Ministry of Higher Education and Scientific Research
University Center of El Bayadh
Institute of Economic Sciences, Commercial Sciences,
and Management Sciences

A Pedagogical Handbook on the Subject of:

Time Series Analysis 2

Intended for Third-Year Undergraduate Students – Specialization in Quantitative economics and Statistics

Prepared by:
Dr :Bourdache Chahrazed
Associate Professor-A-

Academic Year: 2025 – 2026

Table of contents

Table of contents

Introduction

Introduction

Time series analysis plays a fundamental role in quantitative economics and statistics, providing a structured framework for examining how economic and financial variables evolve over time. Understanding these dynamics is crucial for making informed decisions, forecasting future trends, and evaluating the impacts of policy changes or external shocks on economic systems.

This pedagogical work is designed to equip third-year undergraduate students with the theoretical knowledge and practical tools necessary to analyze and model time-dependent data. It emphasizes the development of critical skills for interpreting the underlying patterns of economic variables, identifying stochastic behaviors, and applying forecasting techniques to real-world datasets.

The content of this textbook is organized around four main themes. The first focuses on stochastic time series models, which provide the foundation for modeling random fluctuations in economic data. The second introduces long-memory models, highlighting processes where past observations exert a prolonged influence on future values. The third section addresses non-stationary stochastic processes and unit root testing, enabling students to distinguish between stable and unstable series and to apply appropriate statistical tests. Finally, the fourth part presents the Box-Jenkins methodology for short-term forecasting, offering a systematic approach to model identification, estimation, and prediction.

Through learning this topic, students will get an effective understanding of the theoretical components of time series analysis as well as practical experience applying these approaches to economic and financial data. This knowledge serves as a base to more complex modeling approaches and empirical research, offering an ideal basis for careers in data analysis, economic forecasting, and applied econometrics. This knowledge serves as a stepping stone toward more advanced modeling techniques and empirical research, providing a strong foundation for careers in data analysis, economic forecasting, and applied econometrics.

Chapter One:
Stochastic Time Series Models

From a theoretical standpoint, a time series is not merely a sequence of observed numerical values arranged over time, but rather a realization of an underlying random mechanism. Formally, a time series X is defined as a collection of random variables indexed by time, which constitutes a **stochastic process**. When the evolution of the variable occurs in continuous time, the process is denoted by X(t); however, in most economic and social applications, observations are recorded at discrete time intervals. In such cases, the process is represented by X_t, where[1] $t = 1,2, \dots, T$.

The stochastic nature of time series data implies that the value of the variable at any given time is not deterministic. Instead, it is influenced by a wide range of economic, political, institutional, and behavioral factors. Consequently, at each time period t, the random variable X_t may assume different numerical values, even under similar conditions. The observed value is therefore interpreted as a particular realization of the stochastic process, selected from a set of many possible realizations that the process is capable of generating[2].

This interpretation highlights a fundamental distinction between the stochastic process itself and its realizations. While the stochastic process represents the theoretical data-generating mechanism, the observed time series corresponds to only one realization of that mechanism over a finite time horizon. In this sense, the observed series does not exhaust all the information contained in the process, but rather provides a limited window through which its properties can be studied.

The relationship between a stochastic process and its realizations in time series analysis is analogous to the relationship between a population and a sample in cross-sectional data analysis. Just as a sample is used to infer the characteristics of a population, the observed realization of a time series is used to draw inferences about the

[1] Maddala. G.S, **introduction to econometrics**, second edition, MacMillan publishing company, New York, 1992, P 527.
[2] Gujarati .D.N, **Basic Econometrics** , 4th edition, Mc Graw-Hill / Irwin companies Inc New York, 2003, P 797.

underlying stochastic process and the structure of the system that generates the data. Statistical modeling of time series is therefore concerned with specifying appropriate stochastic models capable of capturing the essential features of the process, such as its mean behavior, variability, dependence over time, and possible structural components[1].

The relationship between a stochastic process and its realizations in time series analysis is analogous to the relationship between a population and a sample in cross-sectional data analysis. Just as a sample is used to infer the characteristics of a population, the observed realization of a time series is used to draw inferences about the underlying stochastic process and the structure of the system that generates the data. Statistical modeling of time series is therefore concerned with specifying appropriate stochastic models capable of capturing the essential features of the process, such as its mean behavior, variability, dependence over time, and possible structural components.

Accordingly, stochastic time series models are designed to describe and analyze the probabilistic structure governing the evolution of economic and social variables over time. These models provide a formal framework for understanding temporal dependence, assessing uncertainty, and producing forecasts based on the estimated characteristics of the underlying stochastic process. The subsequent sections will focus on the key properties of these models, including stationarity, autocorrelation, and the role of random disturbances in shaping time series behavior.

I - Definition of a Stochastic Time Series:

A **stochastic time series** is defined as an ordered sequence of observed values generated by a stochastic process over time. Formally, it represents a single realized path of a collection of random variables

[1] Gujarati .D.N, op.cit, p 797.

$T_{t=1}\{X_t\}$, where each X_t is a random variable associated with a specific point in time.

Unlike deterministic sequences, the evolution of a stochastic time series is governed by probabilistic laws rather than fixed functional relationships. This implies that, even under identical initial conditions, the series may evolve differently due to random disturbances and unobservable factors. Consequently, the observed values of a stochastic time series are subject to uncertainty and variability, reflecting the random nature of the underlying data-generating process.

From an econometric perspective, a stochastic time series is not analyzed as an isolated sequence of numbers, but as an empirical manifestation of a broader probabilistic mechanism. The primary objective of time series analysis is therefore to identify, model, and interpret the statistical properties of this mechanism—such as its mean structure, variance, temporal dependence, and potential structural components—based on the observed realization[1].

This definition provides the conceptual foundation for developing stochastic models in time series analysis, including autoregressive, moving-average, and mixed processes. It also underpins key theoretical concepts such as stationarity, autocorrelation, and forecasting, which rely on the assumption that the observed series faithfully reflects the essential characteristics of the underlying stochastic process[2].

II- Properties of Stochastic Time Series:

Stochastic time series possess several important properties that help understand and analyze their behavior[3]:

1. Temporal Dependence: Current values may depend on past values of the series.

2. Mean and Variance: The series may have a constant or varying mean, and its variance may be constant or change over time.

[1] Maddala. G.S, op.cit, p 527.
[2] Gujarati .D.N, op.cit, p 797.
[3] Christiaan. H & others, **Introduction to Time Series Econometrics**, Oxford University Press, 2004, p p 27-32.

3. **Stationarity:** If the statistical properties of the series, such as mean, variance, and autocovariance, do not change over time, the series is said to be stationary.

4.Randomness: Despite temporal dependence, some components of the series are inherently random and unpredictable.

5. Components: Time series generally consist of four main components: Trend, Seasonality, Cyclical variations, and Irregular (Random) fluctuations.

III- Stationary Stochastic Time Series:

Among the most significant stochastic time series that have received considerable attention from specialists are the so-called *stationary stochastic time series*. In this context, one may ask: why has so much emphasis been placed on these series? The answer lies in the fact that if a time series is non-stationary, analyzing its behavior beyond the specific period under study—or generalizing the findings from that period to other time intervals—becomes practically impossible, since each set of observations pertains to a specific temporal frame. This underscores the critical importance of *stationarity* in stochastic time series, particularly when they are used for prediction.

III- 1- Strict Stationarity (Strong Stationarity):

Consider a stochastic time series X_t consisting of "**t**" observations. A comprehensive and rigorous description of such a series can, in principle, be obtained by specifying the joint probability distribution function of the random variables X_t[1] . Based on this perspective, a time series X_t is said to be strictly stationary if its probability distribution is invariant with respect to time "**t**".

More formally, X_t is strictly stationary if, for any set of time indices.

[1] Kirchgässner. G, wolters. J, **introduction to modern time series analysis**, Springer-verlag Berlin Heidelberg, New York 2007, P 12.

$$\mathrm{T} \geq \mathrm{m} \quad , \quad \mathrm{T}^{\mathrm{m}} \in (\mathrm{t}_1, \mathrm{t}_2, \dots, \mathrm{t}_{\mathrm{m}})$$

and for any time shift $\mathrm{T} \in \tau$ such that $\mathrm{T} \in t_{\mathrm{i}+\tau}$ for all $(i = 1, 2, \dots, m)$, the joint distribution of:

$$\{\mathrm{X}_{\mathrm{t}_1}, \mathrm{X}_{\mathrm{t}_2}, \dots, \mathrm{X}_{\mathrm{t}_{\mathrm{m}}}\}$$

is identical to that of:

$$\{\mathrm{X}_{\mathrm{t}_1+\tau}, \mathrm{X}_{\mathrm{t}_2+\tau}, \dots, \mathrm{X}_{\mathrm{t}_{\mathrm{m}}+\tau}\}$$

That is :

$$f_{\mathrm{X}_{\mathrm{t}_1},\dots,\mathrm{X}_{\mathrm{t}_{\mathrm{m}}}}\left(\mathrm{X}_{\mathrm{t}_1}, \dots, \mathrm{X}_{\mathrm{t}_{\mathrm{m}}}\right) = f_{\mathrm{X}_{\mathrm{t}_1+\tau},\dots,\mathrm{X}_{\mathrm{t}_{\mathrm{m}}+\tau}}\left(\mathrm{X}_{\mathrm{t}_1+\tau}, \dots, \mathrm{X}_{\mathrm{t}_{\mathrm{m}}+\tau}\right)$$

Consequently, a strictly stationary stochastic time series is one whose mean, variance, and all higher-order moments remain constant over time. In other words, the entire probabilistic structure of the series is invariant under time shifts.

From a practical standpoint, however, determining the joint probability distribution of a collection of random variables is an extremely complex task. Moreover, assuming that this distribution is fully independent of time constitutes a very strong and often unrealistic assumption, particularly in applied economic and financial contexts. For this reason, empirical time series analysis typically relies on a weaker and more tractable concept of stationarity, known as weak (or second-order) stationarity, which focuses only on the first two moments of the series[1].

Example:

Let X_{t} be a strictly stationary time series. Prove that:

1. For any integer k :

$$F(X_t) = F(X_{t+k})$$

[1] Maddala. G.S, op.cit, p 527-528.

2. the following joint distributions are identical:

$$F(X_2,\ X_5, X_{10}) = F(X_3,\ X_6, X_{11}) = F(X_1,\ X_4, X_9)$$

3. the following joint distributions are identical:

$$F(y_t, y_{t-1}) = F(y_{t+h}, y_{t-1+h})$$

Solution:

1. Setting $m = 1$ and $t_1 = t$, we obtain:

$$P(X_t \leq C_1) = P(X_{t+k} \leq C_1) \quad \forall\ C_1$$

Hence, their cumulative distribution functions are equal:

$$F(X_t) = F(X_{t+k}) \qquad\qquad \forall\ k$$

2. Setting $m = 3$, $t_1 = 2$, $t_2 = 5$, $t_3 = 10$, $k = \mp 1$, we obtain:

$$P(X_2 \leq C_1, X_5 \leq C_2, X_{10} \leq C_3) = P(X_3 \leq C_1, X_6 \leq C_2, X_{11} \leq C_3)$$
$$= F(X_1 \leq C_1, X_4 \leq C_2, X_9 \leq C_3) \quad \forall (C_1, C_2, C_3)$$

Hence, their cumulative distribution functions are equal:

$$F(X_2, X_5, X_{10}) = F(X_3, X_6, X_{11}) = F(X_1, X_4, X_9)$$

3. Setting $m = 2$, $t_1 = t$, $t_2 = t - 1$, we obtain:

$$P(y_t \leq C_1, y_{t-1} \leq C_2) = P(y_{t+h} \leq C_1, y_{t-1+h} \leq C_2) \quad \forall (C_1, C_2)$$

Hence, their cumulative distribution functions are equal:

$$F(y_t, y_{t-1}) = F(y_{t+h}, y_{t-1+h}) \quad \forall\ h$$

III- 2- Weak Stationarity:

A time series X_t is said to be weakly stationary **(or** stationary in the wide sense**)** if its first two moments are invariant over time and if the dependence structure between observations depends only on the time lag separating them, and not on the specific time at which the observations are taken.

Formally, the process X_t is weakly stationary if the following conditions are satisfied[1]:

$$\begin{cases} 1.\ E(X_t) = \mu = cte \quad E(X_t^2) = \mu' = cte \\ 2.\ Var(X_t) = \mu' - \mu^2 = \gamma_0 = \sigma_X^2 \\ 3.\ Cov(X_t X_K) = E(X_t X_K) - \mu^2 = \gamma_h\ ; \quad K = t \pm h \end{cases}$$

The first condition implies that the series fluctuates around a constant mean level. The second condition ensures that the variance of the process is finite and time-invariant, which excludes explosive or highly unstable behaviors. The third condition states that the autocovariance between two observations depends solely on the lag hhh separating them, reflecting a stable dependence structure over time [2].

As a consequence, weakly stationary time series exhibit a tendency to revert to their long-run mean. Moreover, deviations from this mean—measured by the variance—remain constant over time, implying that the amplitude of fluctuations does not increase or decrease systematically as time evolves[3].

Remark:

As a special case, if the random variables $X_{t_1}, X_{t_2}, \dots, X_{t_m}$ follow a multivariate normal distribution, then the entire joint distribution is fully characterized by the first and second moments. Under this assumption, strict stationarity and weak stationarity become equivalent concepts, since invariance of the mean vector and covariance matrix implies invariance of all finite-dimensional distributions[4].

---

[1] Bresson.G, pirotte. A, **économétrie des séries temporelles, théories et applications**, Presses Universitaires de France, paris, 1995, P 19.
[2] Hamilton. J. D, **Time Series Analysis**, Princeton University Press, 1994, p 17.
[3] Maddala. G.S, op.cit, p 213.
[4] Kirchgässner. G, wolters. J, op.cit, p 13.

However, this equivalence does not generally hold for non-Gaussian processes, where higher-order moments may vary over time even if the first two moments remain constant.

Example 1:

Suppose the time series y_t follows the model:

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t \qquad , t = 1,2, \dots, n$$

where β_0 and β_1 are constants, and ε_t is white noise.

Solution:

$$E(y_t) = \beta_0 + \beta_1 t \quad , \quad t = 1,2, \dots \dots, n$$

This means that the mean (expected value) of the series y_t is not constant over time, so:

- ✓ It increases at a constant rate if $\beta_1 > 0$.
- ✓ It decreases at a constant rate if $\beta_1 < 0$.

In other words, the series exhibits a deterministic trend if $\beta_1 \neq 0$. Therefore, the series y_t is **non-stationary**.

Example 2:

Suppose the random process $\{y_t\}$ follows the model:

$$y_t = y_{t-1} + \varepsilon_t \qquad , t = 1,2, \dots \dots, n$$

Is the process y_t stationary?

Solution:

Step 1: Mean

$$E(y_t) = E(y_{t-1}) + E(\varepsilon_t) = E(y_{t-1})$$

Therefore, the expected value does **not depend on time**, so the mean is **constant**.

Step 2: Variance

$$V(y_t) = V(y_{t-1}) + \sigma^2 + 2cov\,(y_{t-1}, \varepsilon_t)$$

$$= V(y_{t-1}) + \sigma^2$$

This shows that the variance **increases over time**:

$$V(y_t) \neq V(y_{t-1})$$

Although the mean is constant, the variance is **not constant**, and therefore the series y_t **is not stationary**.

the process y_t is non-stationary.

Example 3:
Suppose the time series y_t follows the model:

$$y_t = \beta_0 + \varepsilon_t$$

where β_0 is a constant, and $\varepsilon_1, \varepsilon_2, \ldots$ are uncorrelated random variables with zero mean and variance σ^2.

Is the series y_t stationary.

Solution:

$$E(y_t) = \beta_0 \quad , \quad t = 0, \pm 1, \pm 2, \ldots.$$

Thus, the mean of the series does **not depend on time**.

$$V(y_t) = V(\beta_0 + \varepsilon_t) = V(\varepsilon_t) = \sigma^2$$

Hence, the variance is also **independent of time**.

$$cov(y_t, y_{t-k}) = cov(\beta_0 + \varepsilon_t, \beta_0 + \varepsilon_{t-k}) = 0 \quad , \quad k = \pm 1, \pm 2, \ldots.$$

Therefore, all autocovariances are **independent of time**.

Since the mean, variance, and autocovariances of y_t are constant over time, the series y_t is **weakly stationary**.

III- 3- Autocorrelation Function (ACF):

If we want to study the correlation between pairs (X_t, X_{t+h}), we rely on what is called the autocorrelation function. The autocorrelation coefficient of $lag\ h$ for a stationary time series X_t is defined as[1]:

$$\rho_h = \frac{\text{cov}\,(X_t, X_{t+h})}{[\text{Var}(X_t)]^{1/2}.\ [\text{Var}(X_{t+h})]^{1/2}} = \frac{\gamma_h}{\gamma_0}\ ;\ h = 0, \qquad \pm 1, \pm 2, \dots, \pm T$$

In practice, we usually have a single observed sample of the stochastic process, so we compute the sample autocorrelation function (SACF):

$$\hat{\rho}_h = \frac{\hat{\gamma}_h}{\hat{\gamma}_0}$$

Where:

$$\hat{\gamma}_0 = \frac{1}{T-1}\sum_{t=1}^{T}(X_t - \overline{X})^2$$

$$\hat{\gamma}_h = \frac{1}{T-1}\sum_{t=1}^{T-h}\ (X_t - \overline{X})(X_{t+h} - \overline{X})$$

$$\overline{X} = \frac{1}{T-1}\sum_{t=1}^{T} X_t$$

Properties of the autocorrelation function:

1. $\rho_0 = 1$;

---

[1] Bresson. G, pirotte. A, op.cit, 1995, P 21.

