary(model1_data3)

```
ARIMA(0,0,0)(1,0,0)[12] with non-zero mean

Coefficients:
sar1      intercept
      0.9675     0.9001
s.e.  0.0090     0.8800

sigma^2 estimated as 0.9328:  log likelihood=-599.61
AIC=1205.21   AICc=1205.27   BIC=1217.35
```

```
par(mfrow=c(2,1))
acf(model1_data3$residuals, main="Acf Residuals")
pacf(model1_data3$residuals, main="Pacf Residuals")
```





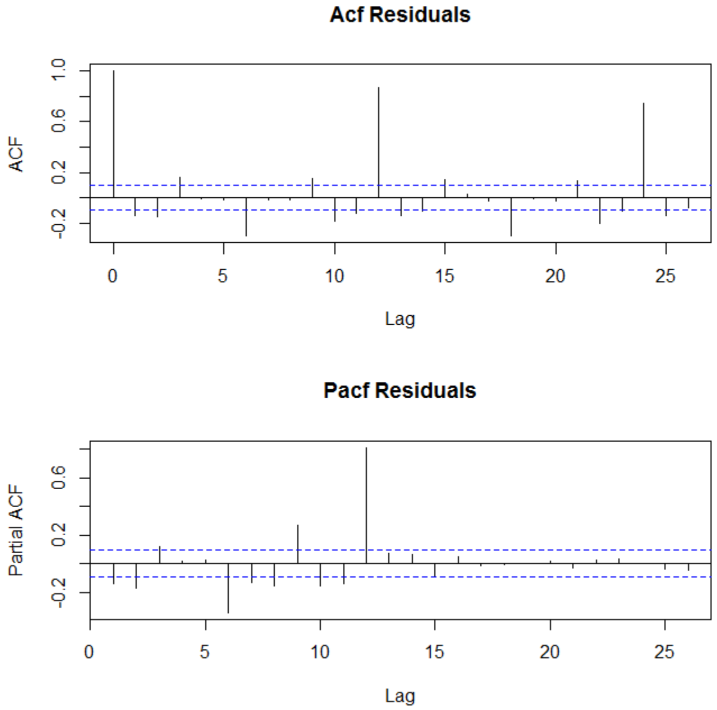
8)<u>Modelling $Y_t=\mu+\alpha Y_{t-d}+\sum_{s=1}^{S} m_s\delta_{st}+\varepsilon_t$ as $Y_t=\mu+\sum_{s=1}^{S} m_s\delta_{st}+\varepsilon_t$</u>

model2_data3<-lm(data3~ dummies)
summary(model2_data3)

```
Residuals:
    Min      1Q  Median      3Q     Max 
-7.8729 -1.6145  0.0396  1.5569  6.9854 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -2.9479     0.3942  -7.479 4.60e-13 ***
dummiesJan    2.3487     0.5536   4.243 2.73e-05 ***
dummiesFeb    4.1398     0.5536   7.478 4.61e-13 ***
dummiesMar    4.8735     0.5574   8.743  < 2e-16 ***
dummiesApr    0.5096     0.5574   0.914 0.361181    
dummiesMay    2.1181     0.5574   3.800 0.000167 ***
dummiesJun   -0.2934     0.5574  -0.526 0.598901    
dummiesJul    6.4563     0.5574  11.582  < 2e-16 ***
dummiesAug    7.8098     0.5574  14.010  < 2e-16 ***
dummiesSep    0.6713     0.5574   1.204 0.229198    
dummiesOct    9.5003     0.5574  17.042  < 2e-16 ***
dummiesNov    7.4565     0.5574  13.376  < 2e-16 ***


Residual standard error: 2.332 on 410 degrees of freedom
```

Multiple R-squared: 0.6661,   Adjusted R-squared: 0.6571
F-statistic: 74.35 on 11 and 410 DF, p-value: < 2.2e-16

par(mfrow=c(2,1))
acf(model2_data3$residuals, main="Acf Residuals")
pacf(model2_data3$residuals, main="Pacf Residuals")



