Unit Root Tests in Finance

A Thesis

presented to

the Faculty of Natural and Applied Sciences

at Notre Dame University-Louaize

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Financial Mathematic

---

by

RIMA JREIJ

JULY 2020

Notre Dame University - Louaize

Faculty of Natural and Applied Science

Department of Mathematics and Statistics

We hereby approve the thesis of

Rima Jreij

Candidate for the degree of Master of Science

Dr. John Haddad	Supervisor, Chair
Dr. Roger Nakad	Committee Member
Dr. Remi El Hage	Committee Member
Dr. Full Name	Committee Member

[Signature]

Abstract of the Thesis

Unit Root Tests in Finance

by

Rima Jreij

Master of Science

in

Financial Mathematics

Notre Dame University-Louaize,

Zouk Mosbeh,
Lebanon

2020

The knowledge of whether a time series contains a unit root or not provides guidance to determine whether the series is stationary or not. This topic is one that covers vast amount of research given to its importance in the analysis of economic and other time series data. To understand the behavior, the properties of the series and the influence of any shock that occur to the series, stationary and unit root tests were constructed. In this thesis, we first present the Box and Jenkins ARMA models, discuss the conditions for stationarity. Then, we display different method to test autocorrelation.

And finally, we examine several unit root tests and discuss their power.

To my family, I would like to express my special gratitude for your moral support and encouragement which helped me in the completion of the thesis.

Contents

Acknowledgements

I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals. I would like to extend my sincere thanks to all of them.

I thank my God for providing me with everything that I required in completing this project.

I am highly indebted to my supervisor, Dr John Haddad for his guidance and constant supervision as well as for providing necessary information regarding the project and also for his support in completing the thesis.

Last but not least, I would like to thank my colleagues for sharing their academic experience, which provided me guidance and confidence.

Chapter 1

Introduction

A gambler's dispute in 1654 led to the creation of a mathematical theory of probability by two famous French mathematicians, Blaise Pascal and Pierre de Fermat. Antoine Gombaud, Chevalier de Méré, a French nobleman with an interest in gaming and gambling questions, called Pascal's attention to an apparent contradiction concerning a popular dice game. The game consisted in throwing a pair of dice 24 times; the problem was to decide whether or not to bet even money on the occurrence of at least one "double six" during the 24 throws. A seemingly well-established gambling rule led de Méré to believe that betting on a double six in 24 throws would be profitable, but his own calculations indicated just the opposite. This problem, which was posed by de Méré, led to an exchange of letters between Pascal and Fermat in which the fundamental principles of probability theory were formulated for the first time. Although a few special problems on games of chance had been solved by some Italian mathematicians in the 15th and 16th centuries [1], no general theory was developed before this famous correspondence.

The Dutch scientist Christian Huygens, teacher of Leibniz, learned of this correspondence and shortly thereafter (in 1657) published the first book on probability; entitled De Ratiociniis in Ludo Aleae ("On Reasoning in Games of Chance"), it was a treatise on problems associated with gambling. Because of the inherent appeal of games of chance, probability theory soon became popular, and the subject developed rapidly during the 18th century. The major contributors during this period were Jakob Bernoulli (1654-1705) and Abraham de Moivre (1667-1754).

In 1827, the Scottish botanist Robert Brown reported the observation of a very irregular motion displayed by a pollen particle immersed in a fluid. In 1900, Louis Bachelier give a solution to the Brownian motion, but it was refused by his advisor Henri Poincaré. Later, in 1905 and 1906, Einstein and Smoluchowski discovered the solution. Moreover, Bachelier showed in his thesis that the stock and option markets vary in a random manner and therefore are unpredictable. 60 years later, the Random Walk Theory was found by Paul Samuelson and Eugene Fama.

The theoretical developments in time series analysis started early with stochastic processes. The first actual application of auto regressive models to data can be brought back to the work of G. U Yule and J. Walker in the 1920s and 1930s.

During this time the moving average was introduced to remove periodic fluctuations in the time series, for example fluctuations due to seasonality. Herman Wold introduced ARMA (AutoRegressive Moving Average) models for stationary series, but was unable to derive a likelihood function to enable maximum likelihood (ML) estimation of the parameters.

It took until 1970 before this was accomplished. At that time, the classic book "Time Series Analysis" by G. E. P. Box and G. M. Jenkins came out, containing the full modeling procedure for individual series: specification, estimation, diagnostics and forecasting. Nowadays, the so-called Box-Jenkins models are perhaps the most commonly used and many techniques used for forecasting and seasonal adjustment can be traced back to these models. The first generalization was to accept multivariate ARMA models, among which especially VAR models (Vector AutoRegressive) have become popular. These techniques, however, are only applicable for stationary time series. However, especially economic time series often exhibit a rising trend suggesting non-stationarity, that is, a unit root.

A process is called a unit root process if the characteristic polynomial has at least a root equal to one. This process in not stationary. However, it is possible to make these processes stationary by differencing.
Testing for unit root is important in forecasting and many other fields to understand the behavior of the series and predict the impact of any shocks that may occur to the series. Unit root tests developed mainly during the 1980's. The first unit root test, the Dickey-Fuller test, was developed in 1979. Unfortunately, the DF test was only applicable for simple auto-regressive models (AR(1)). Later on, several tests were developed to accommodate more complicated models.

The aim of this thesis is to give an overview of time series analysis and to point out on an important issue which is the presence of a unit root in a time series and the problems accompanying this. The thesis is composed of 5 chapters. Chapter one gives a short history of probability and time series. Chapter

2 presents basic concepts of time series, stochastic processes, Stationary and non stationary processes and at the end discusses Auto regressive, Moving average and Auto regressive Moving average models and their Autocorrelation and Partial Autocorrelation functions (ACF and PACF). Chapter 3 gives an estimation of the sample autocorrelation functions (SACF)and shows different methods to construct a confidence interval for the ACF. Chapter 4 defines unit root processes, integrated and difference processes, then presents the most important unit root tests and last discusses their power. Finally, chapter 5 is a conclusion.

Chapter 2

Time Series Concepts

2.1 Introduction

In this chapter we will define some basic concepts in time series analysis and give a brief overview of some stochastic process that are mentioned frequently throughout the thesis.

In general, a time series is a sequence of data points indexed in time order.Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values.

A time series could be modeled using stochastic processes or deterministic processes. A stochastic process is simply a random process through time. A good way to think about it, is that a stochastic process is the opposite of a deterministic process.

In a deterministic process, given the initial conditions and the parameters of

the system, we can define the exact "position" of the system at any time. In a stochastic process, we don't know where the process will be, even if we know the initial conditions and parameters. Stochastic Analysis deals with models which involve uncertainties or randomness. Uncertainty, complexity and dynamism have been continuing challenges to our understanding and control of our physical environment. Everyday we encounter signals which cannot be modeled exactly by an analytic expression or in a deterministic way. Examples of such signals are ordinary speech waveforms, biological signals, temperature histories, communication signals etc. In manufacturing domain no machine is totally reliable. Every machine fails at some random time. Thus in a typical manufacturing system which involves a large number of machines, the total number of machines at any time cannot be determined in a deterministic way. In a market driven economy, the stock market is volatile, the interest rates fluctuate in a random fashion. One can give any number of examples from our daily life events where uncertainty prevails in an essential way. This gives us the realization that many real life phenomena require the analysis of a system in a probabilistic setting rather than in a deterministic setting. Thus stochastic models are becoming increasingly important for understanding or making performance evaluation of complex systems in a broad spectrum of fields.

2.2 Preliminaries

2.2.1 Mean and Variance

Definition:

- Let X be a discrete random variable and p_i= P(X=x_i), then the expected value E(X) of X is defined by

$$E(X) = \sum_{i=1}^{\infty} x_i p_i$$

provided that the sum $\sum_{i=1}^{\infty} x_i p_i$ absolutely converges. i.e, $\sum_{i=1}^{\infty} |x_i| p_i < \infty$.

- Let X be a continuous random variable and $f_X(x)$ be its probability density function, then the expected value $E(X)$ of X is defined by

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx$$

provided that the integral $\int_{-\infty}^{+\infty} |x| f_X(x) dx$ is finite (X is integrable).

- The variance of the random variable X is defined by

$$Var(X) = E[(X - E(X))^2]$$

2.2.2 Some important Stochastic Processes

• **Gaussian Process:** A stochastic process $\{X_t\}_{t=1}^{\infty}$ is said to be a Gaussian process if and only if for every finite set of indices $t_1, t_2, \cdots, t_k$, the vector $(X_{t_1}, X_{t_2}, \cdots, X_{t_k})$ is a multivariate Gaussian random variable X_t; i.e., every linear combination $a_1 X_{t_1} + a_2 X_{t_2} + \cdots + a_k X_{t_k}$ is normally distributed ($\sim \mathcal{N}(\mu, \sigma^2)$).

• **Gaussian White Noise:** A stochastic process $\{X_t\}_{t=1}^{\infty}$ is said to be a Gaussian White Noise (GWN) process if for all $t \leq 1$, the random variables X_t are i.i.d., and normally distributed with mean zero i.e., $X_t \sim \mathcal{N}(0, \sigma^2)$.

• **Brownian Motion(Weiner Process)** A stochastic process $\{X_t\}_{t=0}^{\infty}$ is said to be a Brownian motion if it satisfies the following properties:

1. $X_0 = 0$
2. The process has independent increments; i.e., $X_{t_2} - X_{t_1}$ and $X_{t_4} - X_{t_3}$ are independent for $0 \leq t_1 < t_2 \leq t_3 < t_4$.
3. The process has a Gaussian increments; i.e., for $t_1 < t_2$, $X_{t_2} - X_{t_1} \sim \mathcal{N}(0, t_2 - t_1)$.
4. The process is continuous.

2.2.3 Autocovariance Function and Autocorrelation Function

Definition: Let $\{X_t, t = 1, \ldots, T\}$ be a time series of $T < \infty$ observations, and $h \in \mathbb{N}$ such that $t + h < T$, then

• The **Autocovariance function** of X_t is defined by:

$$\gamma(h) = cov(X_t, X_t + h) = E[(X_t - \mu_t)(X_{t+h} - \mu_{t+h})] \tag{2.2.1}$$

$$= E(X_t X_{t+h}) - E(X_t)E(X_{t+h}) \tag{2.2.2}$$

The autocovariance function measures the direction of the linear dependence between the random variables X_t and X_{t+h}.

It's clear that the autocovariance at lag 0 of a variable X_t is equivalent to the variance:

$\gamma(0) = cov(X_t, X_t) = E[(X_t - \mu_t)^2] = var(X_t)$

- The **Autocorrelation function (ACF)** of x_t is defined by:

$$\rho(h) = corr(X_t, X_{t+h}) = \frac{cov(X_t, X_{t+h})}{\sqrt{var(X_t)}\sqrt{var(X_{t+h})}} = \frac{\gamma(h)}{\gamma(0)}. \qquad (2.2.3)$$

The autocorrelation function of X_t measures the direction and the strength of the linear dependence between X_t and X_{t+h}.

It's clear that the autocorrelation at lag 0 of a variable X_t is equal to 1: $\rho(0) = \frac{\gamma(0)}{\gamma(0)} = 1$.

2.3 Stationary and non-Stationary Processes

A process $\{x_t\}_{t=1}^{\infty}$ is said to be:

- **Stationary process:** if the joint distribution of the random variables is time invariant.

- **Weakly Stationary process:**

a- $E(x_t) = \mu$ does not depend on t.

b- $var(x_t) = \sigma^2$

c- $\gamma(h) = cov(x_t, x_{t+h})$ exists, finite and does not depend on t.

• **Stationary increments:** if the distribution of the increments $x_{t+h} - x_t$ are independent of t.

2.4 Time Series Processes

In time series analysis, the moving-average model (MA model), also known as moving-average process, is a common approach for modeling univariate time series. The moving-average model specifies that the output variable depends linearly on the current and various past values of a stochastic (imperfectly predictable) term. The autoregressive model (AR model) specifies that the output variable depends linearly on its own previous values and on a stochastic term (an imperfectly predictable term); thus the model is in the form of a stochastic difference equation. Together with the autoregressive (AR) model, the moving-average model is a special case and key component of the more general ARMA and ARIMA models of time series, which have a more complicated stochastic structure.

2.4.1 Moving Average Process MA(q)

A q-order moving average process, denoted **MA(q)** takes the form
$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$ (this process can also be expressed as:
$Y_t = X_t - \mu = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$).
The value of X at time t+1 is a linear function of past errors. We assume that the error terms are independently distributed with a normal distribution with

mean zero and a constant variance σ^2. Thus,

$\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ and $cov(\varepsilon_i, \varepsilon_j) = 0$ if $i \neq j$.

- **Property 1:** The mean of an MA(q) process is μ.

  Proof: $E(X_t) = \mu + E(\varepsilon_t) + \theta_1 E(\varepsilon_{t-1}) + ... + \theta_q E(\varepsilon_{t-q})$

  $= \mu + 0 + \theta_1 \cdot 0 + \cdots + \theta_q \cdot 0 = \mu.$

- **Property 2:** The variance of an MA(q) process is

  $var(X_t) = \sigma^2(1 + \theta_1^2 + \cdots + \theta_q^2)$.

  Proof: $var(X_t) = 0 + var(\varepsilon_t) + \theta_1^2 var(\varepsilon_{t-1}) + \cdots + \theta_q^2 var(\varepsilon_{t-q})$

  $= \sigma^2 + \theta_1^2\sigma^2 + \cdots + \theta_q^2\sigma^2 = \sigma^2(1 + \theta_1^2 + \cdots + \theta_q^2)$.

- **Property 3:** The autocorrelation function of an MA(1) process is

  $$\rho_k = \begin{cases} \frac{\theta_1}{1+\theta_1^2} \text{ if } k = 1 \\ 0 \text{ if } k \geq 2 \end{cases}$$

  Proof: An MA(1) process takes the form

  $X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$

  $\gamma(0) = cov(X_t, X_t) = var(X_t) = \sigma^2(1 + \theta_1^2)$ **(property 2).**

  $\gamma(1) = cov(X_t, X_{t+1}) = cov(\varepsilon_t + \theta_1\varepsilon_{t-1}, \varepsilon_{t+1} + \theta_1\varepsilon_t)$

  $= cov(\varepsilon_t, \theta_1\varepsilon_t) = \theta_1 cov(\varepsilon_t, \varepsilon_t) = \theta_1\sigma^2.$

  $\gamma(2) = cov(X_t, X_{t+2}) = cov(\varepsilon_t + \theta_1\varepsilon_{t-1}, \varepsilon_{t+2} + \theta_1\varepsilon_{t+1}) = 0.$

  $(cov(\varepsilon_i, \varepsilon_j) = 0$ **for** $i \neq j)$

  $\gamma(k) = cov(X_t, X_{t+k}) = cov(\varepsilon_t + \theta_1\varepsilon_{t-1}, \varepsilon_{t+k} + \theta_1\varepsilon_{t+k-1}) = 0$ for $k > 3$.