2. $|\rho_h| \leq \rho_0$;

3. $\rho_h = \rho_{-h}$ (Even function);

4. For a given ACF, there may be several non-natural series with the same ACF[1].

III- 3- 1- Statistical Significance of Individual Autocorrelations:

After estimating the SACF of a time series X_t, the main question is: which of these coefficients significantly differ from zero? This can be answered by constructing confidence intervals for the estimates at a chosen significance level.

Bartlett (1946) showed that if X_t is a pure random (white noise) series, the sample autocorrelations $\hat{\rho}_h$ are approximately normally distributed with zero mean and variance $\frac{1}{T}$[2]:

$$\hat{\rho}_h \overset{a}{\rightarrow} N\left(0 \,.\, \frac{1}{T}\right)$$

This implies that in large samples, the sample autocorrelations are normally distributed. The confidence interval for $\hat{\rho}_h$ at a significance level $\alpha\%$ is given by:

$$prob\left[\hat{\rho}_h - z_{\alpha/2}\,(1/\sqrt{T}\,) \leq \rho_h \leq \hat{\rho}_h + z_{\alpha/2}\,(1/\sqrt{T})\right] = (1-\alpha)$$

where $z_{\alpha/2}$ is the critical value from the standard normal distribution. If zero belongs to this interval, we accept the null hypothesis $H_0: \rho_h = 0$, meaning the true autocorrelation at $lag\ h$ is not significantly different from zero at the $H_0: \rho_h = 0$ confidence level.

---

[1] Jenkins. G.M, watts.D.G., Spectral **Analysis and Its Applications**, San Francisco, Holden-Day, 1968, P 170.

[2] Bartlett. M. S, **On The Theoretical Specification of Sampling Properties of Autocorrelated Time Series**, Journal of the Royal Statistical society, series B, vol.27, 1946, P P27-41.

III- 3- 2- Box–Pierce (1970) and Ljung–Box (1978) Tests:

Instead of testing the statistical significance of the autocorrelation coefficients individually, as described above, it is possible to test joint hypotheses concerning whether a group of autocorrelation coefficients ρ_h, up to a given lag, are simultaneously equal to zero. This amounts to testing the following hypotheses[1]:

$$\begin{cases} H_0 : \quad \rho_1 = \rho_2 = \cdots = \rho_h \,, (h < T) \\ H_A : \ \exists k \ \in 1, 2, \ldots, h \,; \ \ \rho_k \neq 0 \end{cases}$$

This can be carried out using the statistic Q proposed by D. A. Box and G. E. P. Pierce (1970), which is defined as[2]:

$$\mathrm{Q} = \mathrm{T}.\sum_{\mathrm{k}=1}^{\mathrm{h}} \hat{\rho}_{\mathrm{k}}^2 \xrightarrow{\mathrm{a}} \chi_h^2$$

Alternatively, the modified statistic $Q^\star$, proposed by G. M. Ljung and G. E. P. Box (1978), is given by[3]:

$$\mathrm{Q}^\star = \mathrm{T}(\mathrm{T}+2)\sum_{k=1}^{h} (\mathrm{T}-\mathrm{K})^{-1} \,.\, \hat{\rho}_{\mathrm{k}}^2 \xrightarrow{\mathrm{a}} \chi_h^2$$

In large samples and under the null hypothesis H_0, both statistics Q and $Q^\star$ follow asymptotically a chi-square distribution with h degrees of freedom. However, in small samples, the statistic $Q^\star$ generally exhibits better performance than Q, as it possesses greater statistical power.

---

[1] Hayashi. F, **Econometrics**, Princeton University Press, Prenceton, N.J. 2000, P P 142-147.

[2] Box G. E, Pierce. D. A, **Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models**, Journal of the American Statistical Association, vol. 65, 1970, P P 1509-1526.

[3] Ljung. G. M. , Box. G. P, **on the Measure of Lack of Fit in Time Series Models**, Biometrica, vol 66, 1978, P P 66-72.

If the computed value of Q (or $Q*$) exceeds the critical value of the chi-square distribution χ_h^2 at a given significance level, the null hypothesis—stating that all true autocorrelation coefficients ρ_k for $(k = 1, \dots, h)$ are not significantly different from zero—is rejected; otherwise, it is not rejected.

IV – Basic Stochastic Processes:

Stochastic processes constitute the theoretical foundation of modern time series analysis. A time series is viewed as a realization of an underlying stochastic mechanism that evolves over time. Consequently, the objective of time series modeling is not limited to describing observed data, but rather to characterizing the probabilistic structure that generates the observations[1].

✓ White Noise Process

The white noise process is the simplest and most fundamental stochastic process in time series analysis. It represents a sequence of purely random shocks with no systematic pattern.

A stochastic process $\{\varepsilon_t\}$ is said to be a white noise process if it satisfies the following conditions:

$$E(\varepsilon_t) = 0, \;\; V(\varepsilon_t) = \sigma^2 < \infty\,, cov(\varepsilon_t, \varepsilon_{t-h}) = 0 \quad for\ h \neq 0$$

These assumptions imply that the process is weakly stationary, with a constant mean and variance, and an autocovariance function equal to zero at all non-zero lags[2]. In applied work, white noise is interpreted as a sequence of unpredictable innovations affecting the system.

✓ Random Walk Process:

---

[1] Hamilton, op.cit, p 3.

[2] Brockwell. P. J, Davis. R. A, **Time Series: Theory and Methods**, 2nd ed, New York: Springer-Verlag, 1991, p 15.

One of the most important non-stationary stochastic processes is the random walk. It is defined by the recursive relation:

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

where $\{\varepsilon_t\}$ is a white noise process.

The random walk plays a central role in economics and finance, particularly in modeling asset prices. Although the expected value of X_t may be constant over time, its variance increases linearly with ttt, which violates the stationarity condition[1]. As a result, the random walk is classified as a non-stationary process.

- ✓ **Random Walk with Drift:**

A generalization of the random walk is obtained by adding a constant drift term:

$$X_t = \beta_0 + X_{t-1} + \varepsilon_t$$

In this case, the process exhibits a deterministic trend driven by the constant β_0. The presence of drift reinforces the non-stationary nature of the process, since both the mean and the variance depend on time. Such processes are commonly encountered in macroeconomic time series[2].

- ✓ **Moving Average Process :**

The moving average process of order q, denoted $MA(q)$, is defined as:

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

---

[1] Hamilton, op.cit, p 47.
[2] Maddala. G. S, Kim. I. M, **Unit Roots, Cointegration, and Structural Change**, Cambridge: Cambridge University Press, 1998, p 23.

where $\{\varepsilon_t\}$ is a white noise process.

Moving average processes are weakly stationary by construction, since they are finite linear combinations of white noise terms. Their autocovariance function is non-zero only up to lag qqq, after which it becomes identically zero[1] .

- ✓ **Autoregressive Process :**

An autoregressive process of order p, denoted AR(p), is defined as:

$$X_t = \emptyset_1 X_{t-1} + \emptyset_2 X_{t-2} + \cdots + \emptyset_p X_{t-p} + \varepsilon_t$$

The stationarity of an $AR(p)$ process depends on the values of its parameters. Specifically, the process is weakly stationary if the roots of the characteristic polynomial lie outside the unit circle[2]. Autoregressive models are widely used to capture persistence and dynamic dependence in economic time series.

- ✓ **Integrated Processes :**

Integrated processes arise when differencing is required to achieve stationarity. A process X_t is said to be integrated of order one, denoted $I(1)$, if its first difference $\Delta X_t = X_t - X_{t-1}$ is stationary.

Random walks are typical examples of $I(1)$ processes. The concept of integration is fundamental in the analysis of non-stationary time series and forms the basis of unit root testing procedures[3].

---

[1] Box. G. E. P, Jenkins. G. M, Reinsel. G. C, **Time Series Analysis: Forecasting and Control,** 4th ed, Hoboken: John Wiley & Sons, 2008, p 31.
[2] Hamilton, op.cit, p 56.
[3] Dickey. D. A, Fuller, W. A, **Distribution of the Estimators for Autoregressive Time Series with a Unit Root**, Journal of the American Statistical Association, 74(366), United States, 1979, p 429.

Chapter Two:
Long Memory Models

Traditional time series analysis often assumes that economic processes exhibit short memory, meaning shocks dissipate rapidly and correlations decay exponentially, so the effect of past shocks diminishes over time. Under this framework, correlations between distant observations decay exponentially, a property that constitutes the theoretical foundation of classical ARMA and ARIMA models[1]. This assumption is valid for many practical applications; however, empirical evidence from macroeconomic and financial time series often contradicts it.

A substantial number of economic series display **persistent behavior**, where shocks have long-lasting effects and autocorrelations decay slowly. This phenomenon is commonly referred to as **long memory**, **long-range dependence**, or **long-term persistence**. In such processes, the autocorrelation function remains statistically significant even at large lags, indicating that distant past values continue to influence future outcomes[2].

The presence of long memory has been documented in various economic and financial variables. For example, inflation rates and interest rates often show persistence due to slow-moving expectations and gradual policy transmission mechanisms. In financial markets, volatility and trading volume frequently exhibit long-range dependence, especially in their second moments. Ignoring this feature can lead to biased parameter estimates, misleading inference, and poor forecasting performance[3].

From a methodological standpoint, the introduction of long memory models represents an important extension of classical time series theory. These models provide an intermediate framework

[1] Box. G. E. P, Jenkins. G. M, op.cit, p p 27-31.
[2] Granger. C. W, Joyeux. R, **An introduction to long-memory time series models and fractional differencing**, Journal of Time Series Analysis**,** 1(1), 1980, p p 15–18.
[3] Baillie. R.T, **Long memory processes and fractional integration in econometrics**, Journal of Econometrics**,** 73(1), 1996, p p 6–9.

between stationary short-memory processes and non-stationary unit root processes by allowing for **fractional degrees of integration**. This flexibility enables researchers to model persistence more accurately and to capture gradual decay patterns observed in real-world data[1].

In this context, linear long memory models such as the **ARFIMA (Autoregressive Fractionally Integrated Moving Average)** model have been widely adopted. By allowing the differencing parameter to take non-integer values, ARFIMA models generalize the ARIMA framework and offer a parsimonious way to represent long-term dependence while preserving stationarity under certain conditions[2].

More recently, advances in computational methods and artificial intelligence have introduced alternative approaches for modeling long memory. In particular, **recurrent neural networks** and **Long Short-Term Memory (LSTM)** architectures have been developed to explicitly address the limitations of traditional models in capturing long-term dependencies. LSTM networks incorporate memory cells and gating mechanisms that allow relevant information to be retained or discarded over long time horizons, thereby overcoming the vanishing gradient problem associated with standard recurrent neural networks[3].

These data-driven approaches do not rely on explicit assumptions about the underlying stochastic process, yet they have demonstrated strong empirical performance in modeling complex temporal dynamics. Consequently, LSTM models are increasingly viewed as a complementary tool to econometric long memory models, particularly in forecasting applications where nonlinearities and complex dependence structures are present[4].

---

[1]Beran.J, **Statistics for Long-Memory Processes**, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, V 61, CRC Press, 1994, p p 41-45.
[2] Hosking. J. R. M, **Fractional differencing**, Biometrika, 68(1), 1981, p p 167–169.
[3] Hochreiter. S, Schmidhuber. J, **Long short-term memory,** Neural Computation, 9(8), 1997, p p 1735–1780
[4] Zhang. A, Lipton. Z. C., Smola. A. J, **Dive into Deep Learning**, Release 0.14.X, Interactive deep learning textbook with code and mathematics, d2l.ai Foundation & Contributors, 2020, p p 372–375.

Within this framework, the present chapter aims to provide a structured and pedagogical treatment of long memory modeling. It begins with a conceptual and statistical characterization of long memory processes, followed by an examination of linear fractional integration models. The chapter then extends the analysis to neural network–based approaches, with particular emphasis on LSTM architectures, highlighting their ability to capture persistent dynamics in time series data. This integrated perspective equips the reader with both theoretical foundations and practical tools for analyzing and forecasting time series characterized by long-term dependence.

This chapter presents a structured overview of long memory modeling, starting with the theoretical foundations of fractional integration and then discussing linear models such as ARFIMA. The chapter concludes with a discussion of nonlinear and machine learning-based approaches, with a particular focus on Long Short-Term Memory (LSTM) networks.

I - Expanded Conceptual Introduction to Long Memory and Nonlinear Modeling:

In traditional time series analysis, most models rely on the assumption of **short memory**. This implies that the effect of a system shock dissipates quickly and that the autocorrelation function decays exponentially with increasing lag. Such behavior is consistent with classical ARMA and ARIMA models, where the dependence between observations weakens rapidly over time[1]. In practical terms, this means that distant past values have negligible influence on the present, allowing forecasts to converge rapidly to a steady-state mean.

However, empirical studies across macroeconomic and financial series have increasingly documented a different type of persistence, known as **long memory** or **long-range dependence**. Long memory refers to processes in which autocorrelations decay at a hyperbolic rate,

[1] Hamilton. J. D, **Time Series Analysis,** 2nd ed, Princeton University Press, 2020, p p 24-28.

much more slowly than the exponential decay characteristic of short-memory processes. As a result, correlations between observations remain significant even at long lags, indicating that shocks have prolonged effects that can last for many periods[1].

This phenomenon is particularly important in economics because many key variables exhibit slow adjustment and inertia. For instance, inflation rates, interest rates, GDP growth, unemployment, and volatility measures often show persistent patterns that cannot be explained by simple short-memory dynamics. Such persistence is not merely a statistical curiosity; it reflects structural features of the economic system such as slow-moving expectations, gradual policy transmission, and persistent behavioral patterns[2]. Therefore, accurately modeling long memory is essential for both forecasting and policy analysis.

- ✓ **Long Memory: Definition and Econometric Implications:**

The econometric framework that formalizes long memory is based on **fractional integration**. Unlike the classical ARIMA framework, where the differencing parameter d is restricted to integer values (0 or 1), fractional integration allows d to take **non-integer values**, thus representing a continuum between stationarity and non-stationarity. The fractional ARIMA (ARFIMA) model, for example, can capture long-range dependence through a parameter d that lies between 0 and 0.5 for stationary long-memory processes, and between 0.5 and 1 for non-stationary but mean-reverting processes[3].

From a forecasting perspective, long memory significantly alters the behavior of predictive distributions. Standard short-memory models tend to underestimate persistence, resulting in forecasts that converge

---

[1] Diebold. F. X, **Forecasting in Economics, Business, Finance and Beyond**, Princeton University Press, 2017, p p 215-218.

[2] Franses. P. H, Haldrup. N, **Nonlinear Time Series Models in Empirical Finance**, Cambridge University Press, 2019, p p 5-9.

[3] Brockwell. P. J, Davis. R. A, **Introduction to Time Series and Forecasting** , 3rd ed, Springer, 2016, p p 424-431.

too quickly to the mean and therefore fail to capture the long-run behavior of the series. Long-memory models, in contrast, yield forecasts with **slower decay** and a more realistic representation of long-run uncertainty[1]. This is particularly crucial in macroeconomic forecasting where policy decisions depend on long-horizon projections.

- ✓ **Nonlinearity and Structural Features:**

While long memory captures persistence, real-world time series often display additional complexities such as **nonlinearities**, **regime shifts**, and **structural breaks**. These features may coexist with long-range dependence and can distort the estimation and inference of linear models if ignored. For example, inflation dynamics may change across different monetary regimes, while financial volatility may exhibit asymmetric responses to shocks[2].

Nonlinear models provide a flexible framework to capture such behavior. They can model features such as:

- Regime switching (e.g., Markov-switching models);

- Threshold effects (e.g., TAR models);

- Volatility clustering (e.g., GARCH-type models);

- Asymmetry in responses to positive and negative shocks.

- ✓ **Neural Networks as Nonlinear Learning Models:**

In recent years, **machine learning and deep learning** methods have become prominent alternatives to traditional econometric models, especially for capturing complex nonlinear patterns without requiring explicit parametric assumptions. Neural networks, in particular, are

---

[1] Diebold. F. X, **Forecasting in Economics, Business, Finance and Beyond**, Princeton University Press, 2017, p p 221-224.

[2] Tsay. R. S, **Multivariate Time Series Analysis: With R and Financial Applications** , 3rd ed, Wiley, 2018, p p 438-422.

powerful universal approximators capable of modeling intricate relationships between variables and learning from data directly[1].

For sequential and time-dependent data, recurrent neural networks (RNNs) have been widely used. However, standard RNNs face limitations such as the **vanishing gradient problem**, which hampers learning long-term dependencies. To address this issue, Long Short-Term Memory (LSTM) networks were developed. LSTMs incorporate memory cells and gating mechanisms that allow the network to retain or forget information over long sequences, making them particularly suitable for long-term forecasting tasks[2].

✓ **Integrating Long Memory and Nonlinear Modeling:**

The integration of long memory and nonlinear modeling represents a modern direction in time series analysis. It recognizes that persistence and nonlinearities may occur simultaneously and that hybrid models may provide superior forecasting performance. The combined approach can take several forms:

- Hybrid econometric-machine learning models (e.g., ARFIMA + LSTM);

- Nonlinear fractional models that incorporate both long-range dependence and nonlinear structures;

- Ensemble methods that combine linear and nonlinear forecasts.