9)Modelling $Y_t=\mu+\alpha Y_{t-d}+\sum_{s=1}^{S} m_s \delta_{st}+\varepsilon_t$ as$Y_t=\mu+\alpha Y_{t-d}+\sum_{s=1}^{S} m_s \delta_{st}+\varepsilon_t$

model3_data3<- Arima(data3, order=c(0,0,0), seasonal=c(1,0,0),xreg=dummies)
summary(model3_data3)

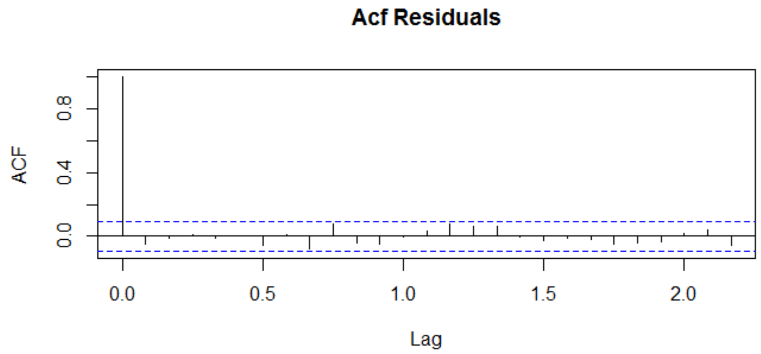
ARIMA(0,0,0)(1,0,0)[12] with non-zero mean

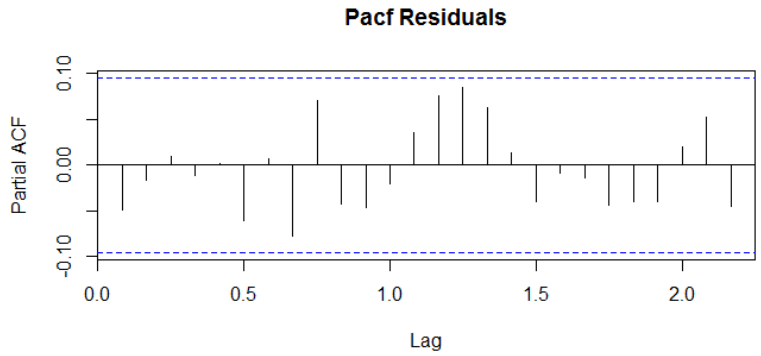
Coefficients:
sar1 intercept   Jan   Feb   Mar   Apr   May   Jun   Jul   Aug
      0.9220   -2.1126 2.0956 3.7627 2.9316 0.4737 1.5922 -1.0939 5.6264
6.6276
s.e. 0.0175   1.6025 2.2505 2.2509 2.2772 2.2597 2.2610   2.2627 2.2629
2.2662
          Sep   Oct   Nov
      -0.0689 8.2615 5.7243
s.e.2.2623 2.2669 2.2737

sigma^2 estimated as 0.9403: log likelihood=-590.59

```
AIC=1209.19   AICc=1210.22   BIC=1265.82
```

```
par(mfrow=c(2,1))
acf(model3_data3$residuals, main="Acf Residuals")
pacf(model3_data3$residuals, main="Pacf Residuals")
```





<u>Code in R for time series plot and ACF/PACF plots using the dataset of the inflation</u>

```
##Setingwd##
setwd("C:/Users/Georgia/Documents/thesis")

##Read Data##
cpi<-read.table("cpi_gr.txt")
cpi_val<-ts(cpi[,2],start=c(1977,1),end=c(2012,2),freq=12)


##Creating Inflation Series##
infl<-log(cpi_val)-log(lag(cpi_val,-1))
inf<-ts(infl,start=c(1977,1),end=c(2012,2),freq=12)


##FirstPart###
inf191<-ts(infl,start=c(1977,1),end=c(1991,12),freq=12)

plot(infl91,type="l")
```

```
acf(inf191 ,lag.max =NULL)
pacf(inf191 ,lag.max =NULL)
```

Code in R of the seasonal ADF test using the dataset of the inflation-Chapter 6

(1)

```
model1<-ar(diff(infl91,12),aic="TRUE",order.max=24,method="ols")
resid<-model1$res
```