Thus, $\rho_1 = \frac{\gamma(1)}{\gamma(0)} = \frac{\theta_1\sigma^2}{(1+\theta_1^2)\sigma^2} = \frac{\theta_1}{1+\theta_1^2}$

and $\rho_k = \frac{\gamma(k)}{\gamma(0)} = 0$ for $k \geq 2$.

- **Property 4:** The autocorrelation function of an MA(2) process is

$$\rho_k = \begin{cases} \frac{\theta_1+\theta_1\theta_2}{1+\theta_1^2+\theta_2^2} \text{ if } k = 1 \\ \frac{\theta_2}{1+\theta_1^2+\theta_2^2} \text{ if } k = 2 \\ 0 \text{ if } k \geq 3 \end{cases}$$

Proof: An MA(2) process takes the form

$X_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}$

$\gamma(0) = cov(X_t, X_t) = var(X_t) = \sigma^2(1 + \theta_1^2 + \theta_2^2)$ **(property 2)**.

$\gamma(1) = cov(X_t, X_{t+1}) = cov(\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}, \varepsilon_{t+1} + \theta_1\varepsilon_t + \theta_2\varepsilon_{t-1})$

$= \theta_1 cov(\varepsilon_t, \varepsilon_t) + \theta_1\theta_2 cov(\varepsilon_{t-1}, \varepsilon_{t-1}) = \theta_1\sigma^2 + \theta_1\theta_2\sigma^2 = (\theta_1 + \theta_1\theta_2)\sigma^2$.

$\gamma(2) = cov(X_t, X_{t+2}) = cov(\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}, \varepsilon_{t+2} + \theta_1\varepsilon_{t+1} + \theta_2\varepsilon_t)$

$= \theta_2 cov(\varepsilon_t, \varepsilon_t) = \theta_2\sigma^2$.

$\gamma(k) = cov(X_t, X_{t+k}) = cov(\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}, \varepsilon_{t+k} + \theta_1\varepsilon_{t+k-1} +$

$\theta_2\varepsilon_{t+k-2}) = 0$ for $k \geq 3$.

Thus,

$\rho_1 = \frac{\gamma(1)}{\gamma(0)} = \frac{(\theta_1+\theta_1\theta_2)\sigma^2}{(1+\theta_1^2\theta_2^2)+\sigma^2} = \frac{\theta_1+\theta_1\theta_2}{1+\theta_1^2+\theta_2^2}$.

$\rho_2 = \frac{\gamma(2)}{\gamma(0)} = \frac{\theta_2\sigma^2}{(1+\theta_1^2+\theta_2^2)\sigma^2} = \frac{\theta_2}{1+\theta_1^2+\theta_2^2}$.

$\rho_k = \frac{\gamma(k)}{\gamma(0)} = 0$ for $k \geq 3$.

- **Property 5:** The autocorrelation function of an MA(q) process is

$$\rho_h = \begin{cases} \frac{\theta_h+\sum_{j=1}^{q-h}\theta_j\theta_{j+h}}{1+\sum_{j=1}^{q}\theta_j^2}, \text{if } h \leq q \\ 0, \text{if } h > q \end{cases}$$

Proof: As in the case of MA(1) and MA(2),

we first calculate $\gamma(1), \gamma(2), \cdots, \gamma(q), \cdots$

$\gamma(1) = cov(X_t, X_{t+1})$

$= cov(\varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q},\ \varepsilon_{t+1} + \theta_1\varepsilon_t + \cdots + \theta_q\varepsilon_{t-q+1})$

$= (\theta_1 + \theta_1\theta_2 + \cdots + \theta_q\theta_{q+1})\sigma^2$

$\gamma(2) = cov(X_t, X_{t+2})$

$= cov(\varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q},\ \varepsilon_{t+2} + \theta_1\varepsilon_{t+1} + \cdots + \theta_q\varepsilon_{t-q+2}$

$= (\theta_2 + \theta_2\theta_4 + \cdots + \theta_{q-2}\theta_q)\sigma^2$

$\vdots$

$\gamma(q) = cov(X_t,\ X_{t+q})$

$= cov(\varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q},\ \varepsilon_{t+q} + \theta_1\varepsilon_{t+q-1} + \cdots + \theta_q\varepsilon_t)$

$= \theta_q\sigma^2$

Now, $\gamma_{q+h} = cov(X_t, X_{t+q+h})$

$= cov(\varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q}, \varepsilon_{t+q+h} + \theta_1\varepsilon_{t+q+h-1} + \cdots + \theta_q\varepsilon_{t+h}) = 0$

Thus, $\gamma(h) = cov(X_t,\ X_{t+h}) = \begin{cases} (\theta_h + \sum_{j=1}^{q-h}\theta_j\theta_{j+h})\sigma^2, & \text{if } h \le q \\ 0, & \text{if } h > q \end{cases}$

Note that, $\gamma(0) = cov(X_t,\ X_t) = var(X_t) = (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma^2$

Consequently, $\rho_h = \begin{cases} \frac{\theta_h + \sum_{j=1}^{q-h}\theta_j\theta_{j+h}}{1+\sum_{j=1}^{q}\theta_j^2}, \text{if } h \le q \\ 0, \text{if } h > q \end{cases}$

- **Property 6:** A **finite** MA(q) process is **stationary**.

  Proof: Let $X_t = \varepsilon_t + \theta_1\varepsilon_{t-11} + \cdots + \theta_q\varepsilon_{t-q}$

  In this case, $E(X_t) = 0$ and $var(X_t) = (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma^2$ do not depend on t.

$\gamma(h) = cov(X_t,\, X_{t+h}) = \begin{cases} (\theta_h + \sum_{j=1}^{q-h} \theta_j \theta_{j+h})\sigma^2, & \text{if } h \leq q \\ 0, & \text{if } h > q \end{cases}$ is finite and does not depend on t. Thus, the process is stationary.

Definition: A process X_t is said to be invertible **if** the random disturbance at time t can be expressed as a convergent sum of present and past values of X_t in the form :

$\varepsilon_t = \sum_{j=0}^{\infty} \theta_j X_{t-j}$ where $\sum |\theta_j| < \infty$.

We now define the lag function:

$B^i X_t = X_{t-i}$; for example, $BX_t = X_{t-1}$ and $B^2 X_t = X_{t-2}$.

For a moving average model of orders q, the equation can be re-written using the backshift operator B:

$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$

Then, $X_t = (1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t$

A process is **invertible** if and only if all the roots of

$\beta_0 + \beta_1 B + \beta_2 B^2 + \cdots + \beta_q B^q = 0$ all lie outside the unit circle.

Example: Let $X_t = \varepsilon_t + \theta \varepsilon_{t-1}$

$= (1 + \theta B)\varepsilon_t$

$1 + \theta B = 0 \leftrightarrow B = \frac{-1}{\theta}$

Thus, this process is invertible iff $|B| > 1 \leftrightarrow |\frac{-1}{\theta}| > 1 \leftrightarrow |\theta| < 1$.

2.4.2 Auto Regressive Process AR(p)

A p-order auto regressive process, denoted **AR(p)** takes the form

$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \cdots + \alpha_p Y_{t-p} + \varepsilon_t$;

The value of Y at time t is a linear function of Y at earlier times plus a fixed

constant and a random error term.

We assume that the error terms are independently distributed based on a normal distribution with zero mean and a constant variance σ^2 and that the error terms are independent of the Y values (**i.e** $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, $cov(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$ and $cov(\varepsilon_i, Y_j) = 0$ for all i, j).

A first-order auto regressive process, denoted **AR(1)**, takes the form

$y_t = \alpha_0 + \alpha_1 Y_{t-1} + \varepsilon_t$.

We see that the value of Y at time $t+1$ is a linear function of Y at time t plus a fixed constant and a random error term.

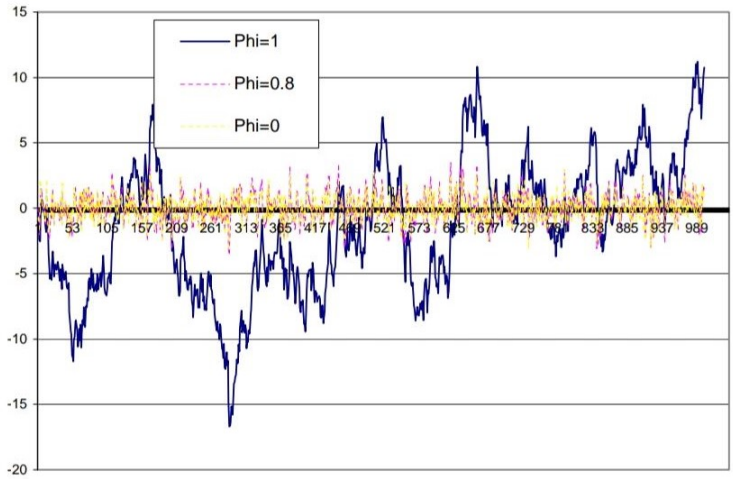


Figure 2.1: Autoregressive Process with differing values of $\phi(0, 0.8, 1)$.

- **Property:**

  An AR(1) process is stationary if $|\alpha_1| < 1$.

  Proof: AR(1): $Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \varepsilon_t$

We can rewrite AR(1) as : $Y_t - \alpha_0 = \alpha_1 Y_{t-1} + \varepsilon_t$

$Y_t - \alpha_0 = \alpha_1(\alpha_1 Y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$

$= \alpha_1^2(\alpha_1 Y_{t-3} + \varepsilon_{t-2}) + \alpha_1\varepsilon_{t-1} + \varepsilon_t$

$= \alpha_1^3(\alpha_1 Y_{t-4} + \varepsilon_{t-3}) + \alpha_1^2\varepsilon_{t-2} + \alpha_1\varepsilon_{t-1} + \varepsilon_t$

$= \varepsilon_t + \alpha_1\varepsilon_{t-1} + \alpha_1^2\varepsilon_{t-2} + \alpha_1^3\varepsilon_{t-3} + \cdots$

$E(Y_t - \alpha_0) = E(Y_t) - \alpha_0 = E(\varepsilon_t) + \alpha_1 E(\varepsilon_{t-1}) + \alpha_1^2 E(\varepsilon_{t-2}) + \alpha_1^3 E(\varepsilon_{t-3}) +$

$\cdots = 0$

Therefore, $E(Y_t) = \alpha_0$.

$\gamma(0) = var(Y_t - \alpha_0) = var(Y_t) = cov(Y_t - \alpha_0,\, Y_t - \alpha_0)$

$= cov(\varepsilon_t + \alpha_1\varepsilon_{t-1} + \alpha_1^2\varepsilon_{t-2} + \alpha_1^3\varepsilon_{t-3} + \cdots ,\, \varepsilon_t + \alpha_1\varepsilon_{t-1} + \alpha_1^2\varepsilon_{t-2} + \alpha_1^3\varepsilon_{t-3} + \cdots)$

$= \sigma^2 + \alpha_1^2\sigma^2 + \alpha_1^4\sigma^2 + \cdots = (1 + \alpha_1^2 + \alpha_1^4 + \cdots)\sigma^2$

$= \frac{\sigma^2}{1-\alpha_1^2}$

First note that for any constant a, $cov(a + x, a + y) = cov(x, y)$

Then, $\gamma(1) = cov(Y_t,\, Y_{t+1}) = cov(Y_t - \alpha_0,\, Y_{t+1} - \alpha_0)$

$= cov(\varepsilon_t + \alpha_1\varepsilon_{t-1} + \alpha_1^2\varepsilon_{t-2} + \alpha_1^3\varepsilon_{t-3} + \cdots ,\, \varepsilon_{t+1} + \alpha_1\varepsilon_t + \alpha_1^2\varepsilon_{t-1} + \alpha_1^3\varepsilon_{t-2} + \cdots)$

$= \alpha_1\sigma^2 + \alpha_1^3\sigma^2 + \alpha_1^5\sigma^2 + \cdots$

$= (\alpha_1 + \alpha_1^3 + \alpha_1^5 + \cdots)\sigma^2$

$= \frac{\alpha_1\sigma^2}{1-\alpha_1^2}$

$\vdots$

$\gamma(k) = \frac{\alpha_1^k\sigma^2}{1-\alpha_1^2}$ (by induction on k).

Then, $\rho_k = \frac{\gamma(k)}{\gamma(0)} = \alpha_1^k$ for $k = 0, 1, 2, \cdots$

when $|\alpha_1| < 1$, the autocovariance do not explode as k increases.

Thus, the process is stationary if $|\alpha_1| < 1$.

Note that if $|\alpha_1| = 1$, we have a random walk.

- **Property:** The mean of the Y_t in a stationary AR(p) process is

  $\mu = \frac{\alpha_0}{1-\sum_{j=1}^{p} \alpha_j}$

  Proof: Since the process is stationary, for any k, $E[Y_t] = E[Y_{t-k}]$, a value which we will denote μ. Since $E[\varepsilon_t] = 0,\ E[\alpha_0] = \alpha_0$ and

  $Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \cdots + \alpha_p Y_{t-p} + \varepsilon_t$

  It follows that

  $\mu = E[Y_t] = E[\alpha_0] + \alpha_1 E[Y_{t-1}] + \cdots + \alpha_p E[Y_{t-p}] + E[\varepsilon_t]$

  $= \alpha_0 + \alpha_1 \mu + \cdots + \alpha_p \mu$

  $\mu = \alpha_0 + \mu(\alpha_1 + \cdots + \alpha_p)$

  $\mu(1 - \sum_{j=1}^{p} \alpha_j) = \alpha_0$

  Thus, $\mu = \frac{\alpha_0}{1-\sum_{j=1}^{p} \alpha_j}$.