This integrated perspective is increasingly relevant in applied economic research because it allows models to adapt to the complex nature of real-world data, capturing both long-term persistence and nonlinear behaviors[3].

---

[1] Goodfellow. I, Bengio. Y, Courville. A, **Deep Learning**, MIT Press, 2016, p p 26-29.
[2] Goodfellow. I, Bengio. Y, Courville. A, op.cit, 361-364.
[3] Hamilton. J. D, op.cit, p p 456-459.

II - Artificial Neural Networks – General Framework:

Artificial Neural Networks (ANNs) constitute a class of computational models inspired by the structural and functional principles of biological neural systems. Their development emerged from the attempt to formalize human cognitive processes—such as learning, memory, and pattern recognition—within a mathematical and algorithmic framework. Unlike traditional statistical models, which rely on explicit assumptions regarding data distributions and linear relationships, neural networks are capable of learning complex, nonlinear mappings directly from observed data.

At their core, artificial neural networks are based on **massive parallel processing** carried out by a large number of simple computational units known as *neurons* or *nodes*. Each neuron performs elementary operations, yet when interconnected within a network, they collectively exhibit powerful learning and generalization capabilities. This distributed structure allows neural networks to store empirical knowledge and experiential information implicitly through the adjustment of connection weights, rather than through explicit symbolic rules[1].

From a formal perspective, artificial neural networks can be defined as **data-processing systems with properties analogous to those of natural neural networks**. They represent mathematical models that simulate aspects of human cognition by mimicking the way biological neurons receive signals, process information, and transmit outputs. In this sense, an ANN may be viewed as a processor composed of numerous simple processing elements whose collective behavior enables learning from data and adaptation to changing environments.

Each artificial neuron receives a set of inputs, multiplies them by adjustable weights, aggregates the resulting signals, and passes the outcome through an activation function. Learning occurs through the

[1] Haykin. S, **Neural Networks and Learning Machines**, 3rd ed, Pearson, New York, 2009, p p 28-31.

iterative modification of these weights in response to observed errors, allowing the network to approximate unknown functional relationships between inputs and outputs[1]. This learning mechanism makes neural networks particularly suitable for applications where the underlying data-generating process is complex or poorly understood.

Within the broader family of neural networks, **Recurrent Neural Networks (RNNs)** play a crucial role in modeling sequential and temporal data. Unlike feedforward architectures, RNNs incorporate feedback connections that allow information from previous time steps to influence current outputs. This property makes them especially relevant for time series analysis, where temporal dependence is a defining characteristic[2].

However, standard RNNs suffer from well-documented limitations, most notably the *vanishing gradient problem*, which hampers their ability to capture long-range dependencies. This limitation motivated the development of more advanced architectures, among which the **Long Short-Term Memory (LSTM)** network is the most prominent.

- ✓ **Long Short-Term Memory Neural Networks (LSTM):**

Long Short-Term Memory (LSTM) networks constitute a specialised class of recurrent neural network architectures explicitly developed to model sequential data exhibiting long-term dependencies. Originally introduced by Hochreiter and Schmidhuber (1997), the LSTM framework was proposed to overcome fundamental limitations of conventional recurrent neural networks, particularly the vanishing gradient problem that hampers effective learning over long sequences.

The core component of the LSTM architecture is the memory cell, which functions as an internal state capable of preserving information

[1] Bishop. C. M, **Pattern Recognition and Machine Learning**, Springer, New York, 2006, p p 227-230.

[2] Goodfellow. I, Bengio. Y, Courville. A, op.cit, p p 369-372.

across extended time horizons. Unlike standard recurrent units, the LSTM cell is a composite structure formed by interconnected processing units that allow information to be selectively stored, updated, or discarded as new observations become available[1].

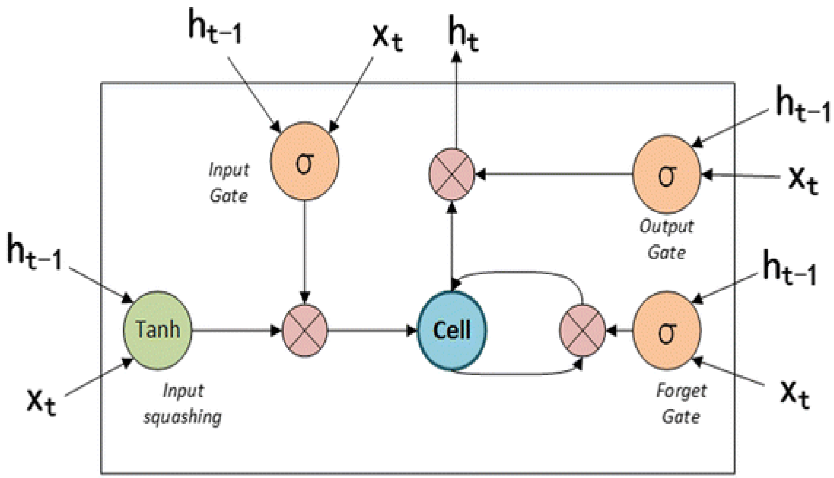
The distinctive feature of the LSTM model lies in its gate-based control mechanism. Each memory cell is equipped with a set of multiplicative gates that regulate the flow of information through the network. These gates determine which information should be incorporated into the cell state, which elements of past information should be forgotten, and which components of the internal state should be exposed as output. This adaptive regulation enables the network to maintain stable learning dynamics and capture long-range temporal dependencies effectively.

Specifically, the LSTM architecture comprises three principal gates: the **input gate**, which controls the incorporation of new information into the memory cell; the **forget gate**, which determines the extent to which previously stored information is retained or discarded; and the **output gate**, which governs the contribution of the internal cell state to the observable network output. Through this structured gating mechanism, LSTM networks achieve a robust balance between memory retention and flexibility, making them particularly suitable for time series forecasting and other sequence modelling tasks.

Figure (II-1) presents a structural representation of a Long Short-Term Memory (LSTM) cell, illustrating the memory cell and the gate-based control structure.

[1] Hochreiter. S, Schmidhuber. J, op.cit, p p 1736–1739.

Figure (II-1): Architecture of a Long Short-Term Memory (LSTM) unit showing the memory cell and gate-based control structure



Source: Hochreiter. S, Schmidhuber. J, **Long short-term memory**, Neural Computation, 9(8), 1997, p p 1736–1739.

✓ **Gates in the LSTM Architecture:**

Each LSTM cell contains three primary gates:

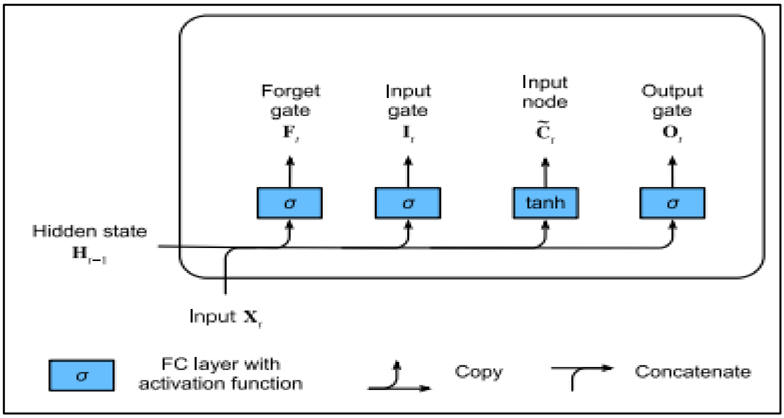
1- **Input Gate**, which controls the extent to which new information influences the internal state of the cell;

2- **Forget Gate**, which determines how much of the previous cell state should be retained or discarded;

3- **Output Gate**, which regulates how much of the internal state contributes to the observable output at a given time step.

Figure (II-2) illustrates the internal architecture of an LSTM cell, highlighting the interaction between the input, forget, and output gates and the cell state. This visual representation facilitates understanding of the gating mechanisms prior to the formal mathematical formulation.

Figure (II-2): Internal structure of an LSTM cell illustrating the input gate, forget gate, output gate, and memory cell state.



Source: Aston Zhang, Zachary C. Lipton, Mu Li, Alexander J. Smola, **Dive into Deep Learning**, Jupyter notebooks, 2020, p373.

Mathematically, let $X_t \in \mathbb{R}^{n \times d}$ denote the input matrix at time t, and let $H_{t-1} \in \mathbb{R}^{n \times h}$ represent the hidden state from the previous time step. The gates are defined as:

$$I_t = \sigma(X_t W_{xi} + H_{t-1} W_{hi} + b_i)$$
$$F_t = \sigma(X_t W_{xf} + H_{t-1} W_{hf} + b_f)$$
$$O_t = \sigma(X_t W_{xo} + H_{t-1} W_{ho} + b_o)$$

where W_x and W_h are weight matrices, b are bias vectors, and $\sigma(\cdot)$ denotes the sigmoid activation function, which maps inputs into the interval [0,1]. The use of sigmoid functions ensures that each gate operates as a soft filter, allowing partial rather than binary control over information flow.

✓ **Relevance to Long Memory and Time Series Modeling**

The ability of LSTM networks to selectively retain or discard information makes them particularly suitable for modeling time series exhibiting long memory or persistent temporal dependence. While classical econometric models—such as ARFIMA processes—capture long memory through fractional differencing, LSTM networks approach the problem from a data-driven and nonlinear perspective. As such, they can be viewed as complementary tools rather than substitutes for traditional long-memory models, especially in forecasting contexts where nonlinear dynamics play a significant role[1].

III- Long Short-Term Memory Neural Networks as Long Memory Models:

The analysis of time series characterized by persistence and long-range dependence has long occupied a central position in econometrics and applied statistics. Long memory processes are typically defined by a slow, hyperbolic decay of autocorrelations, implying that shocks to the system may have effects that persist over extended periods. While classical econometric models, such as fractionally integrated processes, provide a formal framework for capturing this behavior, they often rely on linear structures and strong parametric assumptions. In contrast, Long Short-Term Memory (LSTM) neural networks offer a flexible, nonlinear alternative capable of learning persistent temporal dependencies directly from data.

The ability of LSTM networks to function as long memory models stems from their internal architecture, which allows information to be stored, updated, and retrieved over long time horizons. Unlike standard recurrent neural networks, where repeated multiplication of gradients leads to rapid information decay, LSTM networks maintain a dedicated memory cell whose evolution is regulated through gating mechanisms. This design enables the network to preserve relevant historical information and to selectively forget irrelevant components, thereby generating a form of temporal persistence analogous to long memory behavior observed in economic and financial time series[2].

From a conceptual standpoint, long memory in LSTM models does not arise from an explicit parameter governing the rate of decay of

[1] Beran.J, op.cit, p p 41-45.
[2] Hochreiter. S, Schmidhuber. J, op.cit, p p 1735-1738.

dependence, as in ARFIMA models, but rather from the cumulative effect of learned gating decisions over time. When the forget gate assigns values close to unity across successive periods, past information remains embedded in the cell state, leading to a slow evolution of the internal representation. This mechanism allows LSTM networks to approximate long-range dependence in a data-driven manner, without imposing a predefined structure on the underlying process[1].

In the context of time series modeling, this implicit representation of persistence provides a significant advantage in environments characterized by nonlinear dynamics or structural instability. Economic and financial series often exhibit regime changes, asymmetric adjustments, and varying degrees of persistence over time. Classical long memory models typically assume a constant degree of fractional integration, which may fail to capture such heterogeneity. LSTM networks, by contrast, can adapt their memory retention behavior dynamically, allowing different segments of the series to exhibit distinct persistence patterns depending on the learned temporal context[2].

Empirical studies have demonstrated the effectiveness of LSTM models in capturing long-term dependencies in a wide range of applications, including macroeconomic forecasting, financial volatility modeling, and energy demand analysis. In these settings, LSTM-based models often outperform traditional linear specifications, particularly when the data exhibit nonlinear dependence structures or long-lasting effects of shocks. The superior performance of LSTM networks is commonly attributed to their capacity to integrate information across multiple time scales, thereby accommodating both short-term fluctuations and long-term trends within a unified modeling framework.

Nevertheless, it is important to emphasize that LSTM networks and classical long memory models serve different analytical purposes.

---

[1] Goodfellow. I, Bengio. Y, Courville. A, op.cit, p p 368–372.
[2] Baillie. R.T, op.cit, p p 6-8.

Fractionally integrated models provide a transparent and interpretable representation of persistence, with parameters that can be directly linked to theoretical concepts such as market efficiency or mean reversion. LSTM models, on the other hand, prioritize predictive performance and flexibility, often at the expense of interpretability. Consequently, LSTM networks should be viewed not as replacements for econometric long memory models, but as complementary tools that enrich the analysis of persistent time series phenomena[1].

From a pedagogical perspective, introducing LSTM networks within a long memory framework allows students to bridge traditional time series theory and modern machine learning techniques. By highlighting the conceptual parallels between persistent cell states and long-range dependence, learners can better appreciate how neural networks generalize classical ideas through nonlinear and adaptive mechanisms. This integration contributes to a more comprehensive understanding of contemporary time series analysis, where econometric rigor and computational intelligence increasingly converge.

IV – Comparative and Hybrid Perspectives: LSTM and Econometric Long Memory Models

The growing availability of high-frequency and high-dimensional economic data has intensified the need for modeling frameworks capable of capturing both persistent temporal dependence and complex nonlinear dynamics. Within this evolving landscape, classical econometric long memory models and modern neural network architectures, particularly Long Short-Term Memory (LSTM) networks, represent two complementary paradigms rather than competing alternatives. A comparative perspective is therefore essential to clarify their respective strengths, limitations, and potential for integration.

Traditional long memory models, such as ARFIMA processes, offer a parsimonious and theoretically grounded representation of

[1] Beran.J, op.cit, p p 41–44.

persistence through fractional integration. The fractional differencing parameter provides a direct and interpretable measure of long-range dependence, allowing researchers to assess the degree of persistence and mean reversion within a rigorous statistical framework[1]. This interpretability is particularly valuable in economic analysis, where parameters are often linked to behavioral mechanisms, institutional features, or policy implications. However, such models typically rely on linear structures and assume a constant degree of persistence over time, which may be restrictive in the presence of nonlinear dynamics or structural change.

LSTM networks approach the modeling of persistence from a fundamentally different perspective. Rather than imposing a predefined memory structure, they learn temporal dependence adaptively through internal state dynamics regulated by gating mechanisms. This allows LSTM models to accommodate time-varying persistence, nonlinear interactions, and regime-dependent behavior without requiring explicit assumptions about the data-generating process. As a result, LSTM networks often exhibit superior predictive performance in environments characterized by complex and evolving dynamics, such as financial markets or macroeconomic systems subject to policy shifts[2].

Despite their flexibility, LSTM networks present challenges that limit their standalone use in economic analysis. Most notably, their black-box nature complicates economic interpretation and inference. Unlike fractional integration parameters, the memory behavior of an LSTM model is embedded in high-dimensional weight matrices and nonlinear transformations, making it difficult to extract explicit measures of persistence. Moreover, neural networks typically require large datasets for reliable training and may be sensitive to hyperparameter choices, which introduces additional sources of uncertainty[3].

[1] Beran.J, op.cit, p p 41–45.
[2] Goodfellow. I, Bengio. Y, Courville. A, op.cit, 369-372.
[3] Bishop. C. M, op.cit, p p 231–234

These considerations have motivated a growing interest in hybrid modeling strategies that combine the interpretability of econometric long memory models with the flexibility of neural networks. One prominent approach involves integrating ARFIMA-type models with LSTM architectures, where the linear long memory component captures persistent dependence, while the neural network models residual nonlinear patterns. Such hybrid frameworks have been shown to improve forecasting accuracy while preserving a degree of economic interpretability[1].

Another avenue of integration is to use LSTM networks as nonlinear filters applied to fractionally differenced series. By first removing long-range dependence through fractional differencing and then modeling the transformed series with LSTM networks, researchers can isolate nonlinear dynamics without conflating them with persistence effects. This sequential strategy allows for a clearer separation between long memory and nonlinearity, facilitating both interpretation and prediction.

From a methodological standpoint, the coexistence of econometric and machine learning approaches reflects a broader shift in time series analysis toward pluralism. Rather than seeking a single dominant framework, contemporary research increasingly emphasizes the complementary use of models tailored to different aspects of the data. In this context, long memory models provide theoretical clarity and inferential rigor, while LSTM networks contribute adaptability and predictive power.

In educational settings, presenting these approaches side by side enables students to appreciate the evolution of time series modeling from linear, parametric structures to adaptive, data-driven systems. By understanding how LSTM networks generalize the notion of memory through nonlinear and dynamic mechanisms, learners can better grasp the conceptual continuity between classical econometric theory and modern artificial intelligence techniques. This integrative perspective not only enhances methodological competence but also prepares students to engage critically with the diverse modeling tools employed in contemporary economic research.

[1] Baillie. R.T, op.cit, p p 6–9.

Chapter Three: Non-Stationary Stochastic processes and Unit root tests

Chapter three: Non-Stationary Stochastic processes and Unit root tests

In time series analysis, a fundamental distinction exists between **stationary** and **non-stationary processes**, as the statistical properties of the series—such as the mean, variance, and autocovariances—determine the choice of appropriate modeling techniques. While stationary time series exhibit constant mean and variance over time, many real-world economic and financial variables do not. For example, macroeconomic aggregates, interest rates, exchange rates, and stock prices often display trends or persistent patterns that evolve over time.

Non-stationarity in a time series arises when its mean or variance (or both) depend on time. Such behavior may manifest as a **deterministic trend**, where the series follows a predictable path over time, or as a **stochastic trend**, in which the series evolves according to the accumulation of random shocks. Distinguishing between these two forms of non-stationarity is essential because it dictates the proper methods for transformation and subsequent statistical inference.