(2)

```
##creation of the first independent variable##

df<-NULL
for(i in 12:31){
k<-diff(infl91,i)
df<-cbind(df,k)
  }

ncol(df)
nrow(df)

coefs<-as.matrix(c(1,-model1$ar))

ncol(coefs)
nrow(coefs)

Ydot<-df%*%coefs

Y<-NULL
for(i in 1:19){
k1<-diff(diff(infl91,12),i)
Y<-cbind(Y,k1)
}

Y<-as.matrix(Y)
nrow(Y)
ncol(Y)

variab<-cbind(Ydot[-1],Y)


model2<-lm(resid[-1]~variab)
model2$coef
```

Code in R of the HEGY test using the dataset of the inflation

```
library(uroot)
hegy.out<-HEGY.test (infl91, itsd=c(1,0,c(1,2,3,4,5,6,7,8,9,10,11)), regvar=0,
selectlags=list(mode="signf", Pmax=NULL))

hegy.out

summary(hegy.out)
```

References

1. **Definition of Stationarity** http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc442.htm
2. **James Douglas Hamilton (1994)** -Time Series Analysis
3. **Diaz-Emparanza I and Lopez-De-Lacalle J(2006)-**Testing for Unit Root in Seasonal Time Series with R: The uroot Package
4. **Σημειώσεις Ε. Ιωαννίδης (2005)-'**Ελεγχοιμοναδιαίας ρίζας I&II
5. **Dickey D.A, Fuller W.A (1979)-**Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*
6. **Using R for Time Series Analysis-**http://a-little-book-of-r-for-time-series.readthedocs.io/en/latest/src/timeseries.html
7. Forecasting Functions for Time Series and Linear Models -**Package "Forecast"**
8. **Unit Root Tests-** https://faculty.washington.edu/ezivot/econ584/notes/unitroot.pdf
9. **Wikipedia-** https://en.wikipedia.org
10. **Investopedia-**http://www.investopedia.com
11. **Engineering Statistics-**6.4.4.3 Seasonality- http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc443.htm
12. **Ghysels E. and Osborn D.R (2001)**- The Econometric Analysis of Time Series
13. **Brockwell P.J and Davis R.A (1991)** -Time Series: Theory and Methods
14. **Seasonal ARIMA-online courses-**PennStateEberly College of Science
15. **Ghysels E, Osborn D.R, Rodrigues P.M.M (1999)-**Seasonal Nonstationarity and Near-Nonstationarity, *CIRANO, Centre interuniversitaire de recherché en analyse des organization.*
16. **Dickey D.A, Fuller W.A, Hasza D.P (1984)-**Testing for Unit Roots in **Seasonal** Time Series. *Journal of the American Statistical Association*
17. **Hylleberg S, Engle R.F, Granger C.W.J, Yoo B.S (1990)** - Seasonal Integration and Cointegration, *Journal of Econometrics*
18. **Ghysels E, Lee H.S, Noh J (1994)-**Testing for unit roots in Seasonal Time Series- Some theoretical extension and a Monte Carlo investigation, *Journal of Econometrics*

19. **Beaulieu J.J, Miron J.A (1993)-**Seasonal Unit roots in aggregate U.S data, *Journal of Econometrics*

20. **Carlo De Michele and Harry Pavlopoulos (2007)-** On the Stationarity of Rainfall Time Series , *Technical Report No 232, November 2007*

21. **Ahamada I. and M. Boutahar (2002)-**Tests for covariance stationarity and white noise, with an application to euro/us dollar exchange rate. An approach based on the evolutionary spectral density, *Economics Letters*.

22. **Pagan A. and G. Schwert (1990)-** Testing for covariance stationarity in stoch market data, *Economics Letters*.

23. **Phillips P.C.B. (1987) -** Time series regression with a unit root, *Econometrica* , 55, 277-301.

24. **Priestley M.B, (1965) -** Evolutionary spectra and non-stationary processes, J. R. Statist. Soc, Series B, 27, 204-237.