- **Property:** For any stationary AR(p) process. The autocovariance at lag $k > 0$ can be calculated as

  $\gamma_k = \alpha_1 \gamma_{k-1} + \alpha_2 \gamma_{k-2} + \cdots + \alpha_p \gamma_{k-p}$

  Similarly the autocorrelation at lag $k > 0$ can be calculated as

  $\rho_k = \alpha_1 \rho_{k-1} + \alpha_2 \rho_{k-2} + \cdots + \alpha_p \rho_{k-p}$

  Here we assume that $\gamma_h = \gamma_{-h}$ and $\rho_h = \rho_{-h}$ if $h < 0$, and $\rho_0 = 1$. These are known as **the Yule-Walker equations**.

- **Property:** The Yule-Walker equations also hold where $k = 0$ provided we add a σ^2 term to the sum. This is equivalent to

  $\gamma_0 = \alpha_1 \gamma_1 + \alpha_2 \gamma_2 + \cdots + \alpha_p \gamma_p + \sigma^2$

  For example, is the case of AR(1): $\rho_0 = 1; \rho_1 = \alpha_1 \rho_0 = \alpha_1;$

  $\rho_2 = \alpha_1 \rho_1 = \alpha_1 \alpha_1 = \alpha_1^2; \cdots; \rho_k = \alpha_1 \rho_{k-1} = \alpha_1^k.$

  And, $\gamma_0 = \alpha_1 \gamma_1 + \sigma^2 = \alpha_1 \rho_1 \gamma_0 + \sigma^2 = \alpha_1^2 \gamma_0 + \sigma^2$

Then, $(1-\alpha_1^2)\gamma_0 = \sigma^2$ and $\gamma_0 = \frac{\sigma^2}{1-\alpha_1^2}$

Now, in the AR(2) case, we have:

$\rho_0 = 1$

$\rho_1 = \alpha_1\rho_0 + \alpha_2\rho_1 \leftrightarrow (1-\alpha_2)\rho_1 = \alpha_1 \leftrightarrow \rho_1 = \frac{\alpha_1}{1-\alpha_2}$

$\rho_2 = \alpha_1\rho_1 + \alpha_2\rho_0 = \alpha_1\frac{\alpha_1}{1-\alpha_2} + \alpha_2 = \frac{\alpha_1^2}{1-\alpha_2} + \alpha_2$

now for the variance, $\gamma_0 = \alpha_1\gamma_1 + \alpha_2\gamma_2 + \sigma^2 = \alpha_1\rho_1\gamma_0 + \alpha_2\rho_2\gamma_0 + \sigma^2$

Then, $(1-\alpha_1\rho_1-\alpha_2\rho_2)\gamma_0 = \sigma^2 \leftrightarrow \gamma_0 = \frac{\sigma^2}{1-\frac{\alpha_1^2}{1-\alpha_2}-(\frac{\alpha_2\alpha_1^2}{1-\alpha_2}+\alpha_2^2)} = \frac{(1-\alpha_2)\sigma^2}{(1-\alpha_2)-\alpha_1^2-\alpha_2\alpha_1^2-(1-\alpha_2)\alpha_2^2}$.

2.4.3 Auto Regressive Moving Average Process ARMA(p,q)

An autoregressive moving average (ARMA) process consists of both autoregressive and moving average terms. If the process has terms from both an AR(p) and MA(q) process, then the process is called ARMA(p, q) and can be expressed as

$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}$

where $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$.

We can define an ARMA(p, q) process with zero mean by removing the constant term (i.e. ϕ_0) and saying that $X_1, X_2, \cdots, X_n$ has an ARMA(p, q) process with mean μ if the time series $z_1, \ldots, z_n$ has an ARMA(p, q) process with zero mean where $z_i = X_i - \mu$.

If we include the constant term, then as in the AR(p) case, for a stationary ARMA(p, q) process

$\mu = \frac{\phi_0}{1-\sum_{j=1}^{p}\phi_j}$.

An equivalent expression for an ARMA(p, q) process with zero mean is

$X_t - \sum_{j=1}^{p}\phi_j X_{t-j} = \varepsilon_t + \sum_{j=1}^{q}\theta_j\varepsilon_{t-j}$

Or $\phi(B)X_t = \theta(B)\varepsilon_t$.

It must be noted that in this representation, both the polynomial AR and the polynomial MA should not have any common factors. This will ensure that there is no parameter redundancy. Should a common factors exist, it will introduce wrong representations of time dependency. The following example will show how parameter redundancy can occur:

Assume a simple model,

$X_t = \varepsilon_t$

$0.5X_{t-1} = 0.5\varepsilon_{t-1}$

with a simple subtraction manipulation, we obtain the following:

$X_t = 0.5X_{t-1} + \varepsilon_t - 0.5\varepsilon_{t-1}$

which seems to indicate ARMA(1,1) model, but in actual fact it is falsely misrepresented. Thus, one should eliminate any common factors between the AR and MA polynomials to establish the correct time dependency representation.

- A linear process $\{X_t\}$ is **causal** if there is a

$\psi(B) = \psi_0 + \psi_1 B + \psi_2 B^2 + \cdots$

with $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and $X_t = \psi(B)\varepsilon_t$

And it is **invertible** if there is a

$\pi(B) = \pi_0 + \pi_1 B + \pi_2 B^2 + \cdots$

with $\sum_{j=0}^{\infty} |\pi_j| < \infty$ and $\varepsilon_t = \pi(B)X_t$.

Theorem 2.4.1. *If ϕ and θ have no common factors,a (unique) stationary solution to $\phi(B)Xt = \theta(B)\varepsilon_t$ exists if and only if the roots of $\phi(z)$ avoid the*

unit circle: $|z| = 1 \rightarrow \phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p \neq 0.$

This process is causal if and only if the roots of $\phi(z)$ *are outside the unit circle:*

$|z| \leq 1 \rightarrow \phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p \neq 0.$

It is invertible if and only if the roots of $\theta(z)$ *are outside the unit circle:*

$|z| \leq 1 \rightarrow \theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q \neq 0.$

Example 1 : Let $x_t = \alpha x_{t-1} + \varepsilon_t + \beta \varepsilon_{t-1}$

a- Identify the process.

b- Under what condition(s) the process is stationary? is invertible?

Solution:

a- The process is an ARMA(1,1);

$x_t - \alpha x_{t-1} = \varepsilon_t + \beta \varepsilon_{t-1}.$

$(1 - \alpha B) x_t = (1 + \beta B) \varepsilon_t.$

b- The process is stationary :

if the roots of $(1 - \alpha B) = 0$ lie outside the unit circle

equivalent to $|B| = |\frac{1}{\alpha}| > 1 \leftrightarrow |\alpha| < 1.$

The process is invertible :

if the roots of $1 + \beta B = 0$ all lie outside the unit circle

equivalent to $|B| = |\frac{-1}{\beta}| > 1 \leftrightarrow |\beta| < 1.$

Example 2 : Let $x_t = \frac{1}{3} x_{t-1} + \frac{2}{9} x_{t-2} + \varepsilon_t$; Is this process stationary?

Solution: $(1 - \frac{1}{3}B - \frac{2}{9}B^2)x_t = \varepsilon_t$

$(1 - \frac{1}{3}B - \frac{2}{9}B^2) = 0$

$B_1 = \frac{3}{2}$ and $B_2 = 3$, then $|B_1| > 1$ and $|B_2| > 1$ Thus, the process is stationary.

Partial Autocorrelation Function (PACF)

We have seen earlier that the autocorrelation function of MA(q) models is zero for all lags greater than q as these are q-correlated processes. Hence, the ACF is a good indication of the order of the process. However AR(p) and ARMA(p,q) processes are " fully " correlated, their ACF tails off and never becomes zero, though it may be very close to zero. In such cases it is difficult to identify the process on the ACF basis only. In this section we will consider another correlation function, which together with the ACF will help to identify the models. The function is called Partial Autocorrelation Function (PACF). Before introducing a formal definition of PACF we motivate the idea for AR(1).

Let $X_t = \phi X_{t-1} + \varepsilon_t$ be a causal AR(1) process. Then

$\gamma(2) = cov(X_t, X_{t-2})$

$= cov(\phi X_{t-1} + \varepsilon_t, X_{t-2})$

$= cov(\phi(\phi X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t, X_{t-2}) = cov(\phi^2 X_{t-2} + \phi\varepsilon_{t-1} + \varepsilon_t, X_{t-2})$

$= \phi^2 cov(X_{t-2}, X_{t-2}) = \phi^2\gamma(0)$

The autocorrelation is not zero because X_t depends on X_{t-2} through X_{t-1}. Due to the iterative kind of AR models there is a chain of depen-

dence. We can break this dependence by removing the influence of X_{t-1} from both X_t and X_{t-2} to obtain:

$X_t - \phi X_{t-1}$ and $X_{t-2} - \phi X_{t-1}$ for which the covariance is zero, i.e.,

$cov(X_t - \phi X_{t-1},\, X_{t-2} - \phi X_{t-1}) = cov(\varepsilon_t,\, X_{t-2} - \phi X_{t-1}) = 0.$

Similarly, we obtain zero covariance for X_t and X_{t-3} after breaking the chain of dependence, i.e. removing the dependence of the two variables on X_{t-1} and X_{t-2}, i.e. for $X_t - f(X_{t-1}, X_{t-2})$ and $X_{t-3} - f(X_{t-1}, X_{t-2})$ for some function f. Continuing this we would obtain zero covariances for variables $X_t - f(X_{t-1}, X_{t-2}, \cdots, X_{t-\tau+1})$ and $X_{t-\tau} - f(X_{t-1}, X_{t-2}, \cdots, X_{t-\tau+1})$ then the only nonzero covariance is for X_t and X_{t-1}. These covariances with an appropriate function f divided by the variance of the process are the **partial autocorrelation**. Hence, for a causal AR(1) process we would have the PACF at lag 1 equal to $\rho(1)$ and at $lags > 1$ equal to 0.

In other words, to calculate ϕ_{kk}, let:

$X_{t+k} = \phi_{k1} X_{t+k-1} + \phi_{k2} X_{t+k-2} + \cdots + \phi_{kk} X_t + \varepsilon_{t+k}$

then, multiplying X_{t+k} by X_{t+k-i} for $i = 1, \cdots, k$ and taking expectations we obtain:

$E[X_{t+k} X_{t+k-i}] = \phi_{k1} E[X_{t+k-1} X_{t+k-i}] + \phi_{k2} E[X_{t+k-2} X_{t+k-i}] + \cdots + \phi_{kk} E[X_t X_{t+k-i}] + E[\varepsilon_{t+k} X_{t+k-i}]$

Then, $\gamma_i = \phi_{k1}\gamma_{i-1} + \phi_{k2}\gamma_{i-2} + \cdots + \phi_{kk}\gamma_{i-k}$

Therefore, $\rho_i = \phi_{k1}\rho_{i-1} + \phi_{k2}\rho_{i-2} + \cdots + \phi_{kk}\rho_{i-k}$ for $i = 1, \cdots, k$

Now, for $k = 1$ we have $\rho_1 = \phi_{11}\rho_0 = \phi_{11}$

For $k = 2$, we have:

for $i = 1, \rho_1 = \phi_{21}\rho_0 + \phi_{22}\rho_1$

for $i = 2, \rho_2 = \phi_{21}\rho_1 + \phi_{22}\rho_0$

Therefore, $\phi_{22} = \dfrac{\begin{vmatrix} \rho_0 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} \rho_0 & \rho_1 \\ \rho_1 & \rho_0 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$

For $k = 3$, we have:

For $i = 1$, $\rho_1 = \phi_{31}\rho_0 + \phi_{32}\rho_1 + \phi_{33}\rho_2$

For $i = 2$, $\rho_2 = \phi_{31}\rho_1 + \phi_{32}\rho_0 + \phi_{33}\rho_1$

For $i = 3$, $\rho_3 = \phi_{31}\rho_1 + \phi_{32}\rho_0 + \phi_{33}\rho_0$

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}$$

In the same way, we calculate ϕ_{kk} for all k.

Example: The PACF of AR(1)

Consider the AR(1) process

$X_t = \alpha X_{t-1} + \varepsilon_t$ Where $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$

Where $|\alpha| < 1$, i.e., a causal AR(1). We have seen before that $\rho_k = \alpha^k$;

then, $\phi_{11} = \rho_1 = \alpha$

$\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \frac{\alpha^2 - \alpha^2}{1 - \alpha^2} = 0$

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} = \frac{\begin{vmatrix} 1 & \alpha & \alpha \\ \alpha & 1 & \alpha^2 \\ \alpha^2 & \alpha & \alpha^3 \end{vmatrix}}{\begin{vmatrix} 1 & \alpha & \alpha^2 \\ \alpha & 1 & \alpha \\ \alpha^2 & \alpha & 1 \end{vmatrix}} = 0$$

Thus, $\phi_{kk} = \begin{cases} \alpha, & \text{if } k = 1 \\ 0, & \text{if } k \geq 2 \end{cases}$

Example: The PACF of AR(2)

Consider the AR(2) process

$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \varepsilon_t$ Where $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$

We know that $\rho_0 = 1;\ \rho_1 = \frac{\alpha_1}{1-\alpha_2};\ \rho_2 = \alpha_1\rho_1 + \alpha_2\rho_0$

Then,

$\phi_{11} = \rho_1 = \frac{\alpha_1}{1-\alpha_2}$

$\phi_{22} = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2} = \frac{\alpha_1\frac{\alpha_1}{1-\alpha_2} + \alpha_2 - (\frac{\alpha_1}{1-\alpha_2})^2}{1-(\frac{\alpha_1}{1-\alpha_2})^2} = \frac{\alpha_1^2(1-\alpha_2)+\alpha_2(1-\alpha_2)^2-\alpha_1^2}{(1-\alpha_2)^2-\alpha_1^2} = \frac{-\alpha_1^2\alpha_2+\alpha_2(1-\alpha_2)^2}{(1-\alpha_2)^2-\alpha_1^2}$

$\phi_{kk} = 0$ for $k \geq 3$

Now for **MA(q)**,

the ACF of an MA(1) is given by

$\rho_k = \begin{cases} \frac{\theta}{1+\theta^2}, \text{if } k = 1 \\ 0, \text{if } k \geq 2 \end{cases}$

Using $\rho_k = 0$ for $k > 1$ we can show that the PACF of MA(1) is

$\phi_{kk} = \frac{\theta^k(1-\theta^2)}{1-\theta^{2(k+1)}}$ for $k > 1$

Contrary to its ACF, which cuts off after lag 1, the PACF of an MA(1) model decays exponentially.