The presence of non-stationarity has critical implications for empirical modeling. Failure to account for it may lead to spurious regressions, unreliable parameter estimates, and misleading hypothesis tests, even when apparent correlations appear strong. Therefore, testing for non-stationarity is a crucial step before applying any time series modeling techniques.

In this chapter, we provide a systematic treatment of non-stationary stochastic processes. We first examine the concept of **non-stationarity** and its manifestations in economic time series. We then discuss the **types of trends**, distinguishing between deterministic and stochastic trends, and illustrate how each affects the statistical properties of the series. This naturally leads to the introduction of the **unit root concept**, which provides a formal framework to identify stochastic non-stationarity. Finally, we present the main **unit root testing procedures**, such as the Dickey-Fuller (DF) and Augmented Dickey-Fuller (ADF) tests, which allow researchers to determine whether a time series exhibits a unit root and to select the appropriate transformations to achieve stationarity.

I- Non-stationary Stochastic Time Series:

The analysis of stochastic time series should not be restricted exclusively to stationary processes. In practice, many economic and financial variables do not satisfy the stationarity property because their mean and/or variance depend on time t and therefore evolve over time. Such behavior constitutes a direct violation of the conditions of weak stationarity and is frequently encountered in applied econometric studies[1].

When the evolution of a time series shows a systematic upward or downward trend, the series is said to **exhibit a trend**. The presence of a trend is a major source of non-stationarity in time series data and has important implications for statistical inference, as it can lead to spurious results if not properly accounted for[2].

In this context, the econometric literature distinguishes between **two fundamental approaches** to modeling the trend component in non-stationary time series. The first approach is based on a deterministic time trend, where the series evolves according to a non-random function of time, typically specified as a linear or polynomial trend. In this case, non-stationarity arises from a predictable and systematic component that can, in principle, be removed by detrending[3].

The second approach considers the case in which the trend is of a stochastic nature, commonly referred to as a stochastic trend. Under this framework, the evolution of the series is driven by the accumulation of

---

[1] Hamilton, op.cit, p 47.
[2] Enders. W, **Applied Econometric Time Series**, 4th ed, Wiley, Hoboken, United States, 2015, p 181.
[3] Gujarati. D, Porter. C, **Basic Econometrics**, 5th ed, McGraw-Hill, New York, United States, 2009,

random shocks over time, as in the case of random walk processes. Such series are characterized by the presence of a unit root, and their non-stationarity cannot be eliminated by simple detrending but instead requires differencing to achieve stationarity[1].

The distinction between deterministic and stochastic trends is of central importance in time series analysis, as it determines both the appropriate transformation needed to induce stationarity and the choice of suitable statistical tests. This distinction naturally motivates the introduction of the unit root concept and its associated testing procedures, which constitute the core subject of the subsequent sections of this chapter.

II- Types of Non-stationarity in Time Series:

Non-stationary time series can exhibit different forms of behavior that violate the conditions of weak stationarity. Understanding the type of non-stationarity is crucial for choosing appropriate modeling and transformation techniques. Broadly, non-stationarity in time series is classified into two main types: deterministic trends and stochastic trends.

II- 1- Time Series Stationary around a Deterministic Trend (TS):

A time series "X_t" can often be represented as the sum of a deterministic trend function $f(t)$ and a stationary stochastic component v_t, such that[2]:

$$X_t = f(t) + v_t$$

In this framework, the function $f(t)$ captures the systematic and predictable evolution of the series over time, while the component v_t represents random fluctuations that are stationary in nature. A time

[1] Hamilton, op.cit, p 48.
[2] Maddala. G.S, op.cit, p 528.

series of this form is said to be trend-stationary (TS), since its non-stationarity is entirely due to the deterministic trend component [1].

- ✓ **Linear Deterministic Trend:**

The simplest and most frequently encountered case arises when the deterministic trend $f(t)$ is linear in time and the stochastic component v_t is a white noise process with zero mean and constant variance "σ_v^2". In this case, the series can be written as:

$$X_t = \alpha + \beta t + v_t \dots\dots\dots (*)$$

where α denotes the intercept, β measures the slope of the deterministic trend, and v_t is a purely random disturbance term[2].

It is straightforward to verify that the series X_t defined in equation $(*)$ is **non-stationary**, since its expected value depends explicitly on time:

$$(X_t) = E(\alpha + \beta t + v_t) = \alpha + \beta t$$

Hence, the mean of the process varies systematically with t, violating the condition of mean stationarity[3].

- ✓ **Detrending and Stationarity:**

However, if the parameters α and β are known or consistently estimated—typically by estimating equation $(*)$ using **ordinary least squares (OLS)**—the series X_t can be transformed into a stationary one by subtracting the estimated deterministic trend $(\hat{\alpha} + \hat{\beta}t)$ from the original observations:

[1] Enders. W, op.cit, p 181.

[2] Gujarati .D.N, op.cit, p 746.

[3] Hamilton, op.cit, p 47.

$$\tilde{X}_t = X_t - (\hat{\alpha} + \hat{\beta}t) = v_t$$

The resulting series, known as the **detrended series**, corresponds to the OLS residuals and is stationary by construction. In particular, it satisfies the orthogonality conditions :

$$\sum \hat{v}_t = 0 \quad , \quad \sum t\hat{v}_t = 0$$

which reflect the elimination of both the intercept and the linear trend components[1].

- ✓ **Nature of Shocks in Trend-Stationary Processes:**

An important characteristic of trend-stationary series is that the impact of a random shock affecting v_t is **temporary (transient)**. Although such a shock may cause short-run deviations from the deterministic trend, its effect diminishes over time, and the series eventually reverts to its long-run deterministic path[2].

This property sharply distinguishes trend-stationary processes from series characterized by a stochastic trend (unit root processes), in which shocks have permanent effects on the level of the series.

- ✓ **Higher-Order Deterministic Trends:**

In practice, the deterministic trend $f(t)$ need not be linear. It may take the form of a higher-order polynomial in time, such as a quadratic or cubic trend:

$$X_t = \alpha + \beta_1 t + \beta_2^2 t + \dots\dots\dots. + v_t$$

Even in such cases, provided that the stochastic component v_t is

[1] Enders. W, op.cit, p 183.
[2] Maddala. G. S, Kim. I. M, op.cit, p 24.

Even in such cases, provided that the stochastic component v_t is stationary, the series remains trend-stationary. Once the deterministic component is properly estimated and removed, the residual series exhibits stable statistical properties over time[1].

II- 2- Difference Stationary Time Series (DS):

As noted earlier, non-stationarity in stochastic time series does not arise solely from a deterministic time trend. In many cases, the source of non-stationarity is stochastic in nature and cannot be represented by a deterministic function of time. Under this approach, a large number of economic and financial time series that do not satisfy stationarity conditions can be transformed into stationary series by applying a first-difference filter multiple times.

Such time series are referred to as difference stationary processes**,** or integrated processes of order "d". They can be expressed as follows[2]:

$$(1 - L)^d X_t = \beta + v_t$$

where:

L: denotes the lag operator;

d: represents the number of times the first difference is taken to achieve stationarity and is known as the order of integration;

v_t: is a stationary stochastic process, typically assumed to be white noise.

This framework is particularly important in applied econometric analysis, as many macroeconomic variables—such as output, consumption, prices, or exchange rates—exhibit non-stationary

[1] Enders. W, op.cit, p 184.
[2] Hamilton, op.cit, p 47.

behavior that cannot be adequately captured by deterministic trends alone[1].

- ✓ **Random Walk with Drift:**

To highlight the fundamental properties of difference stationary processes, the random walk with drift is commonly used as a benchmark model. It is defined as:

$$X_t = X_{t-1} + \beta + v_t \dots \dots \dots . (**)$$

where:

β: is a constant drift parameter;

v_t: is a white noise process with zero mean and variance σ_v^2.

By recursive substitution, equation (∗∗) can be rewritten as:

$$X_t = X_0 + \beta t + \sum_{j=1}^{t} v_j \dots \dots \dots (***)$$

This representation shows that X_t consists of an initial value, a deterministic linear component induced by the drift, and a cumulative sum of random shocks. This cumulative nature explains the non-stationary behavior of the process.

- ✓ **Statistical Properties:**

From equation (∗∗∗), the mean and variance of X_t can be derived as follows:

$$E(X_t) = X_0 + t\,\beta$$

$$\text{var}(X_t) = t\ \sigma_v^2$$

[1] Maddala. G. S, Kim. I. M, op.cit, p 17.

It is clear from these expressions that both the mean and the variance depend explicitly on time. This constitutes a direct violation of the stationarity conditions, since neither moment remains constant over time[1].

✓ **First Differences and Stationarity:**

Using the lag operator, equation $(**)$ can be rewritten as:

$$X_t - X_{t-1} = (1 - L)X_t = \Delta X_t = \beta + v_t$$

This Equation clearly shows that the first difference of the series X_t, denoted by ΔX_t, is stationary provided that v_t is white noise. Consequently, the process X_t is said to be integrated of order one, $I(1)$, while its first difference is an $I(0)$ process.

✓ **Persistence of Random Shocks:**

Unlike trend-stationary (TS) processes, the error term in difference stationary processes is not transitory. Instead, it consists of an accumulation of past random shocks over time. This feature gives rise to one of the most important characteristics of DS processes, namely the persistence of random shocks, whereby the effect of a shock does not dissipate as time passes. In other words, such processes exhibit infinite memory[2].

III- The Concept of Unit Root:

The concept of a **unit root** is central to the analysis of non-stationary stochastic time series and provides a formal statistical characterization of difference stationary processes. In time series econometrics, non-stationarity is often associated with the presence of a unit root in the data-generating process[3].

---

[1] Gujarati. D, Porter. C, op.cit, p 740.
[2] Enders. W, op.cit, p 182.
[3] Hamilton, op.cit, p 488.

To illustrate this concept, consider the first-order autoregressive process AR(1):

$$X_t = \phi X_{t-1} + v_t$$

where v_t is a white noise process with zero mean and constant variance.

The stationarity properties of this process depend crucially on the value of the autoregressive parameter ϕ. If $|\phi| > 1$, the process is weakly stationary, with a constant mean and variance, and shocks to the system have only temporary effects. However, when $\phi = 1$, the process becomes non-stationary and is said to contain a **unit root**.

- ✓ **Unit Root and Random Walk:**

The model indicates that the current value is equal to the previous value plus a random disturbance. This leads to[1]:

$$X_t = X_{t-1} + v_t$$

And by recursively substituting X_{t-1}, we can express X_t in terms of earlier values:

$$X_{t-1} = X_{t-2} + v_{t-1}$$

When we substitute in X_t, we find:

$$X_t = X_{t-2} + v_{t-1} + v_t \Rightarrow X_{t-2} = X_{t-3} + v_{t-2}$$

$$X_t = X_{t-2} + X_{t-3} + v_t + v_{t-1} + v_{t-2}$$

And by continuing this substitution, it leads to:

[1] Gujarati. D. N , **Basic Econometrics**, 4th , The Mc graw-hill higher education, America,2003, p p 814-816.

$$X_t = X_0 + \sum_{i=1}^{t} (v_i)$$

And The variance of Y_t can be expressed as:

$$Var(X_t) = t * \sigma_v^2$$

The covariance between different time points X_t and X_s is given by:

$$\text{Cov}(X_t ; X_s) = \sigma_v^2 \min(t, s) \quad \text{for } t \neq s$$

This indicates that the variance increases with time, leading to non-constant variance and instability.

The instability in difference stationary models arises from random disturbances. If a shock occurs at any point in time, it has a lasting effect on the series, deviating it from its mean of zero. This phenomenon is known as random instability.

As a result, the Xt series is unstable, and based on the last formula of the Xt series, Xt is the sum of the errors for the series from 1 to t. If a disturbance occurs at a certain time, it continues afterwards and deviates the series away from the mean zero. This causes the variance to be non-constant over time and makes the series unstable. Given that the cause of instability in this type of model is due to random disturbances, it is called random instability. and we write:

$$X_t = \beta + X_{t-1} + v_t$$

Where: β represents the constant term.

- If: $\beta = 0$, The model is a simple difference stationary model without a drift.

- If: $\beta \neq 0$, then the model includes a drift, indicating a systematic trend in the series.

By taking the first difference of the series, we obtain:

$$X_t - X_{t-1} = \beta + v_t$$

$$\Delta X_t = \beta + v_t$$

The differenced series ΔX_t is stationary. Consequently, we say that the original series X_t is an integrated series of the first degree, denoted as:

$$X_t \sim I(1)$$

In case the series X_t is integrated of degree d , we write:

$$\nabla^d X_t = \beta + v_t$$

This indicates that the series X_t is integrated of order d, written as:

$$X_t \sim I(d)$$

For example : if $d = 2$

$$\nabla^2 X_t = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2})$$

Trend stationarity and difference-stationary models are two fundamental concepts in time series analysis. Trend stationary models exhibit a deterministic trend that can be removed to achieve stationarity, while difference stationary models require differencing to stabilize the series. Understanding these models is crucial for effective time series forecasting and analysis, as they provide insights into the underlying structure and behavior of the data. By identifying the type of model that best describes the series, analysts can apply appropriate methods for stabilization and forecasting.

- ✓ **Characteristic Polynomial Interpretation:**

From a polynomial perspective, the $AR(1)$ model can be written using the lag operator L as:

$$\mathrm{v}_t = (1 - \phi \mathrm{L})\mathrm{X}_t$$

The characteristic equation associated with this process is:

$$1 - \phi z = 0$$

The root of this equation is $z = 1/\phi$. When $\phi = 1$, the root lies exactly on the unit circle, hence the term **unit root**. In this situation, the process does not satisfy the stationarity condition, which requires all roots to lie outside the unit circle[1].

- ✓ **Unit Root and Differencing:**

A defining feature of unit root processes is that they can be transformed into stationary processes by differencing. Applying the first-difference operator yields:

$$\Delta X_t = (1 - L)X_t = v_t$$

which is stationary if v_t is white noise. Therefore, a time series containing a single unit root is integrated of order one, denoted $I(1)$. More generally, a process containing d unit roots is integrated of order d, written as $I(d)$ [2].

- ✓ **Economic Interpretation:**

[1] Box. G. E. P, Jenkins. G. M, Reinsel. G. C, op.cit, p 92.
[2] Maddala. G. S, Kim. I. M, op.cit, p 21.

From an economic perspective, the presence of a unit root implies that shocks have **permanent effects** on the level of the series. Unlike trend-stationary processes, where deviations from the trend eventually vanish, unit root processes exhibit **shock persistence**, meaning that the impact of random disturbances does not diminish over time[1].

This distinction has important implications for economic modeling, forecasting, and policy analysis, particularly in macroeconomics, where many key variables—such as GDP, inflation, and exchange rates—are often found to be integrated processes rather than trend-stationary ones.

In practice, distinguishing among stationary, trend-stationary, and unit-root processes is a fundamental step in time-series analysis. This motivates the development of formal statistical procedures—known as **unit root tests**—designed to assess the presence of unit roots in observed data. These tests are discussed in the following subsection.

IV - Unit Root Tests:

In time series econometrics, many economic and financial variables exhibit persistent behavior over time, which often implies non-stationarity. A common source of non-stationarity is the presence of a unit root in the stochastic process generating the data. Ignoring this characteristic may lead to spurious regressions and misleading statistical inference, especially in empirical economic analysis.

Unit root tests are therefore essential tools for examining the stochastic properties of time series and determining their order of integration. These tests allow researchers to distinguish between stationary and non-stationary processes, with the latter becoming stationary only after differencing. The distinction is crucial for model specification, hypothesis testing, and forecasting.

[1] Gujarati. D, Porter. C, op.cit, p 748.

This unit presents the main theoretical foundations of unit root processes, beginning with the random walk model, and introduces the most widely used unit root tests in applied econometrics, including the Dickey–Fuller test, the Augmented Dickey–Fuller test, the Phillips–Perron test, and the KPSS test. Special attention is given to the formulation of hypotheses, the computation of test statistics, and the practical interpretation of results in empirical applications.

IV -1- The Simple Dicky Fuller Test(1979) :

The Simple Dickey-Fuller Test, developed by David Dickey and Wayne Fuller in 1979, is a widely used statistical method for testing the presence of a unit root in a time series. This test helps determine whether a series is stationary or non-stationary, which is crucial for effective time series analysis and modeling.

The primary goal of the Dickey-Fuller Test is to identify the nature of instability in a time series. By incorporating different components—such as a constant term and a trend—analysts can discern whether the instability is due to deterministic trends or random fluctuations. The test is based on three distinct models, each accommodating different characteristics of the time series.

The Dickey-Fuller Test utilizes three models to assess the presence of a unit root[1]:

Model 1: No Constant or Trend

$$\Delta X_t = \varphi X_{t-1} + \varepsilon_t \dots\dots\dots\dots \ (\mathbf{I})$$

This model represents a simple autoregressive process where the current value X_t depends solely on its lagged value X_{t-1} and a white

[1] Régis bourbonnais, Michel Terraza, **Analyse des séries temporelles**, 3rd edition, Dunod, France, 2010 , p 163.

noise error term ε_t . The first difference ΔX_t is the change between two consecutive observations.

Model 2: With Constant

$$\Delta X_t = \varphi X_{t-1} + \gamma + \varepsilon_t \dots\dots\dots\dots (\textbf{ II })$$

This model includes a constant term γ, which accounts for a non-zero mean in the time series. The presence of the constant allows for shifts in the mean level of the series, indicating that the series can have a deterministic trend.