For a general MA(q) process, the ACF " cuts down " to zero after lag q and the PACF will have exponential behavior depending on the characteristic roots of

$\Theta(B) = (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q) = 0$

The PACf of ARMA(p,q) model:

An invertible ARMA model has an infinite AR representation, hence the PACF will not cut off. The following table summarizes the behavior of the PACF of the causal and invertible ARMA models

	AR(p)	**MA(q)**	**ARMA(p,q)**
ACF	Decays exponentially after lag p	cuts off after lag q	exponential decay starts after max(p,q)
PACF	cuts off after lag p	decays exponentially after lag q	exponential decay starts after max (p,q)

Chapter 3

Estimation of the Auto-Correlation Function

3.1 Introduction

In chapter 2, we have seen that the basic characterization of an ARMA (p,q) process is in its ACF and PACF functions.

Here, we consider the estimation of both quantities by their sample counterpart. Mainly, we will investigate the probability distribution of the sample correlation and hence a confidence interval for a function of the ACF will be constructed. We conclude with a strategy for ARMA (p,q) identification.

3.2 Point Estimation:

The sample auto correlation coefficient, SACF, based on a realization of n measurements from the process $\{X_t\}$, $\underline{X} = (X_1, \cdots, X_n)'$, is defined as the

following:

$$r_k = \frac{\sum_{t=1}^{n-k}(X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^{n-k}(X_{t+k} - \bar{X})^2} \tag{3.2.1}$$

and the k^{th}-lag SPACF, $\hat{\phi}_{kk}$, is given by

$\hat{\phi}_1 = r_1$ and

$$\hat{\phi}_{kk} = \frac{\begin{vmatrix} 1 & r_1 & \cdots & r_k \\ r_1 & 1 & \cdots & r_{k-1} \\ \vdots & \vdots & \vdots & \vdots \\ r_{k-1} & r_{k-2} & \cdots & r_1 \end{vmatrix}}{\begin{vmatrix} 1 & r_1 & \cdots & r_{k-1} \\ r_1 & 1 & \cdots & r_{k-2} \\ \vdots & \vdots & \vdots & \vdots \\ r_{k-1} & r_{k-2} & \cdots & r_1 \end{vmatrix}} \tag{3.2.2}$$

r_k is the sample auto correlation coefficient of the random variables X_t and X_{t+k}. This may be considered as a direct application to the estimation of the population correlation coefficient in bivariate normal random variables. In this context, let $\{(x_i,\, y_i), i = 1, \cdots, n\}$ be a random sample from a bivariate normal distribution with the joint probability function

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} exp \frac{-1}{2(1-\rho^2)} \{(\frac{x-\mu_x}{\sigma_x})^2 - 2\rho(\frac{x-\mu_x}{\sigma_x})(\frac{y-\mu_y}{\sigma_y}) + (\frac{y-\mu_y}{\sigma_y})^2\} \tag{3.2.3}$$

where $\mu_x = E(x), \mu_y = E(y), \sigma_x = \sqrt{var(x)}, \sigma_y = \sqrt{var(y)}$ and ρ is the population correlation coefficient.

It is in common to estimate ρ by the sample correlation which is known as

Pearson's product moment:

$$\hat{\rho} = \frac{1}{n-1}\sum_{i=1}^{n-1}(\frac{x_i - \bar{x}}{S_x})(\frac{y_i - \bar{y}}{S_y}) \tag{3.2.4}$$

where $\bar{x} = \sum_{i=1}^{n} x_i/n$, $\bar{y} = \sum_{i=1}^{n} y_i/n$, $S_x^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/n - 1$ and $S_y^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2/n - 1$.

Fisher(1915) obtained the following representation of the exact density function of $\hat{\rho}$ given in equation (3.2.4)

$$f_\rho(x) = \frac{2^{n-3}}{\pi(n-3)!}(1-\rho^2)^{(n-1)/2}(1-x^2)^{(n-2)/4}\sum_{i=0}^{\infty}\Gamma^2(\frac{n-i+1}{2})\frac{(2\rho x)^i}{i!} \tag{3.2.5}$$

Where $-1 < x < 1$. The series representation in equation (3.2.5) converges very slowly. Moreover, the normal distribution approximation of $\hat{\rho}$ is markedly skewed when n is small and $|\rho|$ is large. Winterbottom (1979) showed that normal approximation of the distribution of $\hat{\rho}$ requires large sample size n and the variance of $\hat{\rho}$ changes with the mean $E(\hat{\rho})$. We shall survey the population most useful methods to construct a $(1-\alpha)\%$ confidence interval for ρ.

3.3 Fisher's z-Transformation:

Fisher's z-transformation also known as variance stabilizing transformation. It's given by $z = \frac{1}{2}log(\frac{1+r}{1-r}) = tanh^{-1}$. Fisher(1915) has shown that z has an asymptotic distribution with mean $\zeta = \frac{1}{2}log\frac{1+\rho}{1-\rho}$ and variance $\frac{1}{n-3}$. Therefore, a $(1-\alpha)\%$ confidence interval for z is given by

$$\left(tanh(z - \frac{z_{\alpha/2}}{\sqrt{n-3}}), tanh(z + \frac{z_{\alpha/2}}{\sqrt{n-3}})\right) \tag{3.3.1}$$

where z_γ is the γ^{th} upper quantile of a standard normal distribution.

3.4 Hotelling's Approximations:

Hotelling gave four modification of Fisher's z-transformation:

$z_1 = z - \frac{7z+r}{8(n-1)}, \zeta_1 = \zeta - \frac{7\zeta+\rho}{8(n-1)}$

$z_2 = z - \frac{7z+r}{8(n-1)} - \frac{119z+57r+3r^2}{384(n-1)^2}, \zeta_2 = \zeta - \frac{7\zeta+\rho}{8(n-1)} - \frac{119\zeta+57\rho+3\rho^2}{384(n-1)^2}$

$z_3 = z - \frac{3z+r}{4(n-1)}, \zeta_3 = \zeta - \frac{3\zeta+\rho}{4(n-1)}$

$z_4 = z - \frac{3z+r}{4(n-1)} - \frac{23z+33r-5r^2}{96(n-1)^2}, \zeta_4 = \zeta - \frac{3\zeta+\rho}{4(n-1)} - \frac{23\zeta+33\rho-\rho^2}{96(n-1)^2}$

and showed that the $z_i \sim \mathcal{N}(\zeta_i, \frac{1}{n-1})$ for $i = 1 \cdots 4$. Consequently, we can construct four confidence intervals for ρ based on these approximations. However, there is no closed form for each of these confidence intervals, and they are obtained numerically.

3.5 Ruben's Approximation:

Ruben (1996) shows that for large n,

$z_{hr} = \frac{(\frac{2n-5}{2})^{\frac{1}{2}}\bar{r} - (\frac{2n-3}{2})^{\frac{1}{2}}\bar{\rho}}{(1+\frac{1}{2}(\bar{r}^2+\bar{\rho}^2))^{1/2}}$ where $\bar{r} = \frac{r}{\sqrt{(1-r^2)}}$ and $\bar{\rho} = \frac{\rho}{\sqrt{(1-\rho^2)}}$ is asymptotically distributed as a standard normal distribution. Therefore, we can construct a confidence interval for ρ numerically based on this approximation.

3.6 Muddapur's Method:

Muddapur(1988) considered the test statistic:

$f = \frac{(1+r)(1-\rho)}{(1-r)(1+\rho)}$,

He showed that it has an approximate F distribution with n-2 and n-2 degrees of freedom. Note that f is related to Fisher's z-transform through:

$log(f) = 2(z - \zeta)$,

Therefore, a $100(1-\alpha)\%$ confidence interval for ρ is

$\left(\frac{(1+F_{\alpha/2})r+(1-F_{\alpha/2})}{(1+F_{\alpha/2})+(1-F_{\alpha/2})r}, \frac{(1+F_{\alpha/2})r-(1-F_{\alpha/2})}{(1+F_{\alpha/2})-(1-F_{\alpha/2})r}\right)$

where F_γ is the γ^{th} upper quantile of the F distribution with n-2 and n-2 degrees of freedom.

3.7 Haddad and Provost's Method:

Let $D^+ = \sum_{i=1}^{n}(X_{i1}^* + X_{i2}^*)^2$ and $D^- = \sum_{i=1}^{n}(X_{i1}^* - X_{i2}^*)^2$, where $X_{ij}^* = \frac{X_{ij}-\bar{X}_i}{S_i}$, $i = 1, 2$, $j = 1, \cdots, n$ are the standard values. Haddad and Provost's (2011) proposed an approximately $100(1-\alpha)\%$ confidence interval for ρ as $\left(\frac{D^+-D^-F^*_{\alpha/2}}{D^++D^-F^*_{\alpha/2}}, \frac{D^+-D^-F^*_{1-\alpha/2}}{D^++D^-F^*_{1-\alpha/2}}\right)$ where F_γ^* is the γ^{th} upper quantile of the F distribution with n-1 and n-1 degrees of freedom.

In conclusion, a basic and important concern in science is to test the hypothesis of uncorrelation versus significant correlation between two quantitative variables. That is the hypothesis of $H_0 : \rho = 0$ vs. $H_a : \rho \neq 0$. Although transform methods seem to stabilize the probability distribution of the sample

correlation coefficient, there is a good deal of interest in the probability distribution of the sample correlation coefficient. The methods that have been considered in this chapter show powerful tests for the uncorrelation hypothesis even when n is small as show in the article of Kazemi and Jafari [2]. Unfortunately, these methods fall short of answering the hypothesis of linear dependence, that is, $H_0 : \rho = 1$. However, there is a model that is capable of testing both hypothesis is the autoregressive process model. Thus if one considers the AR(1) model: $X_t = \phi X_{t-1} + w_t$, then the hypothesis of $\phi = 0$ can be easily answered through the methods of this chapter, while the hypothesis of $\phi = 1$ will be dealt with in the next chapter.

Chapter 4

Unit Root Tests

4.1 Introduction:

A unit root process is a linear stochastic process that has a unit root, i.e. the process characteristic equation has a root equal to one. Such process is non-stationary.

In probability theory and statistics, this may cause problems in statistical inference involving time series models: for non-stationary series, it's impossible to predict the behavior of the series.

It's important to distinguish between a trend stationary process and a unit root process. Note that, both processes are non-stationary. However, the impact of a shock on both series are different: in the case of a trend stationary process, the time series will converge again towards the growing mean, while in the case of a unit root process, the impact of a shock will be permanent. If a root of the process's characteristic equation is larger than 1, then it is called an explosive process.

The presence of a unit root can be tested using a unit root test.

This chapter is organized as follows. In the section 4.2, we are going to define the difference and integrated processes. The sections 4.3 till 4.7 describe some popular unit root tests. Section 4.8 discusses some problems associated with traditional unit root and stationarity tests and at the end section 4.9 will talk in brief on the importance of unit root tests in finance.

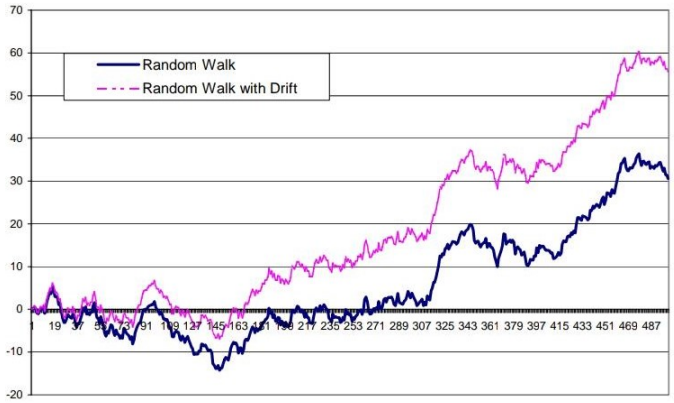


Figure 4.1: Random Walk Processes with and without drift.

Remark: As it is mentioned in the introduction it's important to differentiate between two important cases:

• A non stationary process with a deterministic trend where Y_t is a function of time. $Y_t = \alpha + \beta t + \varepsilon_t$ (β is a parameter, $\varepsilon_t \sim \mathcal{WN}(0, \sigma^2)$, and α a constant). Such process become stationary by detrending i.e. subtract Y_t and Y_{t-1} to obtain $\Delta Y_t = \beta + (\varepsilon_t - \varepsilon_{t-1})$ which is stationary.

• A process with a stochastic trend i.e. Y_t is a function of the past value. Let's consider an AR(1) process

$Y_t = c + Y_{t-1} + \varepsilon_t$ Where c is a constant and $\varepsilon_t \sim \mathcal{WN}(0, \sigma^2)$

As long as, $|\phi| < 1$, the process is stationary (chapter 2). In this chapter we will discuss the case where $|\phi| = 1$.

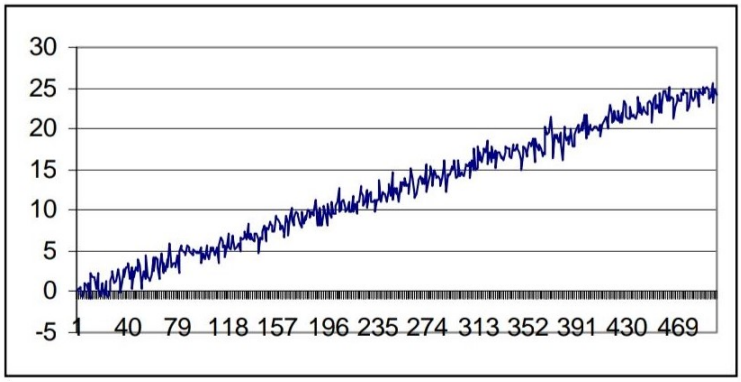


Figure 4.2: A Deterministic Trend Process.

4.2 Difference and Integrated Processes

In the Box-Jenkins approach to analyzing time series, a key question is whether to difference the data , i.e., to replace the raw data $\{X_t\}$ by the differenced series $\{X_t - X_{t-1}\}$, or not. Experience indicates that most economic time series tend to wander and are not stationary, but that differencing often yields a stationary result. A key example, which often provides a fairly good description of actual data, is the random walk, $X_t = X_{t-1} + \varepsilon_t$, where ε_t is white noise, assumed here to be independent, each having the same distribution. The random walk is said to have a **unit root**.

To understand more what this means, let's recall the condition of stationarity of an AR(p) model. We said in Chapter 2 that an AR(p) process of the form $x_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p}$ is stationary if the roots of $(1 - \alpha_1 z - \alpha_2 z^2 - \cdots - \alpha_p z^p) = 0$ lie outside the unit circle.