Model 3: With Constant and Trend

$$\Delta X_t = \varphi X_{t-1} + \lambda + \delta t + \varepsilon_t \dots\dots\dots\dots (\textbf{ III })$$

This model incorporates both a constant term λ and a linear trend δt. The inclusion of the trend component allows for a more flexible representation of the time series, accommodating cases where the series exhibits a deterministic trend over time.

- ✓ **Error Term Assumption:**

In all three models, the error term ε_t is assumed to follow a Brownian motion process:

$$\varepsilon_t \rightsquigarrow BB(0, \sigma_\varepsilon^2)$$

This assumption ensures that the error term has zero mean and constant variance.

- ✓ **Null Hypothesis:**

The null hypothesis for this test says that the series X_t has a unit root, implying that it is non-stationary:

$$\{H_0 : \phi_1 = 1$$

So If the null hypothesis is true, it suggests that shocks to the time series have a permanent effect, and the series does not revert to a mean over time.

- ✓ **Test Statistic:**

The test statistic is computed as:

$$t_{\widehat{\phi_1}} = \frac{\widehat{\phi_1} - 1}{\hat{\sigma}_{\widehat{\phi_1}}}$$

Where :

$\widehat{\phi_1}$: is the estimated coefficient from the regression;

$\hat{\sigma}_{\widehat{\phi_1}}$: The standard error of the estimated coefficient $\widehat{\phi_1}$.

The calculated statistic is then compared with critical values derived from the DF distribution to determine whether to reject the null hypothesis

The Simple Dickey-Fuller Test is a fundamental tool in time series analysis for assessing stationarity. By using different models that account for constants and trends, the test provides insights into the nature of instability in a series. The results of the test help analysts determine whether to apply differencing or other techniques to stabilize the series for further analysis. Understanding the implications of the null hypothesis and the calculation of the test statistic is essential for interpreting the outcomes of the Dickey-Fuller Test effectively.

IV -2- Augmented Dickey–Fuller Test (ADF):

The Augmented Dickey-Fuller (ADF) test is an extension of the Simple Dickey-Fuller test, designed to address the limitations of the original test by accounting for potential autocorrelation in the error term. While the Simple Dickey-Fuller test assumes that the error term is white noise, this assumption may not hold true in practice. The ADF test incorporates lagged differences of the dependent variable to correct for autocorrelation, making it a more robust method for testing the presence of a unit root in a time series.

The ADF test is used to determine whether a time series is stationary or non-stationary by testing the null hypothesis that the series has a unit root. If the null hypothesis is rejected, it indicates that the series is stationary. The ADF test employs three models, each incorporating lagged differences of the dependent variable to account for autocorrelation[1]:

- ✓ **ADF Test Models:**

Model 4: No Constant or Trend

$$\Delta X_t = \varphi X_{t-1} + \sum_{j=1}^{p} \alpha_j + \Delta X_{t-j} + \eta_t \dots\dots\dots (\mathbf{IV})$$

This specification includes the lagged level X_{t-1} and a finite number of lagged first differences ΔX_{t-j} in order to capture serial correlation in the disturbance term η_t. It is appropriate when the series fluctuates around a zero mean without any deterministic trend.

Model 5: With Constant

$$\Delta X_t = \varphi X_{t-1} + \gamma + \sum_{j=1}^{p} \alpha_j + \Delta X_{t-j} + \eta_t \dots\dots\dots (\mathbf{V})$$

[1] John D. Levendis, **Time Series Econometrics Learning Through Replication**, Springer Texts in Business and Economics, New Orleans, LA, USA, 2018, p151.

This model introduces a constant term γ, allowing the series to have a non-zero mean. It is suitable for series that do not exhibit a deterministic trend but fluctuate around a constant level.

Model 6: With Constant and Trend

$$\Delta X_t = \varphi X_{t-1} + \lambda + \delta t + \sum_{j=1}^{p} \alpha_j + \Delta X_{t-j} + \eta_t \dots\dots\dots (\textbf{VI})$$

In this specification, both a constant λ and a linear time trend δt are included. This formulation is appropriate when the series exhibits a deterministic trend in addition to stochastic dynamics and autocorrelation.

- ✓ **Assumption on the Error Term:**

In all three models, the error term η_t is assumed to be a white noise process with zero mean and constant variance.

$$\eta_t \sim BB(0, \sigma_\eta^2)$$

The inclusion of lagged differences ensures that any autocorrelation in the residuals is adequately captured, leading to well-behaved error terms.

- ✓ **Lag Length Selection:**

The number of lagged differences p plays a crucial role in the specification of the ADF test. An insufficient number of lags may leave residual autocorrelation, while too many lags reduce the power of the test. Therefore, information criteria are commonly used to determine the optimal lag length[1].

---

[1] Enders. W, op.cit, p 260.

- Akaike Information Criterion (AIC):

$$AIC = n\, ln\left(\frac{SCR}{n}\right) + 2(p+q)$$

Where:

n: Sample size;

SCR: Sum of squared residuals;

p: Number of lagged terms included in the model;

q: Number of parameters in the model.

- Schwarz Criterion (SC):

$$SC = n\, ln\left(\frac{SCR}{n}\right) + (p+q)Ln\,(n)$$

The SC penalizes the number of parameters more heavily than the AIC, often leading to a simpler model.

- ✓ **Null Hypothesis and Test Statistic:**

The null hypothesis of the ADF test is given by:

$$\{H_0 : \phi_1 = 1$$

Under this hypothesis, the series X_t contains a unit root and is non-stationary. If the null hypothesis holds, shocks to the series have permanent effects and the series does not revert to a long-run mean[1].

The test statistic is computed as:

$$t_{\widehat{\phi_1}} = \frac{\widehat{\phi_1} - 1}{\hat{\sigma}_{\widehat{\phi_1}}}$$

[1] Hamilton, op.cit, p 517.

Where:

$\widehat{\phi_1}$: is the estimated coefficient of X_t ;
$\hat{\sigma}_{\widehat{\phi_1}}$: is the standard error of $\widehat{\phi_1}$.

The calculated statistic is compared with the critical values tabulated by Dickey and Fuller, which do not follow the standard Student's t-distribution.

The Augmented Dickey–Fuller test constitutes a powerful and flexible extension of the Simple Dickey–Fuller test. By incorporating lagged differences of the dependent variable, it effectively addresses the problem of autocorrelation in the error term and provides a more reliable assessment of stationarity in time series data. Proper lag selection based on information criteria such as AIC and SC ensures correct model specification and enhances the validity of the test results.

example:

We tried to test the stasionarity of. for that we used the The Dickey-Fuller test using the previous example series, than we represented the results in the following tables:

Table (III-1): Results of the Dickey-Fuller test for the original time series

Null Hypothesis: IPC_GLOBAL has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic - based on SIC, maxlag=0)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-0.643238	0.9752
Test critical values:	1% level	-3.996918	
	5% level	-3.428739	
	10% level	-3.137804	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(IPC_GLOBAL)
Method: Least Squares
Date: 05/10/24 Time: 19:48
Sample (adjusted): 2004M02 2023M12
Included observations: 239 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
IPC_GLOBAL(-1)	-0.005469	0.008503	-0.643238	0.5207
C	-0.000236	0.000442	-0.534067	0.5938
@TREND("2004M01")	5.67E-06	2.74E-06	2.065548	0.0400

R-squared	0.017909	Mean dependent var	0.000185
Adjusted R-squared	0.009586	S.D. dependent var	0.002699
S.E. of regression	0.002686	Akaike info criterion	-8.988920
Sum squared resid	0.001703	Schwarz criterion	-8.945283
Log likelihood	1077.176	Hannan-Quinn criter.	-8.971336
F-statistic	2.151787	Durbin-Watson stat	0.314278
Prob(F-statistic)	0.118552		

Based on the results of the Dickey-Fuller tests for the series shown in Table (III-1), there is an auto-correlation of errors and the model is not significate so we accept hypothesis H_0, and admit that the series is not stationary, so we will apply the first-order difference to the series. as the following:

Table (III-2): Results of the Dickey-Fuller test for the time series (first-order difference)

Null Hypothesis: IPC_GLOBAL has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic - based on SIC, maxlag=1)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.687411	0.0250
Test critical values:	1% level	-3.997083	
	5% level	-3.428819	
	10% level	-3.137851	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(IPC_GLOBAL)
Method: Least Squares
Date: 05/10/24 Time: 19:49
Sample (adjusted): 2004M03 2023M12
Included observations: 238 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
IPC_GLOBAL(-1)	-0.016582	0.004497	-3.687411	0.0003
D(IPC_GLOBAL(-1))	0.851022	0.034392	24.74484	0.0000
C	0.000421	0.000235	1.792615	0.0743
@TREND("2004M01")	3.10E-06	1.46E-06	2.124829	0.0346

R-squared	0.729282	Mean dependent var	0.000172
Adjusted R-squared	0.725811	S.D. dependent var	0.002697
S.E. of regression	0.001412	Akaike info criterion	-10.27046
Sum squared resid	0.000467	Schwarz criterion	-10.21210
Log likelihood	1226.185	Hannan-Quinn criter.	-10.24694
F-statistic	210.1222	Durbin-Watson stat	1.839371
Prob(F-statistic)	0.000000		

Based on the results of the Dickey-Fuller tests for the series, first-order difference shown in Table (II-2), we reject hypothesis H_0, and assume that the series is stationner.

IV -3- Perron and Phillips test (1988):

The Phillips–Perron (PP) test, introduced by Phillips and Perron in 1988, is a widely used statistical procedure for testing the presence of a unit root in time series data. The test constitutes an extension of the Dickey–Fuller framework and was specifically developed to overcome some of its limitations, particularly those related to serial correlation and heteroskedasticity in the error term.

Unlike the Augmented Dickey–Fuller (ADF) test, which addresses autocorrelation by adding lagged differences of the dependent variable, the Phillips–Perron test applies non-parametric corrections directly to the test statistic, without modifying the underlying regression model [1].

Unit root testing plays a central role in time series analysis, as the use of non-stationary variables may lead to spurious regression results and invalid statistical inference. The existence of a unit root implies that random shocks have permanent effects on the series, preventing it from reverting to a stable long-run mean[2].

- ✓ **Model Specification:**

The Phillips–Perron test is based on the same three regression models used in the Dickey–Fuller test, all estimated by Ordinary Least Squares (OLS):

Model (1): No Constant and No Trend

$$X_t = \phi_1 X_{t-1} + \varepsilon_t$$

Model (2): With Constant

$$X_t = \phi_1 X_{t-1} + c + \varepsilon_t$$

Model (3): With Constant and Linear Trend

$$X_t = \phi_1 X_{t-1} + c + bt + \varepsilon_t$$

where ε_t denotes the stochastic error term.

---

[1] Phillips P. C, Perron .P, **Testing for a unit root in time series regression**, Biometrika, 75(2), 1988, p p 335–346.

[2] Hamilton, op.cit , p 514.

After estimating these models, the coefficient $\hat{\phi}_1$ and the corresponding residuals $\hat{\varepsilon}_t$ are obtained. These residuals are subsequently used to compute non-parametric corrections that account for serial correlation and heteroskedasticity.

✓ **Steps of the Phillips–Perron Test:**

The PP test follows a systematic four-step process designed to estimate the presence of a unit root while correcting for autocorrelation and heteroscedasticity in the residuals. Below is a detailed breakdown of each step[1] :

Step 1: Estimation of the Dickey–Fuller Regression

One of the three models above is estimated using OLS, and the conventional Dickey–Fuller test statistic associated with $\hat{\phi}_1$ is computed.

Step 2: Estimation of Short-Run Variance

The short-run variance of the residuals is calculated as follows:

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n}\sum_{t=1}^{n} \hat{\varepsilon}_t^2$$

Where:

n : is the sample size;

$\hat{\varepsilon}_t$: are the estimated residuals.

This calculation provides an estimate of the variance of the residuals, which are crucial for the subsequent steps in the testing

[1] Elagueb Mohammed, **time series analysis Lectures and Applications in Economics**, economic department, university ziane Achour, elDjelfa, 2017, p p 45-46.

process.

Step 3: Estimation of Long-Run Variance

The third step focuses on estimating the adjustment coefficient or the long-term variance of the residuals. This is necessary to account for potential autocorrelation in the residuals, which may bias the test results. The long-term variance is estimated using the Newey–West type estimator:

$$\hat{s}_t^2 = \hat{\sigma}_\varepsilon^2 + 2\sum_{i=1}^{l}\left(1 - \frac{i}{\ell + 1}\right)\frac{1}{n}\sum_{t=i+1}^{n} \hat{\varepsilon}_t \hat{\varepsilon}_{t-i}$$

$$\hat{s}_t^2 = \frac{1}{n}\sum_{t=1}^{n} e_t^2 + 2\sum_{i=1}^{l}\left(1 - \frac{i}{\ell + 1}\right)\frac{1}{n}\sum_{t=i+1}^{n} \hat{\varepsilon}_t \hat{\varepsilon}_{t-i}$$

The truncation $lag\ \ell$ is commonly selected according to the rule:

$$\ell \approx 4\left(\frac{n}{100}\right)^{2/9}$$

This choice ensures a consistent estimate of the long-run variance.

Step 4: Computation of the Phillips–Perron Test Statistic

The corrected Phillips–Perron test statistic is given by:

$$t_{\widehat{\phi_1}}{}^{*} = \sqrt{\theta}\ \frac{\widehat{\phi_1} - 1}{\hat{\sigma}_{\widehat{\phi_1}}} + \frac{n(\theta - 1)\hat{\sigma}_{\widehat{\phi_1}}}{\sqrt{\theta}}$$

Where:

$$\theta = \frac{\sigma_\varepsilon^2}{\hat{s}_t^2}$$

This correction adjusts the conventional Dickey–Fuller statistic to reflect the presence of serial correlation and heteroskedasticity in the error process.

- ✓ **Hypotheses and Decision Rule:**

The hypotheses of the Phillips–Perron test are formulated as:

$$\begin{cases} H_0: \phi_1 = 1 & (\text{unit root, non} - \text{stationary series}) \\ H_1: \phi_1 < 1 & (\text{stationary series}) \end{cases}$$

The computed statistic $t_{\widehat{\phi_1}}{}^*$ is compared with the critical values tabulated by MacKinnon, which depend on whether a constant and/or a trend is included in the regression.

The Phillips-Perron test is an essential tool in the arsenal of time series analysis, providing a robust method for testing the presence of unit roots in data. By correcting for autocorrelation and heteroscedasticity, the PP test enhances the reliability of statistical inferences drawn from time series models. The systematic four-step process, including model estimation, variance calculations, adjustment for autocorrelation, and the final test statistic computation, ensures that analysts can make informed decisions regarding the stationarity of their time series data.

This test is particularly valuable in empirical research across various fields, including economics, finance, and environmental studies, where understanding the underlying properties of time series data is crucial for effective modeling and forecasting. By employing the Phillips-Perron test, researchers can better navigate the complexities of non-stationary data and implement appropriate corrective measures to achieve reliable results.

Example:

In this example we will test the stastionarity of the same series with the pp test and display the results in the following tables:

Table (III-3): Results of the Phillips–Perron Test at Level

The decision	Prob-p	T calculated	The value of regression marameter's	Auto-correlation	The model
Model rejected	0.52	0.0085	-0.005	Doesn't exist	(1)
	0. 59	0.00044	-0.0002		
	0.04	2.74^{E}-06	5.67^{E}-06		
Model accepted series non-statioraire	0.84	0.191266	0.001503	Doesn't exist	(2)
	0.78	0.27662	0.00011		
Model accepted series non-statioraire	0.299	0.00336	0.00347	Doesn't exist	(3)

From the results reported in the table above, it can be observed that the series is **non-stationary at level**. Consequently, the first-order difference of the series is computed, and the Phillips–Perron test is applied again.

Table (III-4): Results of the Phillips–Perron Test after First-Order Differencing

The decision	Prob-p	T calculated	The value of regression marameter's	Auto-correlation	The model
Model rejected series non-statioraire	0.00	0.035	-0.1608	Doesn't exist	(4)
	0. 56	0.00019	-0.00011		
	0.46	1.38^{E}-06	1.1^{E}-06		
Model accepted series non-statioraire	0.00	-4.5255	-0.157	Doesn't exist	(5)
	0.9	0.11452	1.08^{E}-05		
Model accepted series statioraire	0.00	-4.538	-0.15722	Doesn't exist	(6)

IV -4- KPSS test (Kwiatkowski, Phillips, Schmidt and Shin, 1992):

The KPSS test, developed by Kwiatkowski, Phillips, Schmidt, and Shin in 1992, is a statistical procedure designed to test the stationarity of a time series. Unlike traditional unit root tests such as the Dickey–Fuller and Phillips–Perron tests, which adopt non-stationarity as the null hypothesis, the KPSS test assumes stationarity under the null. This fundamental difference allows the KPSS test to serve as a complementary tool, particularly useful when combined with unit root tests, in order to avoid ambiguous conclusions regarding the stochastic properties of time series data.

In empirical time series analysis, relying on a single test may lead to misleading inferences, especially in small samples or in the presence of structural features such as persistence or autocorrelation. For this reason, the KPSS test is frequently employed alongside ADF and PP tests to strengthen the robustness of stationarity diagnostics[1].

The KPSS test is formulated under the null hypothesis that the series is stationary and is evaluated using only two model specifications.

- ✓ **Null Hypothesis and Models:**

The KPSS test is based on two primary models[2]:

- **Model 2**: with a constant.