In the case of a Random Walk which is an AR(1): $X_t = \alpha X_{t-1} + \varepsilon_t$ with $\alpha = 1$. Indeed, for an AR(1) to be stationary, it is necessary that all the roots of the equation $1 - \alpha z = 0$ lie outside the unit circle. We see that the AR(1) is stationary if and only if $-1 < \alpha < 1$. For the Random Walk, we have a unit root, that is, a root equal to one. The first difference of a random walk is Stationary since $X_t - X_{t-1} = \varepsilon_t$ a white noise process.

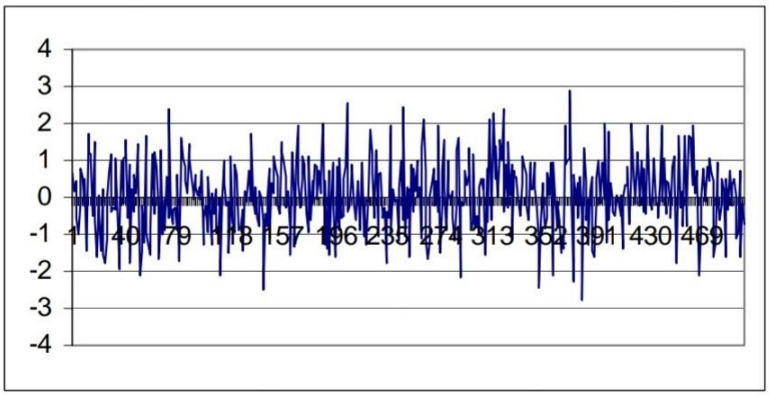


Figure 4.3: A White Noise Process.

4.2.1 Integrated process

Definition:

The series is **Integrated of order d**, denoted by **I(d)**, where d is an integer

with $d \geq 1$, if the series and all its differences up to the $(d-1)^{th}$ are non stationary, but the d^{th} difference is stationary.

A series is said to be integrated of order zero, denoted by I(0), if the series is both stationary and invertible.

if the series $\{X_t\}$ is I(d) with $d \geq 1$, then $\{X_t - X_{t-1}\}$ is I(d-1).

Example: Consider an AR(2) process

$X_t = 2X_{t-1} - X_{t-2} + \varepsilon_t$;

Then, the process is stationary if all the roots of the equation $(1 - 2z + z^2) = (z-1)^2 = 0$ lie outside the unit circle. In this example, the process has two unit roots, thus the process is not stationary. Also, the first difference is not stationary : $X_t - X_{t-1} = 2X_{t-1} - X_{t-2} + \varepsilon_t - 2X_{t-2} + X_{t-3} - \varepsilon_{t-1} = 2X_{t-1} - 3X_{t-2} + X_{t-3} + \varepsilon_t - \varepsilon_{t-1}$. The roots of the equation $2z - 3z^2 + z^3 = 0$ are $z_1 = \frac{-3-\sqrt{17}}{2}, z_2 = 0, z_3 = \frac{-3+\sqrt{17}}{2}$. $|z_2| < 1$ and $|z_3| < 1$ do not lie outside the unit circle. Consequently, the first difference is not stationary.

Now, the second difference is $(X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) = X_t - 2X_{t-1} + X_{t-2}$ which is equal to ε_t by definition of the given AR(2). Thus the second difference is stationary. Consequently, $\{x_t\}$ is I(2).

4.3 Dickey-Fuller and Augmented Dickey-Fuller(ADF) Tests

Named for American statisticians David Dickey and Wayne Fuller, who developed the test in 1979, the Dickey-Fuller test is used to determine whether a unit root is present in an autoregressive model. It is the simplest approach to

test for a unit root, but most economic and financial times series have a more complicated and dynamic structure than what can be captured by a simple autoregressive model, which is where the augmented Dickey-Fuller test comes into play.

With a basic understanding of that underlying concept of the Dickey-Fuller test, it is not difficult to jump to the conclusion that an augmented Dickey-Fuller test (ADF) is just that: an augmented version of the original Dickey-Fuller test. In 1984, the very same statisticians expanded their basic autoregressive unit root test (the Dickey-Fuller test) to accommodate more complex models with unknown orders (the augmented Dickey-Fuller test).

Similar to the original Dickey-Fuller test, the augmented Dickey-Fuller test is one that tests for a unit root in a time series sample. The test is used in statistical research and econometric, or the application of mathematics, statistics, and computer science to economic data.

4.3.1 Dickey-Fuller Test:

The early work on testing for a unit root in time series was done by Dickey and Fuller (Dickey and Fuller 1979, Fuller 1976). The basic objective of the test is to test the null hypothesis that $\phi = 1$ in $Y_t = \phi Y_{t-1} + \varepsilon_t \;\; (\varepsilon_t \sim \mathcal{WN}(0, \sigma^2))$ against the one-sided alternative $\phi < 1$. The hypotheses of interest are:

$H_0 : \phi = 1$ (unit root in $\phi(z) = 0) \Rightarrow Y_t \sim I(1)$

$H_1 : |\phi| < 1 \Rightarrow Y_t \sim I(0)$

Subtracting Y_{t-1} from both sides, we can rewrite the AR(1) model as:

$\Delta(Y_t) = Y_t - Y_{t-1} = (\phi - 1)y_{t-1} + \varepsilon_t = \delta Y_{t-1} + \varepsilon_t.$

Now a test of $\phi = 1$ is a simple t-test of whether the parameter on the lagged level of y is equal to zero.

$H_0 : \delta = 0 \Rightarrow Y_t \sim I(1)$

$H_1 : |\delta| < 0 \Rightarrow Y_t \sim I(0)$

There are three versions of the Dickey-Fuller unit root test:

1- Test for a unit root without drift (constant) and without trend:

$\Delta(X_t) = \delta X_{t-1} + \varepsilon_t$

2- Test for a unit root with drift only: $\Delta X_t = a_0 + \delta X_{t-1} + \varepsilon_t$

3- Test for a unit root with drift and deterministic time trend:

$\Delta X_t = a_0 + a_1 t + \delta X_{t-1} + \varepsilon_t$

The test statistic is: $t_{\phi=1} = \frac{\hat{\phi}-1}{SE(\hat{\phi})} = \frac{\hat{\delta}}{SE(\hat{\delta})}$

where $\hat{\phi}$ is the least squares estimate and SE($\hat{\phi}$) is the usual standard error estimate.

If $\{Y_t\}$ is stationary ($|\phi| < 1$) then it can be shown (c.f Hamilton (1994) pg. 216)

$\sqrt{T}(\hat{\phi} - \phi) \xrightarrow{d} \mathcal{N}(0,\, (1 - \phi^2))$ [1]. Then, $\hat{\phi} \overset{A}{\sim} \mathcal{N}(\phi,\, \frac{1}{T}(1 - \phi^2))$

And it follows that $t_{\phi=1} \overset{A}{\sim} \mathcal{N}(0,\, 1)$. However, under the null hypothesis of non-stationarity the above result gives $\hat{\phi} \overset{A}{\sim} \mathcal{N}(1,\, 0)$

which clearly does not make any sense. The problem is that under the unit root null, $\{Y_t\}$ is not stationary and ergodic , and the usual sample moments do not converge to fixed constants. Instead, Phillips (1987) showed that the sample moments of $\{Y_t\}$ converge to random functions of Brownian and that under the null hypothesis $H_0 : \phi = 1$:

$T(\hat{\phi} - 1) \xrightarrow{d} \frac{\int_0^1 W(r)dW(r)}{\int_0^1 W(r)^2 dr}$ and then, $t_{\phi=1} \xrightarrow{d} \frac{\int_0^1 W(r)dW(r)}{(\int_0^1 W(r)^2 dr)^{1/2}}$

[1] the sign $\xrightarrow{d}$ means converges in distribution

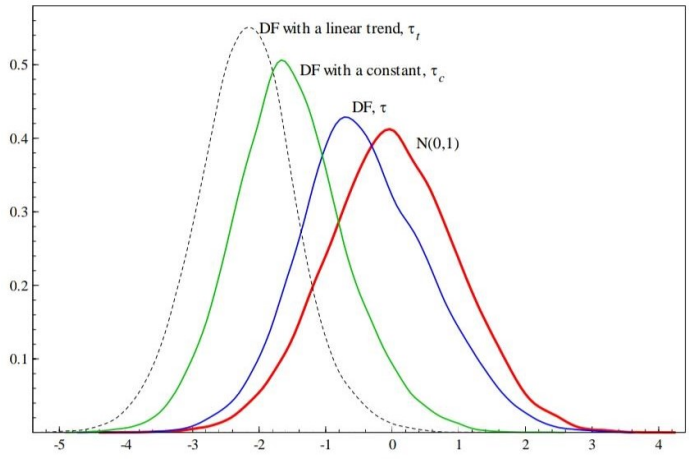


Figure 4.4: The Dickey-Fuller Distributions.

Consequently, $\hat{\phi}$ is not asymptotically normally distributed and $t_{\phi=1}$ is not asymptotically standard normal.

The limiting distribution of $t_{\phi=1}$ is called the Dickey-Fuller(DF) distribution and does not have a closed form representation. Critical values are derived from Monte Carlo experiments in, for example, Fuller (1976).

The null hypothesis of a unit root is rejected in favor of the stationary alternative in each case if the test statistic is smaller than the critical value.

Probability to the Right of Critical Value:

Sample size	**0.99**	**0.975**	**0.95**	**0.99**	**0.10**	**0.05**	**0.025**	**0.01**
25	-2.66	-2.26	-1.95	-1.60	0.92	1.33	1.70	2.16
50	-2.62	-2.25	-1.95	-1.61	0.91	1.31	1.66	2.08
100	-2.60	-2.24	-1.95	-1.61	0.90	1.29	1.64	2.03
250	-2.58	-2.23	-1.95	-1.62	0.89	1.29	1.63	2.01
500	-2.58	-2.23	-1.95	-1.62	0.89	1.28	1.62	2.00
∞	-2.58	-2.23	-1.95	-1.62	0.89	1.28	1.62	2.00

Table 4.1: Testing $\rho = 1$ in $Y_t = \rho Y_{t-1} + \varepsilon_t$

Sample size	0.99	0.975	0.95	0.99	0.10	0.05	0.025	0.01
25	-3.75	-3.33	-3.00	-2.62	-0.37	0.00	0.34	0.72
50	-3.58	-3.22	-2.93	-2.60	-0.40	-0.03	0.29	0.66
100	-3.51	-3.17	-2.89	-2.58	-0.42	-0.05	0.26	0.63
250	-3.46	-3.14	-2.88	-2.57	-0.42	-0.06	0.24	0.62
500	-3.44	-3.13	-2.87	-2.57	-0.43	-0.07	0.24	0.61
∞	-3.43	-3.12	-2.86	-2.57	-0.44	-0.07	0.23	0.60

Table 4.2: Testing $\rho = 1$ in $Y_t = \alpha + \rho Y_{t-1} + \varepsilon_t$

Sample size	0.99	0.975	0.95	0.99	0.10	0.05	0.025	0.01
25	-4.38	-3.95	-3.60	-3.24	-1.14	-0.80	-0.50	-0.15
50	-4.15	-3.80	-3.50	-3.18	-1.19	-0.87	-0.58	-0.24
100	-4.04	-3.73	-3.45	-3.15	-1.22	-0.90	-0.62	-0.28
250	-3.99	-3.69	-3.43	-3.13	-1.23	-0.92	-0.64	-031
500	-3.98	-3.68	-3.42	-3.13	-1.24	-0.93	-0.65	-0.32
∞	-3.96	-3.66	-3.41	-3.12	-1.25	-0.94	-0.66	-0.33

Table 4.3: Testing $\rho = 1$ in $Y_t = \alpha + \beta t + \rho Y_{t-1} + \varepsilon_t$

4.3.2 Augmented Dickey-Fuller Test(ADF):

The Dickey-Fuller test described above is only used for an AR(1) process with white noise errors. However, many time series have a more complicated structure. Said and Dickey (1984) augment the basic autoregressive unit root test to accommodate general ARMA(p,q) models with unknown orders and this test is called Augmented Dickey-Fuller(ADF).

The testing procedure for the ADF test is the same as for the Dickey–Fuller test : the hypothesis that a time series Y_t is I(1) against the alternative that it is I(0) assuming that Y_t is an ARMA process. To test for a unit root using the ADF test, one estimates the following model:

$Y_t = \alpha + \beta t + \phi Y_{t-1} + \sum_{i=1}^{p} \delta_i \Delta Y_{t-i} + \varepsilon_t$; (Appendix A)

An alternative formulation of the ADF test regression is

$\Delta Y_t = \alpha + \beta t + \pi Y_{t-1} + \sum_{i=1}^{p} \delta_i \Delta Y_{t-i} + \varepsilon_t$ where $\pi = \phi - 1$;

where α is a constant,β the coefficient on a time trend and the p lagged difference terms, ΔY_{t-i}, are used to approximate the ARMA structure of the errors, and the value of p is set so that the error ε_t is serially uncorrelated. The error term is also assumed to be homoskedastic. [2] the ADF t-statistic is given by:

$ADF_t = t_{\phi=1} = \frac{\hat{\phi}-1}{SE(\hat{\phi})} = \frac{\hat{\pi}}{SE(\hat{\pi})}$.

Under the null hypothesis, ΔY_t is I(0) which implies that $\pi = 0$. The ADF t-statistic is then the usual t-statistic for testing $\pi = 0$.

The selection of n is made such that the residual term in the model is approximately white noise (e.g. has a mean of zero, a finite variance, and is not serially correlated). The choice of lag length is important. If too few lags are included, there will be remaining auto correlation and size distortion, but if too many lags are included then the power of the test will suffer. In practice, there are several approaches to determining maximum lag length.

Ng. and Perron (1995) suggested to set an upper bound p_{max} for p (in practice often equal to Schwert's rule of thumb) and then estimate the ADF test regression with $p = p_{max}$. If $|t_{\pi(p)}| > 1.6$, set $p = p_{max}$ and perform the ADF test. Otherwise, reduce the lag length by one and repeat the process.

Schwert's (1989) rule of thumb for determining p_{max} is

$p_{max} = \left[12(\frac{T}{100})^{1/4}\right]$ (p_{max} is the integer part).