- **Model 3**: with a constant and a linear trend.

The computation of the KPSS test statistic relies on the long-run variance of the residuals, denoted by s_t^2, which is closely related to the variance correction used in the Phillips–Perron framework. The test is based on the following decomposition:

$$Y_t = \mu + \theta S_t + e_t$$

$$S_t = S_t - 1 + u_t$$

In this representation, S_t captures the stochastic trend component of the series, while e_t represents a stationary error term. The interpretation of the variance σ_u^2 is central to identifying the nature of the process:

[1] Maddala. G. S, Kim. I. M, op.cit, p 122.
[2] Bruce E. Hansen, **Econometrics**, Princeton University Press, New Jersey, 2022, p 582.

- If $\sigma_u^2 = 0$, the component S_t is constant over time, and the series Y_t is stationary.

- If $\sigma_u^2 > 0$, the stochastic component evolves over time, implying that Y_t follows a unit root process and is therefore non-stationary.

Accordingly, the hypotheses of the KPSS test are formulated as:

Null Hypothesis : $H_0 : \sigma_u^2 = 0$

Alternative Hypothesis : $H_1 : \sigma_u^2 > 0$

It is assumed that the error terms (e_t, u_t) are independent and normally distributed, which ensures the validity of the asymptotic distribution of the test statistic[1].

- ✓ **Test Statistic:**

The KPSS test is derived from the Lagrange Multiplier (LM) principle and rejects the null hypothesis of stationarity for sufficiently large values of the following statistic:

$$\frac{1}{n^2\widehat{\sigma^2}}\sum_{i=1}^{n}\left(\sum_{t=1}^{i}\widehat{e_t}\right)^2$$

Where:

$\widehat{e_t}$: are the residuals estimated under the null hypothesis of stationarity;

$\widehat{\sigma^2}$: Is the sample variance of these residuals;

[1] Kwiatkowski. D & others , **Testing the null hypothesis of stationarity against the alternative of a unit root**, Journal of Econometrics, 54, 1992, p 162.

n: denotes the sample size.

This formulation reflects the cumulative behavior of the residuals; under stationarity, their partial sums remain bounded, whereas under a unit root they diverge, leading to large values of the statistic[1].

- ✓ **Adjustment for Serial Correlation:**

When the residuals e_t exhibit serial correlation, the KPSS statistic is modified to account for this dependence structure:

$$\mathrm{KPSS}_1 = \frac{1}{n^2\widehat{\omega^2}}\sum_{i=1}^{n}(\sum_{t=1}^{i}\widehat{e_t})^2$$

The long-run variance $\widehat{\omega}^2$ is estimated using the Newey–West estimator[2] :

$$\widehat{\omega^2} = \sum_{i=1}^{M}(1-\frac{|\ell|}{M+1})\frac{1}{n}\sum_{t=1}^{n}\widehat{e_t}\widehat{e_{t-\ell}})^2$$

This correction ensures consistency of the variance estimator in the presence of autocorrelation and heteroskedasticity, thereby improving the reliability of the KPSS test in practical applications[3].

- ✓ **KPSS Test with Trend:**

When a deterministic linear trend is allowed, the local level model

[1] Hamilton, op.cit, p 506.
[2] Newey. W. K, West. K. D, **A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix**, Econometrica, 55, 1987, p p 703–708.
[3] Maddala. G. S, Kim. I. M, op.cit, p 123.

takes the following form:

$$Y_t = \mu + \beta t + \theta S_t + e_t$$

Under the null hypothesis, the least squares estimator is given by:

$$Y_t = \tilde{\mu} + \tilde{\beta} t + \tilde{e}_t$$

where $\tilde{e}_t$ represents the detrended residuals of the series. The corresponding KPSS test statistic is defined as:

$$\text{KPSS}_2 = \frac{1}{n^2 \widehat{\omega^2}} \sum_{i=1}^{n} \left(\sum_{t=1}^{i} \widehat{e_t}\right)^2$$

As in the previous case, the null hypothesis of stationarity is rejected for large values of this statistic.

- ✓ **Critical Values:**

	1%	5%	10%
Model 2	0.739	0.463	0.347
Model 3	0.216	0.146	0.119

These critical values are non-standard and are obtained from asymptotic simulations, which is why the KPSS statistic cannot be compared to conventional t or χ^2 distributions[1].

The KPSS test constitutes a fundamental component of modern time series analysis by reversing the traditional testing framework and adopting stationarity as the null hypothesis. Through its two

[1] Hamilton, op.cit, p 507.

specifications—one with a constant and another with a constant and linear trend—the test provides a rigorous assessment of the long-run behavior of economic and financial time series. The use of the LM statistic and the Newey–West correction for long-run variance enhances its robustness in the presence of serial correlation. In applied work, the KPSS test is most effective when combined with ADF and PP tests, allowing researchers to draw more reliable conclusions regarding the presence of unit roots and the appropriate modeling strategy for time series data[1].

Example:

Null Hypothesis: IPC_GLOBAL is stationary
Exogenous: Constant, Linear Trend
Bandwidth: 11 (Newey-West automatic) using Bartlett kernel

		LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic		0.087530
Asymptotic critical values*:	1% level	0.216000
	5% level	0.146000
	10% level	0.119000

*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction)	0.000420
HAC corrected variance (Bartlett kernel)	0.004186

KPSS Test Equation
Dependent Variable: IPC_GLOBAL
Method: Least Squares

[1] Maddala. G. S, Kim. I. M, op.cit, p p 124-125.

Null Hypothesis: D(IPC_GLOBAL,2) is stationary
Exogenous: Constant
Bandwidth: 20 (Newey-West automatic) using Bartlett kernel

		LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic		0.045743
Asymptotic critical values*:	1% level	0.739000
	5% level	0.463000
	10% level	0.347000

*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction)	2.26E-06
HAC corrected variance (Bartlett kernel)	9.79E-07

KPSS Test Equation
Dependent Variable: D(IPC_GLOBAL,2)
Method: Least Squares

- ✓ **Procedure for testing for unit-root tests :**

Unit root tests constitute a fundamental set of statistical tools in time series analysis, aimed at determining whether a given series is stationary or exhibits non-stationary behavior due to the presence of a unit root. A unit root implies a random walk–type persistence, whereby shocks have permanent effects on the series, potentially leading to spurious regressions and unreliable econometric inference.

These tests were developed to address critical issues in econometric modeling, particularly those related to persistence, stochastic trends, and non-constant variance in time series data. The most commonly applied unit root tests include the Dickey–Fuller (DF) test, the Augmented Dickey–Fuller (ADF) test, the Phillips–Perron (PP) test, and the KPSS test, each of which relies on a different statistical framework and hypothesis structure.

In general, unit root testing procedures evaluate the null hypothesis that a time series contains a unit root against the alternative hypothesis of stationarity. The testing process typically involves estimating a regression of the first-differenced series on its lagged level, possibly augmented with lagged differences to correct for autocorrelation in the error term. The statistical significance of the estimated coefficients is then assessed using calculated test statistics, which are compared with non-standard critical values.

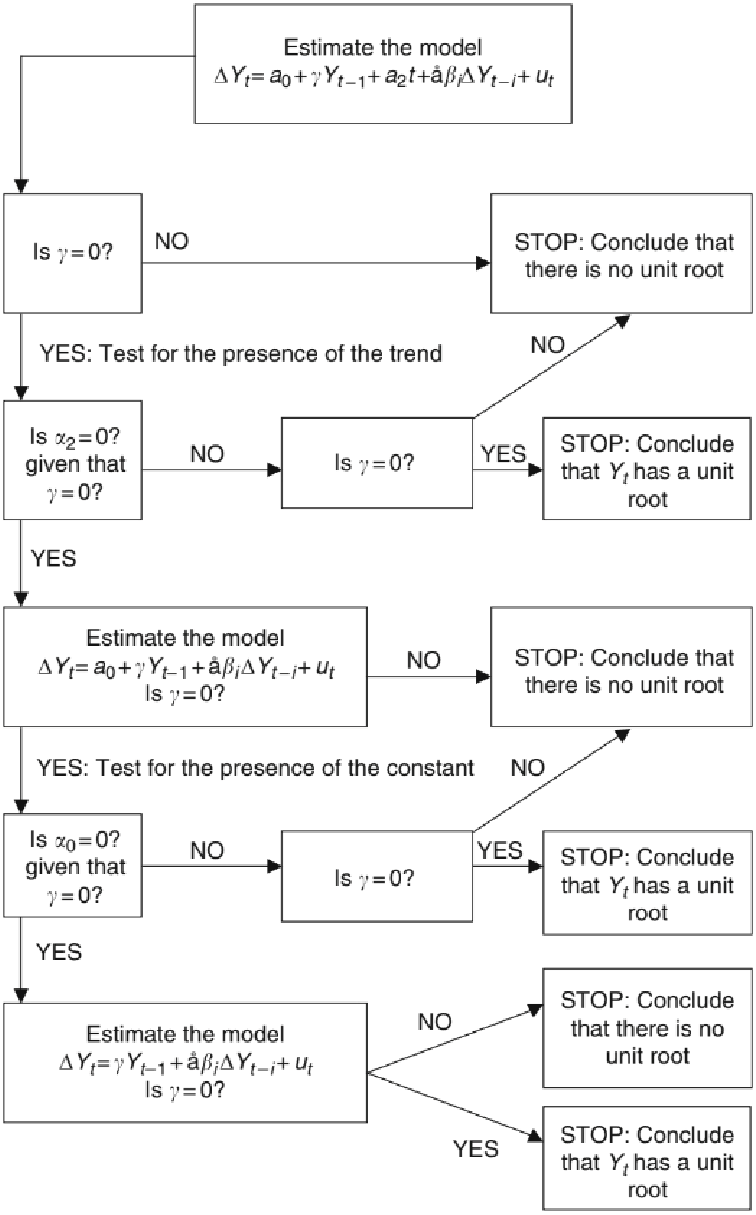
An important aspect of the testing procedure is the consideration of deterministic components in the model, such as the inclusion of a constant term or a linear trend. Depending on the empirical characteristics of the series, different model specifications are examined to ensure that the test results are not biased by omitted deterministic elements. The decision process often follows a structured sequence, where the presence or absence of a trend or intercept is evaluated before drawing conclusions regarding stationarity.

The flowchart diagram presented above illustrates a generalized step-by-step procedure for conducting unit root tests. It outlines the sequential testing strategy and the decision rules based on the outcomes of each test, allowing the researcher to determine whether the series is stationary in levels, stationary after differencing, or requires further transformation.

Unit root tests are widely applied in empirical research across various disciplines, particularly in macroeconomics and finance, where time series data are pervasive. Ensuring stationarity is a prerequisite for the correct specification of econometric models such as ARIMA and VAR models, and for avoiding misleading results caused by spurious correlations. Consequently, unit root testing represents an essential preliminary step in any rigorous time series analysis[1].

[1] Dimitrios Asteriou, Stephen G. Hall, **Applied Econometrics**, Macmillan international higher education, 4th edition, 2021, united kingdom, p 372.

Figure (III-1): Testing Strategy for Unit Roots



Source : Dimitrios Asteriou, Stephen G. Hall, **Applied Econometrics**, Macmillan international higher education, 4th edition, 2021, united kingdom, p 373.

The flowchart presented in Figure (III-1) provides a synthetic and operational representation of the sequential decision procedure commonly adopted in unit root testing, particularly within the Dickey–Fuller and Augmented Dickey–Fuller frameworks. Its inclusion within this chapter aims to clarify the logical structure underlying the choice of the appropriate test specification and the role played by deterministic components such as the constant and the time trend.

In practice, unit root testing does not rely on a single regression equation, but rather on a sequence of nested models whose specification depends on the statistical significance of deterministic terms. As emphasized by Dickey and Fuller (1979) and later formalized in standard textbooks, an incorrect inclusion or exclusion of these components may lead to misleading inferences regarding the stationarity properties of the series.

The procedure begins with the estimation of the most general model, which includes a constant, a deterministic linear trend, and lagged differences in order to account for possible serial correlation in the disturbances. The central parameter of interest is the coefficient associated with the lagged level of the series, whose statistical significance determines whether the null hypothesis of a unit root can be rejected.

If the unit root hypothesis is rejected at this stage, the process terminates and the series is considered stationary. Otherwise, the testing strategy proceeds by examining the relevance of the deterministic trend. When the trend term is statistically insignificant, it is excluded, and the unit root test is repeated within a reduced model containing only a constant. A similar logic applies to the constant term itself, which is tested conditionally before estimating the most restricted specification without any deterministic components.

This sequential testing strategy ensures internal consistency and avoids over-parameterization. By progressively simplifying the model only when the imposed restrictions are statistically justified, the analyst can clearly distinguish between three fundamentally different data-generating processes:

1- trend-stationary processes;
2- difference-stationary processes with drift;
3- pure random walk processes without drift.

Therefore, the flowchart serves not merely as a graphical aid, but as a methodological guide that translates the theoretical principles of unit root testing into a clear empirical procedure. Its integration within this chapter complements the formal presentation of the ADF and PP tests and provides students with a practical roadmap for applied time series analysis.

Chapter Four: Box–Jenkins Methodology and Short-Term Forecasting

Forecasting constitutes a central objective of time series analysis, particularly in applied economic and financial studies where decision-making often relies on accurate short-term predictions. In this context, time series models offer a powerful analytical framework by exploiting the internal temporal structure of data, without the need to introduce external explanatory variables. Such an approach is especially suitable when the underlying economic mechanisms are complex, partially unobservable, or subject to rapid change.

Within the broad spectrum of time series forecasting techniques, the Box–Jenkins methodology represents one of the most rigorous and widely adopted approaches. Developed in the seminal work of Box and Jenkins, this methodology is grounded in stochastic process theory and provides a systematic procedure for modeling, validating, and forecasting time series data. Its methodological strength lies in its emphasis on stationarity, parsimony, and empirical adequacy, making it particularly well suited for short-term forecasting applications.

The Box–Jenkins approach is inherently iterative and rests on a clear sequence of analytical stages. These stages require the analyst to move progressively from preliminary data examination to model identification, parameter estimation, diagnostic checking, and finally forecasting. Each stage builds upon the results of the previous one, ensuring internal consistency and reducing the risk of model misspecification.

A fundamental prerequisite for the application of the Box–Jenkins methodology is the stationarity of the time series under consideration. As established in earlier chapters, non-stationary series can lead to spurious results and unreliable forecasts. Consequently, unit root tests such as the Dickey–Fuller, Augmented Dickey–Fuller, Phillips–Perron, and KPSS tests are systematically employed to determine whether the series is difference-stationary (DS) or trend-stationary (TS). The outcome of these tests directly determines the transformation required to render the series stationary and, hence, suitable for Box–Jenkins modeling[1].

Once stationarity is achieved, the Box–Jenkins methodology provides a coherent framework for capturing the dynamic behavior of the series through autoregressive and moving average components. By carefully analyzing autocorrelation patterns and validating model assumptions, this approach enables the construction of statistically sound models capable of generating reliable short-term forecasts. As such, the Box–Jenkins methodology constitutes a natural continuation of the analytical tools developed in previous sections of this course and serves as a cornerstone for applied forecasting in economics and related fields[2].

The Box-Jenkins methodology, developed by statisticians George Box and Gwilym M. Jenkins, is a systematic approach for identifying, estimating, and diagnosing time series models, specifically those that fall within the Autoregressive Integrated Moving Average (ARIMA) framework[3].

I- General Framework of the Box–Jenkins Methodology:

The Box–Jenkins methodology provides a rigorous and systematic framework for the analysis and short-term forecasting of time series data. Its fundamental objective is to identify an appropriate stochastic model capable of capturing the underlying temporal structure of a series and generating reliable forecasts. Unlike explanatory econometric models that incorporate exogenous variables, the Box–Jenkins approach is purely univariate, relying exclusively on the historical values of the series itself [4].

---

[1] Enders. W, op.cit, p p 48-55.
[2] Hamilton, op.cit, p p 63–66.
[3] Samir Mustafa Al-Shaarawi, **Introduction to Time Series Analysis**, King Abdulaziz University Press, First Edition, Saudi Arabia, 2005, p 261.
[4] Box. G. E, Jenkins. G. M**, Time series analysis: Forecasting and control**, San Francisco, CA: Holden-Day, 1976, p p 5-6.

This methodology is grounded in the theory of stochastic processes and assumes that the observed time series is generated by a stable probabilistic mechanism. As a result, a central prerequisite for applying the Box–Jenkins framework is the stationarity of the series. When the original data exhibit non-stationary behavior, suitable transformations—such as differencing or detrending—must be applied prior to model construction. This requirement establishes a strong conceptual link between the Box–Jenkins methodology and unit root testing procedures, which serve as an essential preliminary step in empirical applications[1].

The general framework of the Box–Jenkins methodology is structured around an iterative process composed of four interdependent stages: model identification, parameter estimation, diagnostic checking, and forecasting. These stages are not executed in a strictly linear fashion; rather, the analyst may return to earlier steps whenever diagnostic tests reveal inadequacies in the specified model. This iterative structure enhances the robustness of the modeling process and reduces the risk of misspecification[2].

A defining principle of the Box–Jenkins framework is parsimony. The methodology emphasizes selecting the simplest model that adequately represents the data, avoiding unnecessary complexity that may undermine forecasting performance. Models with an excessive number of parameters often suffer from overfitting, which limits their predictive accuracy outside the estimation sample. Consequently, the Box–Jenkins approach systematically evaluates alternative specifications and retains only statistically significant components supported by the data[3].