In general, a p-value of less than 5% means you can reject the null hypothesis that there is a unit root. You can also compare the calculated ADF_t statistic with a tabulated critical value. If the ADF_t statistic is more negative than the table value, reject the null hypothesis of a unit root.

[2] Homoskedastic refers to a condition in which the variance of the error term in a regression model is constant.

Note that, the more negative the ADF test statistic, the stronger the evidence for rejecting the null hypothesis of a unit root.

4.4 Phillips-Perron Unit Root Tests:

The Phillips-Perron (PP) unit root tests differ from the ADF tests mainly in how they deal with serial correlation and heteroskedasticity [3] in the errors. The ADF test seeks to approximate the ARMA structure of the errors in the test regression. The PP tests correct for any serial correlation and heteroskedasticity in the errors non-parametrically by modifying the Dickey Fuller test statistics.

The test regression: $\Delta Y_t = \alpha + \beta t + \pi Y_{t-1} + u_t$ where u_t is I(0) and may be heteroskedastic. The modified statistics denoted Z_t and Z_π are given by:

$Z_t = \left(\frac{\hat{\sigma}^2}{\hat{\lambda}^2}\right)^{\frac{1}{2}} .t_{\pi=0} - \frac{1}{2} \left(\frac{\hat{\lambda}^2 - \hat{\sigma}^2}{\hat{\lambda}^2}\right) . \left(\frac{T.SE(\hat{\pi})}{\hat{\sigma}^2}\right)$;

$Z_\pi = T_{\hat{\pi}} - \frac{1}{2} . \frac{T^2 . SE(\hat{\pi})}{\hat{\sigma}^2} (\hat{\lambda}^2 - \hat{\sigma}^2)$;

The terms $\hat{\sigma}^2$ and $\hat{\lambda}^2$ are consistent estimates of the variance parameters:

$$\sigma^2 = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E\left[u_t^2\right]$$

$$\lambda^2 = \lim_{T \to \infty} \sum_{t=1}^{T} E\left[T^{-1} S_T^2\right]$$

[3] Heteroskedastic refers to a condition in which the variance of the residual term, or error term, in a regression model varies widely.

where $S_T = \sum_{t=1}^{T} u_t$. Under $H_0 : \pi = 0$, the PP Z_t and Z_π statistics have the same asymptotic distributions as the DF t-statistic and normalized bias statistics. One advantage of the PP tests over the ADF tests is that the PP tests are robust to general forms of heteroskedasticity in the error term u_t. Also, the PP test does not require one to specify a model and to select the number of lags.

4.5 Kwiatkowski Phillips Schmidt Shin (KPSS) Stationarity Test:

Unlike the ADF and PP tests, the KPSS test has the null hypothesis of stationarity around a deterministic trend and the alternative of a unit root. Kwiatkowski, Phillips, Schmidt and Shin(1992) assume that the series is decomposed into the sum of a deterministic trend, a random walk, and a stationary error:

$$Y_t = \zeta t + r_t + \varepsilon_t,$$

$$r_t = r_{t-1} + u_t, u_t \sim \mathcal{N}(0, \sigma_u^2)$$

where r_t is a pure random walk and r_0 is assumed to be a constant level. Under the null hypothesis, Y_t is assumed to be I(0) so that $\sigma_u^2 = 0$ and $r_t = r_0$. The KPSS test statistic is the Lagrange multiplier (LM) for testing $\sigma_u^2 = 0$ against

the alternative $\sigma_u^2 > 0$, and it's given by: (for proof see [3])

$$KPSS = \frac{\left(T^{-2}\sum_{t=1}^{T}\hat{S}_t^{\,2}\right)}{\hat{\sigma_\varepsilon}^{\,2}}$$

where $\hat{S}_t = \sum_{i=1}^{t}\hat{\varepsilon}_i$, $\hat{\varepsilon}_t$ is the residual of a regression of Y_t on the trend t and $\hat{\sigma_\varepsilon}^{\,2}$ is a consistent estimate of the long run variance of the residuals ε_t using $\hat{\varepsilon}_t$. They showed that this test converges to a function of the standard Brownian motion depending on the form of t but not on their coefficient ζ:
For t=1,$(Y_t = \zeta + r_t + \varepsilon_t)$, then,

$$KPSS \xrightarrow{d} \int_0^1 W(r) - rW(1)dr$$

for $t \in (1, t)$, then,

$$KPSS \xrightarrow{d} \int_0^1 \left[W(r) + r(2-3r)W(1) + 6r(r^2-1)\left(\int_0^1 W(s)ds\right)\right]dr$$

where W(r) is a standard Brownian motion for $r \in [0, 1]$. Accordingly, critical values are calculated by a direct simulation in a Monte Carlo experiment.

	0.90	**0.925**	**0.950**	**0.975**	**0.99**
Constant	0.349	0.396	0.446	0.592	0.762
Linear Trend	0.120	0.133	0.149	0.184	0.229

Table 4.4: Quantiles of the distribution of the KPSS statistic.

In order to reject the null hypothesis, the test statistic should be greater than the provided critical values.

4.6 DF-GLS Test:

The DF-GLS test (ERS test) is a test for a unit root in an economic time series sample. It was developed by Elliott, Rothenberg and Stock in 1996 as a modification of the augmented Dickey-Fuller test (ADF). To increase the power of a unit root test under the null hypothesis of a unit root, ERS proposed a local to unity detrending of the time series. The assumed generating process for the series y_t is as follows:

$$y_t = \beta^{'} D_t + u_t, u_t = \phi u_{t-1} + v_t,$$

where D_t represents a vector of deterministic terms and v_t is a stationary zero-mean process. If $\phi = 1$ the y_t is I(1), but if $|\phi| < 1$ then y_t is I(0). The authors developed feasible point-optimal tests, which take serial correlation of the error term into account. The feasible point-optimal test statistic is defined as:

$$P_t = \frac{(S(\bar{\phi}) - \bar{\phi} S(1))}{\hat{\lambda}^2};\ \bar{\phi} = 1 + \bar{c}/T;$$

where $\hat{\lambda}^2$ is **a consistent estimate of the long run variance of the residuals** v_t and $S(\phi)$ **is the sum of squared residuals** from a least-squares regression of y_ϕ on D_ϕ with

$$y_\phi = (y_1, y_2 - \phi y_1, \cdots, y_T - \phi y_{T-1})'$$

and

$$D_\phi = (D_1, D_2 - \phi D_1, \cdots, D_T - \phi D_{T-1})'$$

and y_ϕ is a T-dimensional vector and D_ϕ is a $(T \times q)$ matrix (T number of observations).

S_ϕ is defined as follows:

$$S_\phi = \bar{y_\phi}' \bar{y_\phi};$$

$$\bar{y}_\phi = y_\phi - D_\phi \hat{\beta}_\phi \; and \; \hat{\beta}_\phi = (D_\phi' D_\phi)^{-1} D_\phi' y_\phi.$$

ERS showed that if $D_t = 1$ (case of a constant) then $\bar{c} = -7$, and if $D_t = (1, t)$ (case of a linear trend) then $\bar{c} = 13.5$;

Now applying the detrended data , $y_t^d = y_t - \hat{\beta}_{\bar{\phi}}' D_t$, in the ADF test regression to get the ADF-GLS test regression:

$$\Delta y_t^d = \pi y_{t-1}^d + \sum_{j=1}^{p} \psi_j \Delta y_{t-j}^d + \varepsilon_t;$$

Then, compute the t-statistic for testing $\pi = 0$. The authors showed that the ADF-GLS test and ADF t-test have the same asymptotic distribution in the case of $D_t = 1$. But the first has higher asymptotic power (against local alternatives). For $D_t = (1, t)$ the asymptotic distribution of the DF-GLS test, however, is different from the ADF t-test. The critical values for the ADF-GLS test obtained on the basis of simulation in Elliot, Rothenberg and Stock (1996) show that for models without constant they are the same as in the case of the ADF test. For the remaining models, critical values of the ADF-GLS test are used as indicated in Elliot, Rothenberg and Stock (1996) as well.

4.7 NGP Test

Ng and Perron (1995, 2001) applied the detrended data Δy_t^d obtained from the ADF-GLS test and modified the Phillips-Perron PP test.

The tests are defined as:

$$\overline{MZ_\phi} = (T^{-1}\Delta y_t^d - \hat{\lambda}^2)\left(2T^{-2}\sum_{t=1}^{T}\Delta y_{t-1}^d\right)^{-1},$$

$$\overline{MSB} = \left(T^{-2}\sum_{t=1}^{T}\Delta y_{t-1}^d/\hat{\lambda}^2\right)^{1/2},$$

$$\overline{MZ_t} = \overline{MZ_\phi} \times \overline{MSB},$$

Where $\hat{\lambda}^2$ is estimated by Ng and Perron(2001) from the ADF-GLS test regression as: $\hat{\lambda}^2 = \left[\sum_{t=p+1}^{T}\varepsilon_t^2\right]\left[(T-p)(1-\sum_{j=1}^{p}\hat{\psi}_j)^2\right]^{-1}$.

4.8 Problems associated with unit root tests and Power comparison of the unit root tests mentioned above :

As Fedorová and Arltová (2016) [4] mentioned in their article that all the unit tests stated above have a common problem that they depend on the length of analyzed time series as stated in Pesaran(2015) and Zivot(2006). The ADF and PP tests are known to suffer potentially severe finite sample power and size problems (PP test is based on an asymptotic theory, i.e. it is designed to test the unit roots in long time series). Pesaran and Zivot also point out

another problem that the power of these tests is low if the process is stationary but with a root close to the non-stationary boundary (where the parameter ϕ in the AR(1) is close to one).

For example, the test suffer(especially in small sample sizes) to reject the null of a unit root in the case of stationary process with $|\phi| = 0.95 < 1$.

The authors also stated that according to Caner and Killian (2001), the KPSS also suffer from similar issues. Moreover, it was proved that the power of these tests is lower in the case where a linear deterministic trend is included in the model of test regression. In contrast, the ADF-GLS and NGP tests should eliminate this problem. Nevertheless, complicated construction and also the fact that they are practically not represented in statistical and econometric software make their application difficult.

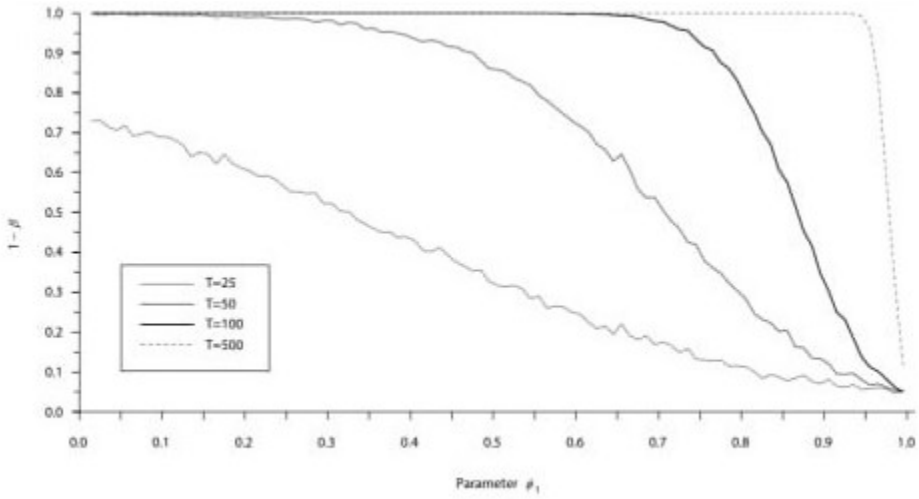
Power comparison:

The results below were obtained from a simulation study by Fedorová(2016) where the criteria were the length of time series and value of parameter ϕ_1 in the autoregressive process AR(1) without drift. The power of the test $(1 - \beta)$ is the probability of rejecting the null hypothesis when it is false. Note that for the KPSS test whose hypotheses are defined reversely, the probability $(1 - \alpha)$ is the probability of not rejecting the null when it is true.

All tests were conducted on a chosen 5% significance level. The simulation study was performed in the statistical software R, R Development Core Team (2008).

The ADF test:

The power of the test decreases gradually when ϕ_1 increases from 0 to 1 in the case of T=25 (i.e. in the case of very short time series). The test is unable to



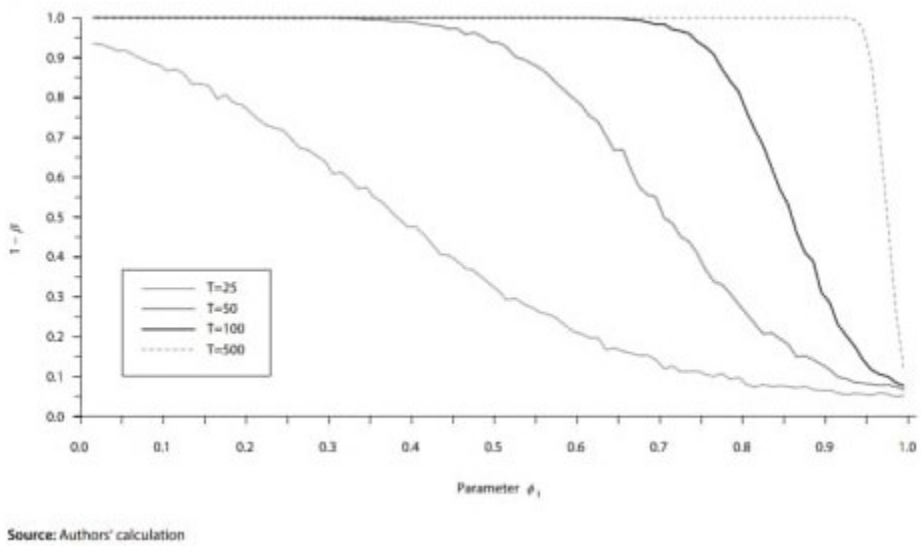
Source: Authors' calculation

Figure 4.5: Power functions $(1-\beta)$ of ADF test for simulated time series of the length T=25,50,100,500 and number of replications n=3000, $\alpha = 0.05$.

prove the stationarity even when ϕ_1 is small: the maximum success achieved is about 70% when ϕ_1 is too small $(0 < \phi_1 < 0.15)$. For $\phi_1 > 0.3$, the test more probably is not going to reject the null of a unit root despite the stationarity of time series. As the length of the time series increase, the power of the test is improved: the sucess of the test reach 1 for $\phi_1 < 0.2$ in the case of(T=50), $\phi_1 < 0.7$ when T=100 and for large time series (T=500), the success is almost 100% when $\phi_1 < 0.9$.

It seems clear that the power of the test increases along with the growing length of time series, but even in the case of time series with 500 observations, which uniquely identifies the test stationarity for $\phi_1 < 0.9$, its power function drops sharply for higher values of parameter .

The PP test: The power of the test remains very low especially for small time series. For T=25, the power of the test is greater than 0.8 for small value of ϕ_1 (less than 0.2). The test reach almost 1 when $\phi_1 < 0.4$ and $\phi_1 < 0.7$ for



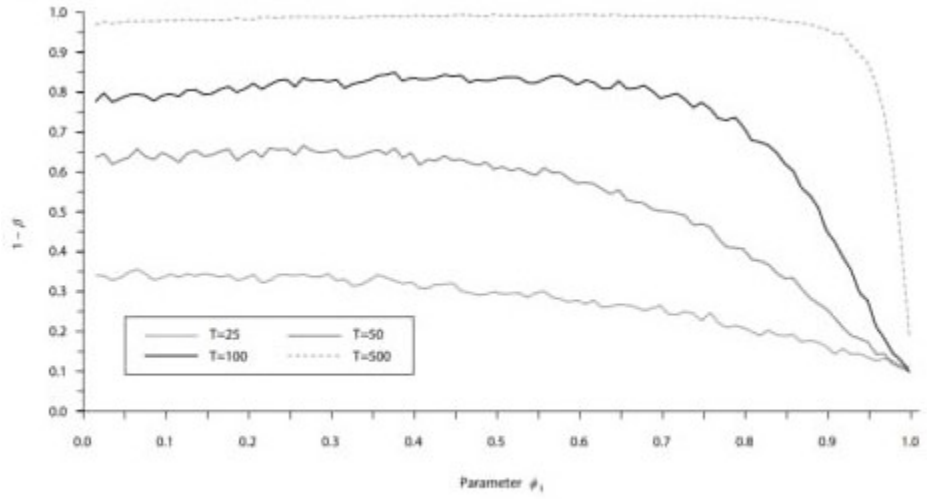
Source: Authors' calculation

Figure 4.6: Power functions$(1-\beta)$ of PP test for simulated time series of lengths T=25,50,100,500 and number of replications n=3000, $\alpha = 0.05$.

T=50 and T=100 respectively. The power of the test is significantly higher if the time series is of length 500. But, similarly to the result of the ADF test, the power decrease rapidly for ϕ_1 close to unity.

The ADF-GLS test:

As the figure below shows, the power of the ADF-GLS remarkably increase with the size of the time series.In addition, we can directly notice that the power of this test is very low in comparison with the results of previous tests for small T and ϕ_1. The power is less than 40% for T=25, similar to the previous test, the power of ADF-GLS increases with the growing length of the time series and decreases every time ϕ_1 moves toward 1. It approaches to 70% for T=50 and it reaches 80% for T=100 and $0.01 < \phi_1 < 0.7$, then it begins to fall again. For the time series of length T = 500, we can observe greater ability to reject the null hypothesis of the unit root also for higher values of the parameter . The power of the test is approaching 1 even for the values

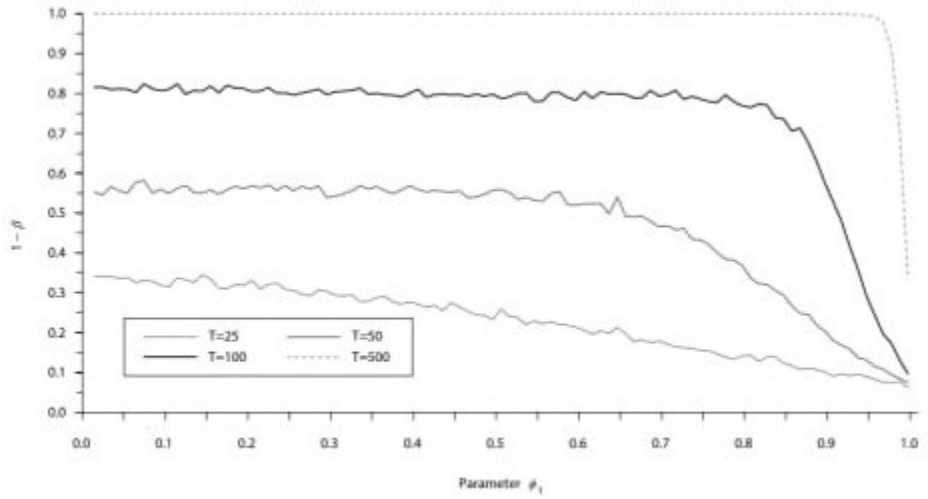


Source: Authors' calculation

Figure 4.7: Power functions $(1-\beta)$ of ADF-GLS test for simulated time series of lengths T=25,50,100,500 and number of replications n=3000, $\alpha = 0.05$.

around $\phi_1 = 0.9$. It, however, declines sharply after this value.

The NGP test:



Source: Authors' calculation

Figure 4.8: Power functions $(1-\beta)$of NGP test for simulated time series of lengths T=25,50,100,500 and number of replications n=3000, $\alpha = 0.05$.

Similarly, the power of the NGP test is very low in the case of small time series, it only reaches around 30% maximum for T=25 and around 50% for T=50, then the power start to decrease slowly when ϕ_1 goes toward 1. The NGP test is assumed to have high power in the case of ϕ_1 close to 1. For large sample sizes, the power of the test increase to 80% for T=100 and then start to decrease after $\phi_1 < 0.8$. Now, for T=500, the power reaches 1 for $\phi_1 < 0.95$ and falls sharply just before ϕ_1 close to 1.

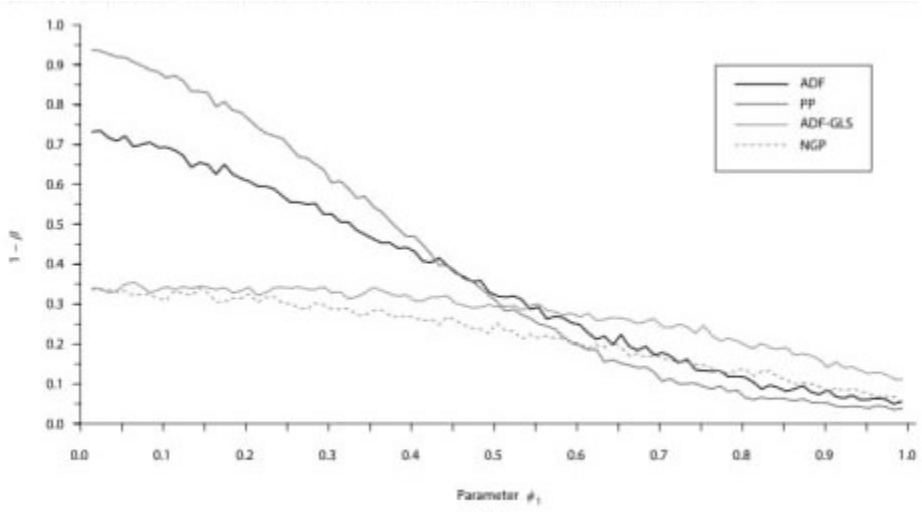
After analyzing the results of each test according to the length of the time series and to the parameter values, the authors intend to compare the power of these tests for each length of the time series.

Very short time series (T = 25):

First, Several tests for time series of length T = 25 will be compared.
In general, it's clear that the power of these unit root tests is very low for very short time series (T=25). For $\phi_1 < 0.5$, the ADF and PP tests have more power than the ADF-GLS and NGP tests. Now, for $\phi_1 \in [0.5; 1]$, the ADF-GLS and the PP tests seems to be more powerful.

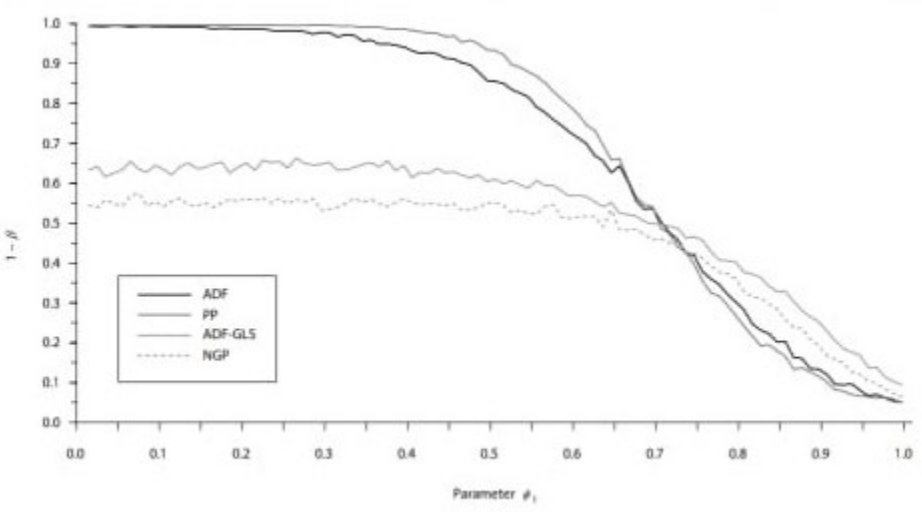
Medium-long time series (T = 50):

Similar to the case T=25, the ADF and PP tests seems to be more powerful than the ADF-GLS and the PP tests for $\phi_1 < 0.7$, the opposite occurs for $\phi_1 > 0.7$.



Source: Authors' calculation

Figure 4.9: Comparison of power functions of selected tests for time series length T = 25, $\alpha = 0.05$.

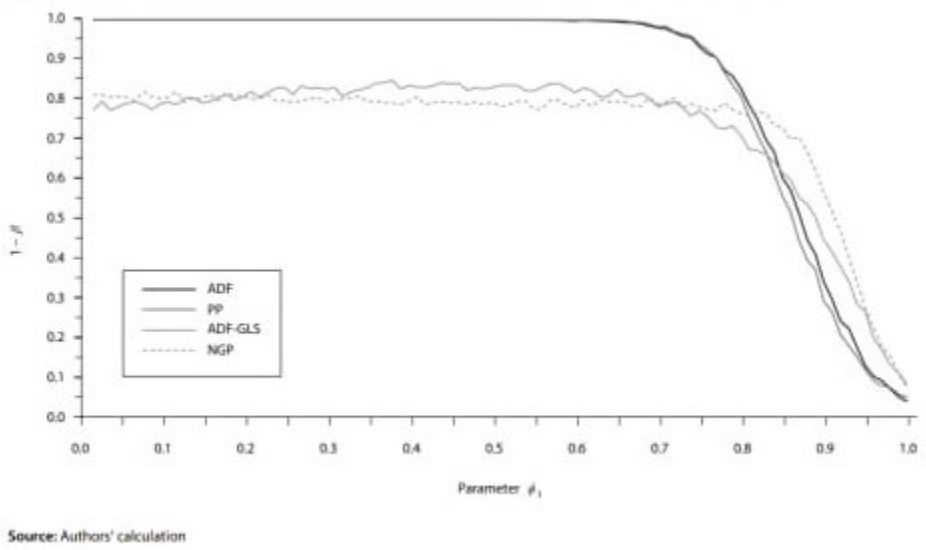


Source: Authors' calculation

Figure 4.10: Comparison of power functions of selected tests for time series length T = 50,$\alpha = 0.05$.

Long time series (T = 100):

The power of these tests increase as the length of the time series increase. For



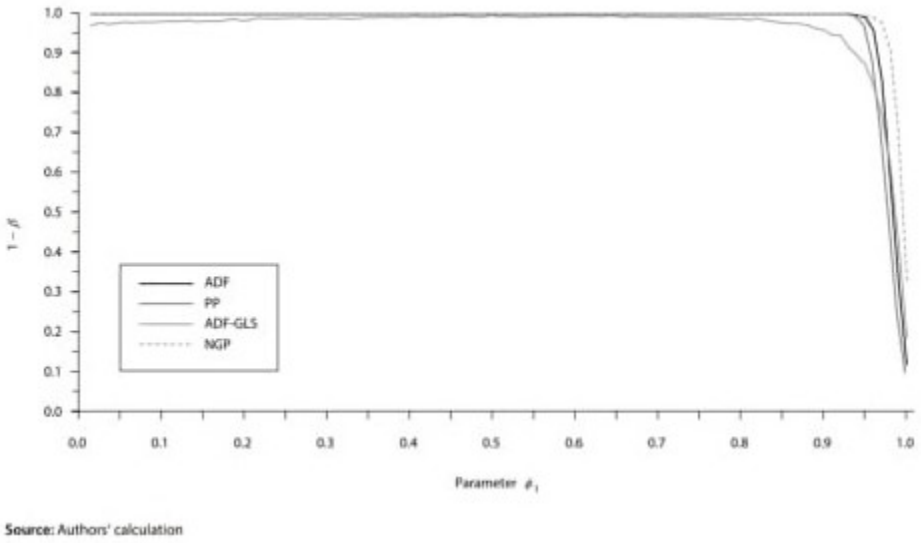
Source: Authors' calculation

Figure 4.11: Comparison of power functions of selected tests for time series length T = 100,$\alpha = 0.05$.

the ADF and PP tests, as we have seen before, for small value of ϕ_1, the power of these two tests reaches the best results. Here, in the case of T=100, the power is 1 for $\phi_1 < 0.6$. However, the other two tests have higher power when compared to ADF and PP tests for $\phi_1 > 0.8$. The advantages of modifications of classic unit root tests, i.e. NGP and ADF-GLS, can already be observed. (see figure 4.11)

Very long time series (T = 500):

The results of all tests for time series of length T=500 are very good. The power functions of the ADF and the PP tests are almost the same. their power reaches 100% for $\phi_1 < 0.9$ and then it drops suddenly. The ADF-GLS test also increases in power and approaches to 100%, then the power suddenly and



Source: Authors' calculation

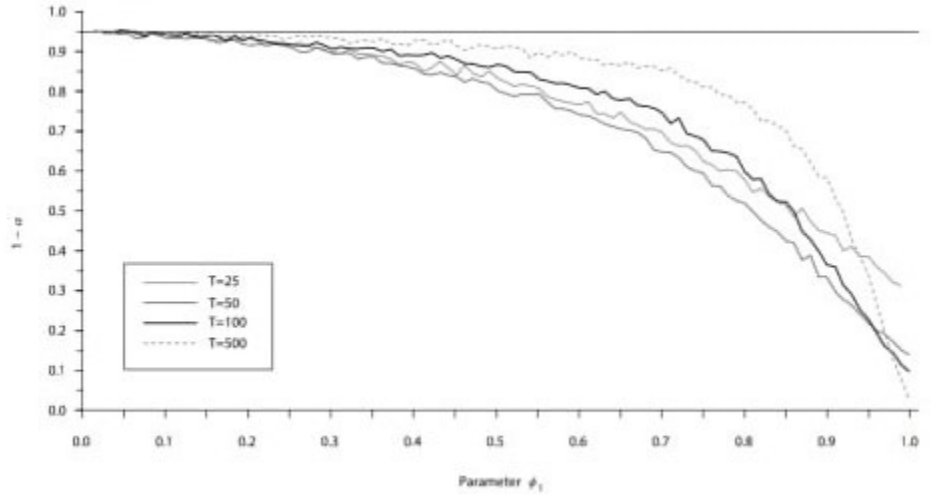
Figure 4.12: Comparison of power functions of selected tests for time series length T = 500, $\alpha = 0.05$.

sharply declines for $\phi_1 > 0.8$. The NGP shows great results in the case of very long time series. It reaches 100% for $\phi_1 < 0.95$ and its power for ϕ_1 close to 1 is still the highest in comparison with other tests.