Another key feature of the Box–Jenkins methodology is its reliance on residual diagnostics as a criterion for model adequacy. Once

---

[1] Enders. W, op.cit, p p 82-85.
[2] Hamilton, op. cit, p p 107–110.
[3] Gujarati. op.cit, p p 740-741.

a candidate model has been estimated, the residuals are examined to ensure that they behave as white noise, exhibiting zero mean, constant variance, and no serial correlation. The absence of systematic patterns in the residuals indicates that the model has successfully captured the relevant dynamics of the time series. If these conditions are not satisfied, the model must be revised and re-estimated[1].

The Box–Jenkins framework is particularly well suited for short-term forecasting, where temporal dependencies and autocorrelation structures play a dominant role. Its effectiveness has been demonstrated across a wide range of empirical applications, including macroeconomic indicators, financial time series, industrial production, and energy demand. By focusing on the internal structure of the data, the methodology provides a flexible and powerful tool for forecasting in environments characterized by relatively stable data-generating processes[2].

In summary, the general framework of the Box–Jenkins methodology offers a coherent and disciplined approach to time series modeling. By integrating theoretical principles with empirical diagnostic tools, it enables the analyst to construct statistically sound models, validate their assumptions, and produce reliable forecasts. This framework remains a cornerstone of modern time series analysis and continues to play a central role in applied economic and statistical research.

II- Stationarity and Preliminary Testing:

Before applying the Box–Jenkins methodology, it is essential to assess the stability of the time series under consideration. The Box–Jenkins framework is built on the assumption that the observed data are generated by a stable probabilistic mechanism over time. Consequently,

---

[1] Box. G. E. P, Jenkins. G. M, Reinsel. G. C, op.cit, p p 67-69.

[2] Pindyck. R. S, Rubinfeld. D. L, **Econometric models and economic forecasts** , 4th ed, Boston, MA: Irwin/McGraw-Hill, 1998, p p 505-508.

stationarity constitutes a fundamental prerequisite for model identification, estimation, and forecasting.

A time series is said to be stationary if its key statistical properties—namely the mean, variance, and autocovariance structure—remain constant over time. This implies the absence of systematic trends, time-varying volatility, or evolving seasonal patterns. Stationarity is crucial because ARIMA models rely on the persistence of temporal dependence structures; when these properties change over time, the estimated parameters lose their economic and statistical interpretability, leading to unreliable forecasts.

- ✓ **The necessity of stationarity in Box–Jenkins modeling:**

Non-stationary time series are known to generate spurious relationships and misleading statistical inferences. In such cases, high goodness-of-fit measures may arise even in the absence of any meaningful dynamic relationship. This issue is particularly relevant for macroeconomic and financial variables—such as gross domestic product, price indices, or consumption—which typically exhibit stochastic trends and structural changes. Applying the Box–Jenkins methodology directly to non-stationary data often results in unstable models with weak forecasting performance[1].

From a forecasting perspective, non-stationarity implies that shocks to the system may have permanent effects, causing the future path of the series to depend heavily on past disturbances. This behavior violates the core assumption underlying ARIMA processes, which require a stable dependence structure over time. Ensuring stationarity therefore enhances both the internal coherence of the model and the reliability of short-term forecasts.

- ✓ **Testing for stationarity and unit roots:**

[1] Enders. W, op.cit, p p 82-85.

To assess whether a time series is stationary, formal unit root tests are commonly employed. Among these, the Dickey–Fuller test and its extended version, the Augmented Dickey–Fuller (ADF) test, are widely used in empirical applications. These tests are designed to detect the presence of a unit root, which indicates non-stationarity. The null hypothesis states that the series contains a unit root, while rejection of the null supports stationarity.

In addition to the ADF test, alternative procedures such as the Phillips–Perron (PP) test are often applied to account for serial correlation and heteroskedasticity in the error terms. In the context of the Box–Jenkins methodology, these tests play a crucial diagnostic role by determining the order of integration of the series prior to model identification[1].

- ✓ **Transformation of non-stationary series:**

When a time series is found to be non-stationary, appropriate transformations must be applied before proceeding with the Box–Jenkins modeling steps. The choice of transformation depends on the nature of the non-stationarity.

If the series is difference-stationary (DS), stationarity can be achieved by differencing the data. First-order differencing removes stochastic trends by computing the change between consecutive observations:

$$\Delta X_t = X_t - X_{t-1}$$

If the series is trend-stationary (TS), non-stationarity arises from a deterministic trend rather than a stochastic one. In this case, detrending is more appropriate and typically involves regressing the series on a time trend and using the residuals for subsequent analysis.

[1] Enders. W, op.cit, p p 82-85.

The distinction between differencing and detrending is methodologically important. Differencing implies that shocks have permanent effects on the level of the series, whereas detrending assumes that shocks are temporary and that the series eventually reverts to its deterministic trend. This choice directly influences both the interpretation of the ARIMA model and the nature of the resulting forecasts[1].

- ✓ **Stationarity as a preliminary step in the Box–Jenkins procedure:**

The requirement of stationarity establishes a strong conceptual and operational link between the Box–Jenkins methodology and unit root testing procedures. In practice, unit root tests are systematically conducted as a preliminary step before model identification, ensuring that the autoregressive and moving average components are specified on a stable series. Only after stationarity has been achieved can the analyst proceed confidently to the identification of ARIMA orders and the subsequent stages of estimation and diagnostic checking.

III- Stages of the Box–Jenkins Methodology and Model Identification:

Applying the Box–Jenkins methodology requires a structured and sequential procedure composed of four main stages: model identification, parameter estimation, diagnostic checking, and short-term forecasting. However, before initiating these stages, it is essential to ensure that the time series under study satisfies the stationarity condition, which constitutes a fundamental assumption of the Box–Jenkins approach.

The methodology is based on the assumption that the observed series is generated by a stable stochastic process. Stationarity implies that the statistical properties of the series—namely its mean, variance,

[1] Hamilton, op.cit, p p 499–501.

and covariance—remain constant over time. If this condition is violated, the estimated parameters may become unstable, leading to misleading inferences and unreliable forecasts.

To assess stationarity, unit root tests are commonly employed, with the Dickey–Fuller test and its augmented version (ADF) being the most widely used. The null hypothesis of these tests assumes the presence of a unit root, indicating non-stationarity. When non-stationarity is detected, the series must be transformed appropriately. If the series is difference-stationary (DS), differencing is applied to remove the stochastic trend. If it is trend-stationary (TS), detrending through regression on a deterministic time trend is more appropriate[1].

Once stationarity has been achieved, the first operational stage of the Box–Jenkins methodology—model identification—can be conducted. This stage is widely regarded as the most challenging and critical, as it determines the general structure of the model to be estimated. Model identification relies primarily on the examination of the autocorrelation function (ACF) and the partial autocorrelation function (PACF), which provide valuable information about the temporal dependence structure of the series.

- ✓ **Stationarity Conditions:**

For a time series to be considered stationary, several conditions must be satisfied:

Constant Mean: The series must fluctuate around a fixed arithmetic mean. Mathematically:

$$E(Y_t) = E(Y_{t+k}) = \mu$$

[1] Enders. W, op.cit, p p 82-85.

Constant Variance: The variance of the series should remain constant over time:

$$VAR(Y_t) = VAR(Y_{t+k}) = \gamma_0$$

Constant Covariance: The covariance between observations at different times must also be constant, which can be represented as:

$$COV(Y_t, Y_{t+k}) = COV(Y_{t+k}, Y_{t+k+s})$$

These criteria ensure that the statistical properties of the time series do not change over time, which is fundamental for the validity of the Box-Jenkins methodology.

The ACF measures the correlation between the series and its lagged values, while the PACF isolates the direct correlation at a given lag after eliminating the effect of intermediate lags. The theoretical patterns of these functions allow the analyst to distinguish between autoregressive (AR), moving average (MA), and mixed ARMA processes[1].

In addition to identifying the appropriate lag structure, this stage involves examining whether the series exhibits randomness, deterministic trends, or seasonal behavior. The presence of seasonality may require extending the basic ARIMA framework to seasonal ARIMA (SARIMA) models, which explicitly account for periodic patterns in the data[2].

- ✓ **Model Classes:**

[1] Gujarati. D, Porter. C, op.cit, p 738-740.
[2] Box. G. E. P, Jenkins. G. M, op.cit, p p 84-90.

Within this framework, several classes of models are defined. A **moving average model MA(q)** expresses the current value of the series as a linear combination of current and past random shocks:

$$MA(q): y_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}$$

Using the lag operator L:

$$Yt = \left(1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q\right)\varepsilon_t$$

$$= \Theta(L)\varepsilon_t$$

Where:

ε_t: represents white noise and εt $\sim BB(0, \sigma_\varepsilon^2)$

$(i = 1,2\ldots, p)$: Model coefficients.

An **autoregressive model AR(p)** explains the current value of the series by its own lagged values. We say that the series follows an autoregressive process of order p when the current value is explained by its previous values through time lags. The model can be written as follows[1]:

$$AR(p): y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \varepsilon_t$$

By using the lag operator L (which has the property $\mathrm{L}^n\mathrm{Y_t} = \mathrm{Y_{t-n}}$) we can write theAR(p) model as:

$$Y_t(1 - \varphi 1L - \varphi 2L^2 - \cdots - \phi pL^p) = \varepsilon_t$$

Where:

[1] Dimitrios Asteriou, Stephen G. Hall, op.cit , p 270.

ε_t: represents white noise and $\varepsilon_t \sim BB(0, \sigma_\varepsilon^2)$

φi $(i = 1,2\ldots, p)$: Model coefficients

(L) Y_t is a polynomial function of Y_t.

A mixed ARMA (p, q) model combines both mechanisms of autoregression and moving average. We say that the series follows an ARMA process when it is explained by both its past values and past error terms. In other words, the current value of the series is modeled as a function of its own lagged values and a weighted average of past random shocks.

The linear mathematical formulation of the ARMA (p, q) model is written as follows:

$$ARMA(p,q): y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

Y_t : is the observed value at time t ;

φ_i $(i = 1, \ldots, p)$: are the autoregressive coefficients;

θ_j $(j = 1, \ldots, q)$: are the moving average coefficients

ε_t : is white noise with $\varepsilon_t \sim BB(0, \sigma_\varepsilon^2)$

If differencing is required to achieve stationarity, the resulting model is an **ARIMA(p,d,q)** model, where d denotes the order of integration. This transformation removes non-stationarity through differencing, usually applied once or twice depending on the presence of trend[1].

- ✓ **Identification Using ACF and PACF:**

[1] Enders.w, 2015, pp 82–85.

The identification of these models is guided by the theoretical behavior of the ACF and PACF, which provide practical rules for distinguishing between AR, MA, and ARMA structures. This procedure is summarized in Table (III-1), which presents the standard correspondence between the patterns of the ACF and PACF and the main classes of ARIMA models.

Table (IV-1): Identification of AR, MA, and ARMA Models Using ACF and PACF

Partial Auto Correlation Functi (PACF)	Autocorrelation function (ACF)	The model
The value of the partial autocorrelation function for the first lag is null, while for the rest of the lags it is non-null.	The function should be - .exponential or sinusoidal	AR (1)
- The first limit of the function is statistically significant. - The other limits greater than p are negligible.	- The function is exponential or sinusoidal depending on the sign of the parameters.	AR (P)
The function should be exponential or sinusoidal	The value of the partial autocorrelation function for the first lag is null, while for the rest of the lags it is non-null.	MA (1)
The function is exponential or sinusoidal depending on the sign of the parameters.	- The first limit of the function is statistically significant. - The other limits greater than q are negligible.	MA (q)
The value of the function for the first lag is null and it decreases exponentially after the first lag.	The value of the function for the first lag is null and it decreases exponentially after the first lag.	ARMA(1,1)
- The function shape is exponentially decreasing or	- The function shape is exponentially decreasing or	ARMA(p,q)

truncated sinusoidal (fading) after the q-p slowdown.	truncated sinusoidal (fading) after the q-p slowdown.	

IV- Parameter Estimation, Diagnostic Checking, and Short-Term Forecasting:

After completing the stages of model identification and preliminary analysis, the Box–Jenkins methodology advances to its most operational and outcome-oriented phase. While the previous stages focus on understanding the structural properties of the time series and selecting an appropriate model form, the present stage is concerned with transforming this theoretical specification into a statistically valid and empirically reliable forecasting tool.

This phase constitutes the core of the Box–Jenkins approach, as it directly links model construction to practical forecasting performance. Once the orders of the ARIMA model (p, d, q), have been determined, attention shifts toward estimating the model parameters, evaluating the adequacy of the estimated specification, and ultimately using the validated model to generate short-term forecasts. Each of these steps plays a critical role in ensuring that the model not only fits the historical data but also captures the underlying stochastic dynamics in a stable and consistent manner.

Parameter estimation is a particularly sensitive stage, as inaccurate or inefficient estimates may distort the dynamic properties of the model, even if the identification stage has been correctly conducted. For this reason, the Box–Jenkins methodology emphasizes the use of statistically sound estimation techniques and the careful evaluation of parameter significance. However, estimation alone is not sufficient. A model that appears satisfactory in terms of parameter estimates may still be inadequate if it fails to account for all systematic patterns in the data.

Consequently, diagnostic checking and model validation represent an indispensable component of this stage. Through residual analysis and stability testing, the analyst assesses whether the estimated model fulfills the fundamental assumptions of the Box–Jenkins framework, particularly the requirement that the residuals behave as white noise and that the parameters remain stable over time. If these conditions are not met, the methodology explicitly calls for a return to earlier stages, highlighting its iterative and self-correcting nature.

Once the model has successfully passed the diagnostic and validation procedures, it can be employed for forecasting purposes. In the context of short-term forecasting, this final step is of particular importance, as it translates the statistical structure of the model into quantitative predictions of future values. The reliability of these forecasts depends critically on the rigor with which the preceding estimation and validation stages have been conducted.

Accordingly, this part of the analysis is devoted to a detailed examination of parameter estimation methods, model selection criteria, diagnostic checking procedures, and short-term forecasting techniques within the Box–Jenkins framework. Together, these elements complete the methodological cycle and provide the foundation for producing robust and credible forecasts based on time series data.

- ✓ **Parameter Estimation:**

Once the appropriate ARIMA (p, d, q) or SARIMA $(\mathrm{p, d, q})(\mathrm{P, D, Q})s$ structure has been identified, the next stage in the Box–Jenkins methodology consists of estimating the model parameters. Parameter estimation aims to obtain numerical values for the autoregressive (AR) and moving average (MA) coefficients that best describe the stochastic process generating the observed time series. The accuracy of this stage is crucial, as it directly affects the reliability of diagnostic checking and the quality of subsequent forecasts.

In general, parameter estimation in Box–Jenkins models is conducted under the assumption that the error term follows a white noise process with zero mean and constant variance. Several estimation methods have been proposed in the literature; however, **Maximum Likelihood Estimation (MLE)** and **conditional or unconditional least squares** remain the most widely used approaches[1].

Maximum Likelihood Estimation is particularly attractive due to its desirable statistical properties. Under regularity conditions, MLE estimators are consistent, asymptotically efficient, and normally distributed. The likelihood function is constructed based on the joint probability distribution of the observed series, conditional on the model parameters. In the case of ARIMA models, the likelihood function is nonlinear and generally requires numerical optimization techniques, such as the Newton–Raphson or BFGS algorithms, to obtain parameter estimates[2].

An alternative approach consists of **conditional least squares estimation**, where the sum of squared residuals is minimized given initial observations. Although this method is computationally simpler, especially for pure autoregressive models, it may lead to biased estimates in small samples when moving average terms are present. For this reason, many econometric software packages initially use conditional least squares to obtain starting values, which are then refined through maximum likelihood procedures[3].

After estimation, the statistical significance of each parameter must be assessed. This is commonly achieved using t-statistics derived from the estimated standard errors. Insignificant coefficients may indicate model over-parameterization and often suggest the need for model simplification. A parsimonious model, characterized by a limited

[1] Box. G. E. P, Jenkins. G. M, op.cit, p p 161-165.
[2] Hamilton, op.cit, pp. 132–136
[3] Enders. W, op.cit, p p 94-97.

number of statistically meaningful parameters, is generally preferred in the Box–Jenkins framework, as it balances goodness of fit with interpretability and forecasting accuracy[1].

In the presence of seasonal components, parameter estimation becomes more complex due to the increased number of parameters and potential multicollinearity between seasonal and non-seasonal terms. Nevertheless, the underlying estimation principles remain the same, and maximum likelihood methods continue to be the standard approach for seasonal ARIMA models[2].

In summary, parameter estimation constitutes a central step in the Box–Jenkins methodology, bridging the gap between theoretical model identification and empirical validation. Careful estimation, combined with rigorous statistical testing and adherence to theoretical constraints, ensures that the resulting model provides a sound basis for diagnostic checking and short-term forecasting.

✓ **Model Selection Criteria:**

Following parameter estimation, the Box–Jenkins methodology requires a systematic evaluation of competing models in order to select the most appropriate specification. Model selection is a critical step, particularly when several ARIMA or SARIMA models exhibit acceptable statistical properties. The objective at this stage is not to identify the most complex model, but rather the one that achieves an optimal balance between goodness of fit, parsimony, and predictive performance.

One of the fundamental principles guiding model selection in the Box–Jenkins framework is **parsimony**. A parsimonious model is one that explains the data adequately using the smallest possible number of parameters. Over-parameterized models may fit the sample data well

[1] Box. G. E. P, Jenkins. G. M, Reinsel. G. C, op.cit, p p 71-73.
[2] Box. G. E. P, Jenkins. G. M, op.cit, p p 186-190.

but often perform poorly in forecasting due to overfitting and increased parameter uncertainty[1].