The KPSS test:

Unlike the tests mentioned above, the KPSS test is a stationarity test. Therefore, we modify the power functions $(1-\beta)$ by $(1-\alpha)$ which is the probability of not rejecting the null hypothesis when it is true.

For small values of ϕ_1 ($\phi_1 < 0.2$), the results of the KPSS test is almost the same for all selected time series sizes, the probability is around 0.95, then it sharply declines. Notice that the probability for T=25 is the highest compared to other sizes for ϕ_1 very close to 1 ($\phi_1 > 0.95$).



Source: Authors' calculation

Figure 4.13: Probability $(1-\alpha)$ of KPSS test for simulated time series of lengths T = 25,50,100,500 and number of replications n=3 000, $\alpha = 0.05$.

As a result, the authors concluded that there is no specified test that could be generally applied in all cases. Thus, they gave some recommendations of which of these tests are the most appropriate to use for a specific T and a given ϕ_1. The results are summarized in the table below by taking into account three basic aspects: power of the test, it's validity and ease of use.

ϕ_1	**T=25**	**T=50**	**T=100**	**T=500**
(0;0.5)	PP,ADF,+KPSS	PP,ADF,+KPSS	ADF,PP	ADF,PP,NGP
(0.5;0.7)	PP,ADF, +KPSS	PP,ADF, +KPSS	ADF,PP	ADF,PP,NGP
(0.7;0.9)	ADF,NGP,PP	ADF,NGP,PP	ADF,NGP	ADF,NGP
(0.9;1)	PP,NGP,ADF	ADF,NGP	ADF,NGP	ADF,NGP

Table 4.5: Overview of appropriate tests for different length of time series and values of parameter ϕ_1, $\alpha = 0.05$.

The authors recommend to use the KPSS (which is suitable for very small values of ϕ_1) as complementary test during the unit root testing of shorter time series.

4.9 Unit Root Tests in Finance:

The importance of unit root tests in macroeconomic time series is to determine the stationarity of the series: in the absence of a unit root, the series is characterized as stationary, and therefore exhibits mean reversion in that it fluctuates around a constant long run mean. Also, in this case, the series has a finite variance that is time-independent and the effects of shocks dissipate over time. Alternatively, the existence of unit root indicates that the series is non-stationary, the process have no tendency to return to a long-run deterministic path. The variance does depend on time and goes to infinity as time approaches to infinity. This will result serious problems in forecasting. Non stationary series suffer permanent effects from random shocks.

Nelson and Plosser (1982) gave statistical evidence that supports the hypothesis of a unit root in the autoregressive representations of macroeconomic time series for the US, including GNP, employment, wages, prices, interest rates, and stock prices. In the literature of unit root testing, numerous economic series have been considered: the real exchange rates (in Narayan and Narayan, 2007(Italy) [5] ; Matsuki and Sugimoto, 2013 (Asia) [6] ; El Montasser, Fry and Apergis, 2016 (US-China) [7]), inflation rate (in Basher and Westerlund, 2008 [8] ; Huang, Lin and Yeh, 2010 (US) [9]), income (in Smyth and Inder, 2004 (China)[10]) and Stock indices (in Tabak, 2007 (Brazil) [11] ; Narayan, 2008 (G7 countries)[12]), etc.

In brief, the existence of a unit root in macroeconomic time series brings about important implications, and this helps to understand why this topic has received a great amount of theoretical and applied research.

Chapter 5

Conclusion

Since the mid 1980' s there has been a veritable explosion of research on the importance of unit roots in the analysis of economic and other time series data. The reasons for this are diverse, but perhaps the most important motivation for this work is the fact that the development of the notion of co-integration by Granger (1981) and Engle and Granger (1987) has stressed the significance of unit roots and the importance of making valid statistical inference in the presence of non stationary time series data. In this thesis, the integrated and difference processes were presented, then several unit root tests were introduced: the Dickey-Fuller and its augmented version the ADF, the Phillips Peron unit root tests, the ADF-GLS, the NGP tests and finally the KPSS tests. All the tests discussed so far have the unit root as the null hypothesis, and (trend) stationarity as the alternative, except the KPSS test which is a stationary test (i.e. the null hypothesis is I(0) against the alternative which is I(1)). In the discussion of the ADF, we have seen that determining the number of lags is important unlike the PP test which corrects this issue. In

the case of the ADF-GLS and NGP tests, it was proposed to detrend the time series before applying the tests. However, frequent criticisms concerning the power of unit root tests have been exhibited. It was shown that these tests have low power for small sample size. In addition, these tests suffer when the process is stationary but with a root close to the non-stationary boundary.

Bibliography

[1] David Bellhouse. Decoding cardano's liber de ludo aleae. *Historia Mathematica*, 32(2):180–202, 2005.

[2] Mohammad Kazemi and Ali Akbar Jafari. Comparing seventeen interval estimates for a bivariate normal correlation coefficient. 10 2014.

[3] Denis Kwiatkowski, Peter CB Phillips, Peter Schmidt, Yongcheol Shin, et al. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of econometrics*, 54(1-3):159–178, 1992.

[4] Darina Fedorová. Selection of unit root test on the basis of length of the time series and value of ar (1) parameter. *Statistika*, 96(3):3, 2016.

[5] Paresh Kumar Narayan and Seema Narayan. Are real exchange rates nonlinear with a unit root? evidence on ppp for italy: a note. *Applied Economics*, 39(19):2483–2488, 2007.

[6] Takashi Matsuki and Kimiko Sugimoto. Stationarity of asian real exchange rates: An empirical application of multiple testing to nonstationary panels with a structural break. *Economic Modelling*, 34:52–58, 2013.

[7] Ghassen El Montasser, John Fry, and Nicholas Apergis. Explosive bubbles in the us–china exchange rate? evidence from right-tailed unit root tests. *China Economic Journal*, 9(1):34–46, 2016.

[8] Syed A Basher and Joakim Westerlund. Is there really a unit root in the inflation rate? more evidence from panel data models. *Applied Economics Letters*, 15(3):161–164, 2008.

[9] Ho-Chuan Huang, Pei-Chien Lin, and Chih-Chuan Yeh. Price level convergence across cities? evidence from panel unit root tests. *Applied Economics Letters*, 18(1):87–93, 2010.

[10] Russell Smyth and Brett Inder. Is chinese provincial real gdp per capita nonstationary?: Evidence from multiple trend break unit root tests. *China Economic Review*, 15(1):1–24, 2004.

[11] Benjamin M Tabak. Testing for unit root bilinearity in the brazilian stock market. *Physica A: Statistical Mechanics and its Applications*, 385 (1):261–269, 2007.

[12] Paresh Kumar Narayan. Do shocks to g7 stock prices have a permanent effect?: Evidence from panel unit root tests with structural change. *Mathematics and Computers in Simulation*, 77(4):369–373, 2008.

[13] Alan Winterbottom. A note on the derivation of fisher's transformation of the correlation coefficient. *The American Statistician*, 33(3):142–143, 1979.

[14] Anindya Banerjee, Juan J Dolado, John W Galbraith, David Hendry, et al. Co-integration, error correction, and the econometric analysis of non-stationary data. *OUP Catalogue*, 1993.

[15] Donald McCarthy. An introduction to testing for unit roots using sas®: the case of us national health expenditures. *Department of Research and Evaluation, Kaiser Permanente*, 2015.

[16] Edward Herranz. Unit root tests. *Wiley Interdisciplinary Reviews: Computational Statistics*, 9(3), Apr 2017.

[17] E Zivot. Economic theory ii: Time series analysis—unit root tests. *Lecture Notes-Econ*, 584:111–139, 2005.

[18] Dan Vlad Metes. Visual, unit root and stationarity tests and their power and accuracy. *Department of Mathematical Sciences, University of Alberta, Edmonton, Canada*, 2005.

[19] Yin-Wong Cheung and Kon S. Lai. Lag order and critical values of the augmented dickey–fuller test. *Journal of Business & Economic Statistics*, 13(3):277–280, 1995.

[20] G. William Schwert. Tests for unit roots. *Journal of Business & Economic Statistics*, 20(1):5–17, 2002.

[21] P. C. B. Phillips. Time series regression with a unit root. *Econometrica*, 55(2):277, 1987.

[22] Said E. Said and David A. Dickey. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3):599–607, 1984.

[23] Peter Phillips and Pierre Perron. Testing for a unit root in time series regression. *Cowles Foundation, Yale University, Cowles Foundation Discussion Papers*, 75, 01 1986.

[24] Robert Dixon. Random walks and the dickey-fuller test statistic. 05 2020.

[25] John N Haddad and Serge B Provost. Approximations to the distribution of the sample correlation coefficient. *World Academy of Science, Engineering and Technology*, 52:910–915, 2011.

[26] Heino Bohn Nielsen. Non-stationary time series and unit root testing. *Engineering statistics hand book. Econometrics*, 2007.

[27] Yongcheol Shin and Peter Schmidt. The kpss stationarity test as a unit root test. *Economics Letters*, 38(4):387–392, 1992.

[28] Mehmet Caner and Lutz Kilian. Size distortions of tests of the null hypothesis of stationarity: evidence and implications for the ppp debate. *Journal of International Money and Finance*, 20(5):639–657, 2001.

[29] Ronald A Fisher. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10(4):507–521, 1915.

[30] Harold Hotelling. New light on the correlation coefficient and its transforms. *Journal of the Royal Statistical Society. Series B (Methodological)*, 15(2):193–232, 1953.

[31] Ramu Ramanathan. Introductory econometrics with applications. 2002.

[32] David A Dickey and Wayne A Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431, 1979.

[33] Serena Ne and Pierre PERRoN. Unit root tests in arma models with doto-dependent methods for the selection of the truncation log. *Journal of the American Statistical Association*, 90(429):268–28, 1995.

[34] Gangadharrao S Maddala and In-Moo Kim. *Unit roots, cointegration, and structural change*. Number 4. Cambridge university press, 1998.

[35] Peter CB Phillips. Time series regression with a unit root. *Econometrica: Journal of the Econometric Society*, pages 277–301, 1987.

[36] M Hashem Pesaran. *Time series and panel data econometrics*. Oxford University Press, 2015.

[37] G William Schwert. Tests for unit roots: A monte carlo investigation. *Journal of Business & Economic Statistics*, 20(1):5–17, 2002.

[38] Peter J Brockwell and Richard A Davis. Nonstationary and seasonal time series models. In *Introduction to Time Series and Forecasting*, pages 157–193. Springer, 2016.

[39] Ronald A Fisher. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10(4):507–521, 1915.

[40] Ronald Aylmer Fisher et al. 014: On the" probable error" of a coefficient of correlation deduced from a small sample. 1921.

[41] Harold Ruben. Some new results on the distribution of the sample correlation coefficient. *Journal of the Royal Statistical Society: Series B (Methodological)*, 28(3):513–525, 1966.

[42] MV Muddapur. A simple test for correlation coefficient in a bivariate normal distribution. *Sankhyā: The Indian Journal of Statistics, Series B*, pages 60–68, 1988.

Appendix A

Formulation of the test regression of the ADF

Given an AR(3) model:

$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \varepsilon_t,$
Then,
$y_t - y_{t-1} = (\phi_1 - 1)y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \varepsilon_t,$
$y_t - y_{t-1} = (\phi_1 - 1)y_{t-1} + (\phi_2 + \phi_3)y_{t-2} + \phi_3(y_{t-3} - y_{y-2}) + \varepsilon_t,$
$y_t - y_{t-1} = (\phi_1 + \phi_2 + \phi_3 - 1)y_{t-1} + (\phi_2 + \phi_3)(y_{t-2} - y_{t-1}) + \phi_3(y_{t-3} - y_{t-2}) + \varepsilon_t,$
Thus, y_t can be written as a function of just y_{t-1} and a series of difference lags terms:

$$\begin{aligned} y_t - y_{t-1} &= (\delta - 1)y_{t-1} + \rho_1 \Delta y_{t-1} + \rho_2 \Delta y_{t-2} + \varepsilon_t, \\ y_t &= \delta y_{t-1} + \rho_1 \Delta y_{t-1} + \rho_2 \Delta y_{t-2} + \varepsilon_t, \\ \Delta y_t &= \pi y_{t-1} + \sum_i \rho_i \Delta y_{t-i} + \varepsilon_t, i = 1, 2. \end{aligned} \tag{A.0.1}$$

Where,

$$\pi = \delta - 1,$$

$$\delta = \phi_1 + \phi_2 + \phi_3,$$

$$\rho_1 = \phi_2 + \phi_3,$$

$$\rho_2 = \phi_3,$$

Note that testing for a unit root in AR(3) model is testing for the existence of B=1 in the AR polynomial equation: $1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 = 0$.

Equivalently, testing for $\phi_1 + \phi_2 + \phi_3 = 1$ i.e. testing for $\delta = 1(\pi = 0$ in (A.0.1)).

In the same way, an ARMA(p,q) process (or AR(p) process) could be written as: $\Delta Y_t = \alpha + \beta t + \pi Y_{t-1} + \sum_{i=1}^{p} \delta_i Y_{t-i} + \varepsilon_t$ Where α is a constant and β the coefficient on a time trend. Here, testing for stationarity is testing for $\pi = 0$. And this is called an Augmented Dickey-Fuller test (ADF).