To operationalize this principle, several **information criteria** have been developed. Among the most widely used are the **Akaike Information Criterion (AIC)**, the **Bayesian Information Criterion (BIC)**—also known as the Schwarz Information Criterion—and the **Hannan–Quinn Criterion (HQ)**. These criteria penalize model complexity by incorporating both the goodness of fit and the number of estimated parameters into a single quantitative measure.

The Akaike Information Criterion is defined as:

$$AIC = n\ ln\left(\frac{SCR}{n}\right) + 2(p+q)$$

Where:

SCR : denotes the sum of squared residuals;

n : is the sample size;

$(p+q)$: represents the number of autoregressive and moving average parameters included in the model.

The AIC favors models with a good fit while imposing a relatively moderate penalty on additional parameters, which may lead to the selection of more complex models, especially in small samples[2].

In contrast, the Bayesian Information Criterion (BIC), also known as Schwarz's Information Criterion, imposes a stronger penalty on

[1] Box. G. E. P, Jenkins. G. M, op.cit, p p 241-243.
[2] Enders. W, op.cit, p p 103-105.

model complexity, particularly as the sample size increases. It is given by:

$$\text{BIC} = n\, ln\left(\frac{SCR}{n}\right) + (p+q)Ln\,(n)$$

where n denotes the number of observations. As a result, the BIC tends to select more parsimonious models and is often preferred in empirical applications where the primary objective is consistent model selection rather than purely predictive accuracy[1].

The Hannan–Quinn Criterion occupies an intermediate position between AIC and BIC, offering a compromise between goodness of fit and parsimony. It is particularly useful in situations where the sample size is neither very small nor very large. It is defined as:

$$HQ = n\, ln\left(\frac{SCR}{n}\right) + 2(p+q)\, Ln(Ln(n))$$

In addition to information criteria, **likelihood-based comparisons** may be employed when models are nested. The likelihood ratio (LR) test allows the researcher to assess whether the inclusion of additional parameters significantly improves the model fit. However, this approach is not applicable when models are non-nested, which is often the case in ARIMA modeling[2].

Beyond statistical measures, model selection must also consider **economic and empirical plausibility**. Estimated coefficients should exhibit signs and magnitudes consistent with theoretical expectations, and the model should provide a coherent interpretation of the underlying data-generating process. A model that performs well

[1] Hamilton, op.cit, p p 142–143.

[2] Box. G. E. P, Jenkins. G. M, Reinsel. G. C, op.cit, p p 74-76.

statistically but lacks interpretability may be of limited practical value, especially in applied economic forecasting[1].

Finally, the ultimate validation of a selected model lies in its **forecasting performance**. In practice, researchers often compare competing models based on out-of-sample forecasts using accuracy measures such as the Mean Absolute Error (MAE) or the Root Mean Squared Error (RMSE). Although these measures fall outside the formal Box–Jenkins identification cycle, they provide valuable complementary evidence in selecting the most appropriate model for short-term forecasting[2].

- ✓ **Diagnostic Checking and Model Validation:**

After the estimation stage, the Box–Jenkins methodology requires a thorough diagnostic checking and model validation process. This step is essential because it ensures that the selected model adequately captures the dynamics of the time series and that the residuals satisfy the assumptions required for reliable inference and forecasting. Diagnostic checking is therefore not merely a formal procedure but a substantive test of model adequacy, and it may lead to model revision and re-estimation when violations are detected[3].

1- Residual Analysis and White Noise Testing:

The primary objective of diagnostic checking is to verify that the residuals of the estimated model behave like a white noise process. Residuals are considered white noise when they have a zero mean, constant variance, and no autocorrelation. In other words, after fitting the ARIMA/SARIMA model, there should be no remaining systematic pattern in the series.

To test this property, the **Ljung–Box Q test** is widely used. The null hypothesis of the test states that the residuals are independently distributed (no autocorrelation up to a specified lag). If the null

[1] Kirchgässner. G, wolters. J, op.cit, p p 33-35.
[2] Enders. W, op.cit, p p 110-112.
[3] Box. G. E. P, Jenkins. G. M, Reinsel. G. C, op.cit, p p 76-80.

hypothesis is rejected, it indicates that the model has not fully captured the autocorrelation structure of the series, and a different model specification should be considered[1].

2- Checking Normality of Residuals:

Although ARIMA models do not strictly require normality for consistency of parameter estimates, normality is important for constructing valid confidence intervals and hypothesis tests. The **Jarque–Bera test** is commonly used to assess whether the residuals follow a normal distribution by examining skewness and kurtosis. Deviations from normality may suggest model misspecification or the presence of outliers, and may prompt the use of robust estimation methods or alternative model forms[2].

3- Homoscedasticity and Stability of Variance:

Another crucial assumption is the homogeneity of variance (homoscedasticity) of residuals. If residual variance changes over time (heteroscedasticity), it can lead to inefficient estimates and misleading inference. In such cases, the analyst may consider models that explicitly account for time-varying volatility, such as ARCH/GARCH models, or apply variance-stabilizing transformations (e.g., logarithmic transformation) before re-estimating the ARIMA model[3].

4- Parameter Stability and Structural Breaks:

Even when residuals appear white noise, the estimated parameters may not remain stable over the sample period. Parameter instability may result from structural breaks, policy changes, or regime shifts, which are common in economic and financial time series. To ensure that the model is valid for forecasting, it is essential to examine the stability of coefficients over time.

[1] Box. G. E. P, Jenkins. G. M, op.cit, p p 242-245.
[2] Enders. W, op.cit, p p 109-110.
[3] Hamilton, op.cit, p p 537–539.

Several tests are used for this purpose[1]:

- **CUSUM Test (Cumulative Sum)**: It detects systematic changes in the regression coefficients over time. If the CUSUM statistic crosses the critical bounds, it indicates parameter instability;

- **CUSUM of Squares (CUSUMSQ)**: This test evaluates the stability of the variance of residuals. It is particularly useful for detecting sudden structural changes;

- **Chow Test**: When a potential break date is known (e.g., a policy change or economic crisis), the Chow test assesses whether there is a significant structural change at that point[2].

5- Model Re-specification:

If diagnostic tests reveal significant problems (e.g., autocorrelation in residuals, non-normality, heteroscedasticity, or parameter instability), the model must be revised. Re-specification may involve[3]:

- Adding or removing AR or MA terms;

- Including seasonal components (for SARIMA);

- Applying additional differencing or transformation;

- Considering alternative models such as Autoregressive Integrated Moving Average with exogenous variables (ARIMAX) or GARCH-type models for volatility.

---

[1] Brown.R. L, Durbin. J, Evans. J. M, **Techniques for testing the constancy of regression relationships over time**, Journal of the Royal Statistical Society: Series B (Methodological), 37(2), 1975, p p 166–171.

[2] Chow. G. C, **Tests of Equality Between Sets of Coefficients in Two Linear Regressions**, Econometrica, 28(3), 1960, p p 391–392.

[3] Box. G. E. P, Jenkins. G. M, Reinsel. G. C, op.cit, p p 80-82.

The iterative nature of the Box–Jenkins methodology implies that the analyst may return to the identification stage and re-evaluate the model structure until diagnostic checks are satisfied.

6- Final Model Validation:

Once the residuals pass the diagnostic tests and parameters are stable, the model is considered validated. The final step is to confirm that the model provides meaningful and interpretable dynamics of the series and that it is suitable for forecasting. A validated model should exhibit:

- residuals that resemble white noise;
- stable parameters over time;
- satisfactory goodness-of-fit and information criteria;
- coherent theoretical interpretation.

Only after this validation can the model be used for short-term forecasting with confidence[1].

[1] Box. G. E. P, Jenkins. G. M, op.cit, p p 246-248.

References

References

1- Aston Zhang, Zachary C. Lipton, Mu Li, Alexander J. Smola, **Dive into Deep Learning**, Jupyter notebooks, 2020.

2- Bresson.G, pirotte. A, **économétrie des séries temporelles, théories et applications**, Presses Universitaires de France, paris, 1995.

3- Bartlett. M. S, **On The Theoretical Specification of Sampling Properties of Autocorrelated Time Series**, Journal of the Royal Statistical society, series B, vol.27, 1946.

4- Brockwell. P. J, Davis. R. A, **Time Series: Theory and Methods**, 2nd ed, New York: Springer-Verlag, 1991, p 15.

5- Brockwell. P. J, Davis. R. A, **Introduction to Time Series and Forecasting** , 3rd ed, Springer, 2016.

6- Box G. E, Pierce. D. A, **Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models**, Journal of the American Statistical Association, vol. 65, 1970.

7- Box. G. E. P, Jenkins. G. M, Reinsel. G. C, **Time Series Analysis: Forecasting and Control,** 4th ed, Hoboken: John Wiley & Sons, 2008.

8- Box. G. E, Jenkins. G. M**, Time series analysis: Forecasting and control**, San Francisco, CA: Holden-Day, 1976.

9- Brown.R. L, Durbin. J, Evans. J. M, **Techniques for testing the constancy of regression relationships over time**, Journal of the Royal Statistical Society: Series B (Methodological), 37(2)**,** 1975.

10- Beran.J, **Statistics for Long-Memory Processes**, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, V 61, CRC Press, 1994.

11- Bishop. C. M, **Pattern Recognition and Machine Learning**, Springer, New York, 2006.

12- Bruce E. Hasen, **ECONOMETRICS**, Princeton University Press, New Jersey, 2022.

References

13- Baillie. R.T, **Long memory processes and fractional integration in econometrics**, Journal of Econometrics**,** 73(1), 1996.

14- Christiaan. H & others, **Introduction to Time Series Econometrics**, Oxford University Press, 2004.

15- Chow. G. C, **Tests of Equality Between Sets of Coefficients in Two Linear Regressions**, Econometrica**,** 28(3), 1960.

16- Dickey. D. A, Fuller, W. A, **Distribution of the Estimators for Autoregressive Time Series with a Unit Root**, Journal of the American Statistical Association, 74(366), United States, 1979.

17- Diebold. F. X, **Forecasting in Economics, Business, Finance and Beyond**, Princeton University Press, 2017.

18- Dimitrios Asteriou, Stephen G. Hall, **Applied Econometrics**, Macmillan international higher education, 4th edition, 2021, united kingdom.

19- Diebold. F. X, **Forecasting in Economics, Business, Finance and Beyond**, Princeton University Press, 2017.

20- Elagueb Mohammed, **time series analysis Lectures and Applications in Economics**, economic department, university ziane Achour, elDjelfa.

21- Enders. W, **Applied Econometric Time Series**, 4th ed, Wiley, Hoboken, United States, 2015.

22- Franses. P. H, Haldrup. N, **Nonlinear Time Series Models in Empirical Finance**, Cambridge University Press, 2019.

23- Granger. C. W, Joyeux. R, **An introduction to long-memory time series models and fractional differencing**, Journal of Time Series Analysis**,** 1(1), 1980.

24- Gujarati .D.N, **Basic Econometrics** , 4th edition, Mc Graw-Hill / Irwin companies Inc New York, 2003.

References

25- Gujarati. D, Porter. C, **Basic Econometrics**, 5th ed, McGraw-Hill, New York, United States, 2009.

26- Goodfellow. I, Bengio. Y, Courville. A, ***Deep Learning*. MIT Press**, Cambridge, 2016.

27- Hamilton. J. D, **Time Series Analysis**, Princeton University Press, 1994.

28- Hamilton. J. D, **Time Series Analysis,** 2nd ed, Princeton University Press, 2020.

29- Hayashi. F, **Econometrics**, Princeton University Press, Prenceton, N.J. 2000.

30- Hosking. J. R. M, **Fractional differencing**, Biometrika, 68(1), 1981.

31- Hochreiter. S, Schmidhuber. J, **Long short-term memory,** Neural Computation, 9(8), 1997.

32- Haykin. S, **Neural Networks and Learning Machines**, 3rd ed, Pearson, New York, 2009.

33- John D. Levendis, **Time Series Econometrics Learning Through Replication**, Springer Texts in Business and Economics, New Orleans, LA, USA, 2018.

34- Jenkins. G.M, watts.D.G., Spectral **Analysis and Its Applications**, San Francisco, Holden-Day, 1968.

35- Kirchgässner. G, wolters. J, **introduction to modern time series analysis**, Springer-verlag Berlin Heidelberg, New York 2007.

36- Kwiatkowski. D & others , **Testing the null hypothesis of stationarity against the alternative of a unit root**, Journal of Econometrics, 54, 1992.

37- Ljung. G. M. , Box. G. P, **on the Measure of Lack of Fit in Time Series Models**, Biometrica, vol 66, 1978.

38- Maddala. G.S, **introduction to econometrics**, second edition, MacMillan publishing company, New York, 1992.

39- Maddala. G. S, Kim. I. M, **Unit Roots, Cointegration, and Structural Change**, Cambridge: Cambridge University Press, 1998.

40- Newey. W. K, West. K. D, **A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix**, Econometrica, 55, 1987.

41- Pindyck. R. S, Rubinfeld. D. L**, Econometric models and economic forecasts** , 4th ed, Boston, MA: Irwin/McGraw-Hill, 1998.

42- Phillips P. C, Perron .P, **Testing for a unit root in time series regression**, Biometrika, 75(2), 1988.

43- Régis bourbonnais, Michel Terraza, **Analyse des séries temporelles**, 3rd edition, Dunod, France , 2010.

44- Samir Mustafa Al-Shaarawi, **Introduction to Time Series Analysis**, King Abdulaziz University Press, First Edition, Saudi Arabia, 2005.

45- Tsay. R. S, **Multivariate Time Series Analysis: With R and Financial Applications** , 3rd ed, Wiley, 2018.

46- Zhang. A, Lipton. Z. C., Smola. A. J, **Dive into Deep Learning**, Release 0.14.X, Interactive deep learning textbook with code and mathematics, d2l.ai Foundation & Contributors, 2020.

Appendices

Augmented Dickey-Fuller Table

	Model 0 - no constant, no trend				Model 1 - constant, no trend				Model 2 - constant, trend			
N	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10
25	-2.661	-2.273	-1.955	-1.609	-3.724	-3.318	-2.986	-2.633	-4.375	-3.943	-3.589	-3.238
50	-2.612	-2.246	-1.947	-1.612	-3.568	-3.213	-2.921	-2.599	-4.152	-3.791	-3.495	-3.181
100	-2.588	-2.234	-1.944	-1.614	-3.498	-3.164	-2.891	-2.582	-4.052	-3.722	-3.452	-3.153
250	-2.575	-2.227	-1.942	-1.616	-3.457	-3.136	-2.873	-2.573	-3.995	-3.683	-3.427	-3.137
500	-2.570	-2.224	-1.942	-1.616	-3.443	-3.127	-2.867	-2.570	-3.977	-3.670	-3.419	-3.132
>500	-2.567	-2.223	-1.941	-1.616	-3.434	-3.120	-2.863	-2.568	-3.963	-3.660	-3.413	-3.128

TABLE 1

Fixed-b critical values of KPSS, Bartlett kernel, number of lags = bT

b	10% cv	5% cv	1% cv	Limit Power		b	10% cv	5% cv	1% cv	Limit Power
0	.347	.463	.739	1		.50	.374	.399	.435	.082
.01	.349	.457	.723	.995		.52	.379	.404	.443	.069
.02	.347	.454	.711	.969		.54	.384	.409	.452	.058
.04	.348	.449	.670	.886		.56	.388	.414	.460	.051
.06	.346	.441	.641	.794		.58	.392	.419	.467	.046
.08	.345	.435	.613	.732		.60	.397	.423	.475	.041
.10	.344	.427	.590	.686		.62	.400	.427	.482	.037
.12	.344	.420	.565	.647		.64	.404	,432	.487	.033
.14	.345	.414	.541	.611		.66	.408	.436	.493	.031
.16	.343	.408	.523	.581		.68	.412	.439	.496	.028
.18	.343	.404	.504	.550		.70	.416	.442	.499	.026
.20	.343	,400	.489	.523		.72	.420	.445	.499	.022
.22	.343	.395	.477	.499		.74	.425	.447	.501	.020
.24	.343	.392	.465	.472		.76	.428	.450	.501	.018
.26	.344	.389	.455	.445		.78	.432	.453	.500	.015
.28	.344	.385	.447	.420		.80	.437	.455	.498	.013
.30	.345	.383	.441	.395		.82	.441	.457	.496	.012
.32	.346	.382	.436	.367		.84	.445	.459	.494	.010
.34	.348	.381	.432	.335		.86	.450	.462	.492	.008
.36	.350	.381	.428	.302		.88	.456	.464	.490	.006
.38	.352	.381	.425	.268		.90	.461	.468	.488	.004
.40	.355	.382	.423	.230		.92	.468	.472	.486	.002
.42	.358	.384	.421	.194		.94	.474	.477	.486	.001
.44	.362	.387	.422	.158		.96	.482	.483	.488	.001
.46	.366	.390	.425	.127		.98	.491	.491	.492	.001
.48	.370	.395	.429	.101		1	.500	.500	.500	NA

Phillips–Perron (PP) Test: Critical Values

1. Without Constant and Trend

Significance Level	Critical Value
1%	-2.66
5%	-1.95
10%	-1.60

2. With Constant (Intercept)

Significance Level	Critical Value
1%	-3.43
5%	-2.86
10%	-2.57

3. With Constant and Trend

Significance Level	Critical Value
1%	-3.96
5%	-3.41
10%	-3.13