Time Series Analysis and Forecasting

Chapter 1: Stochastic Processes and Stationarity

Daniel Traian PELE

Bucharest University of Economic Studies
IDA Institute Digital Assets
Blockchain Research Center
AI4EFin Artificial Intelligence for Energy Finance
Romanian Academy, Institute for Economic Forecasting
MSCA Digital Finance

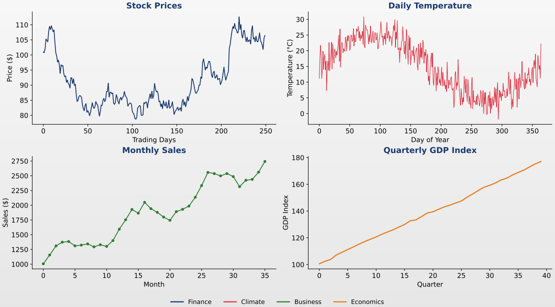
Learning Objectives

By the end of this chapter, you will be able to:

1. **Define** stochastic processes and understand their properties
2. **Distinguish** between strict and weak (covariance) stationarity
3. **Identify** white noise and random walk processes
4. **Compute** and interpret ACF and PACF
5. **Apply** the lag operator and differencing
6. **Conduct** stationarity tests (ADF, KPSS)
7. **Analyze** financial time series data
8. **Distinguish** between unit root and trend-stationary processes

Chapter Outline

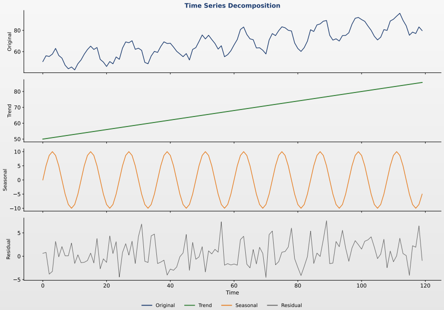
Why Study Time Series?



Time Series Are Everywhere

Finance (stock prices), Economics (GDP, inflation), Climate (temperature), Healthcare (patient monitoring), Energy (demand forecasting), and more.

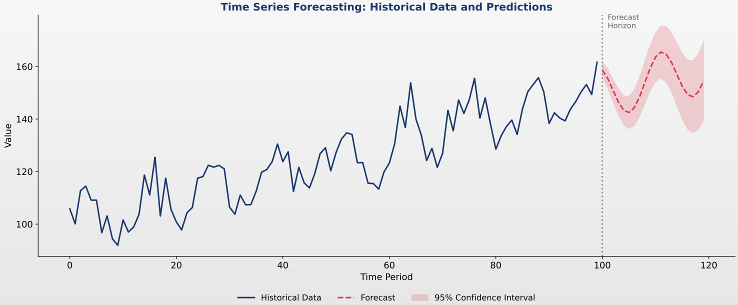
Time Series Components



Understanding Structure

Real time series often contain: **Trend** (long-term direction), **Seasonality** (repeating patterns), and **Irregular** (random fluctuations).

The Ultimate Goal: Forecasting



Why Forecasting Matters

Accurate predictions enable better decisions: inventory management, financial planning, risk assessment, and resource allocation.

Stochastic Process: Definition

Definition 1 (Stochastic Process)

A **stochastic process** is a collection of random variables indexed by time:

$$\{X_t(\omega) : t \in \mathcal{T}, \omega \in \Omega\}$$

where Ω is the sample space of possible outcomes.

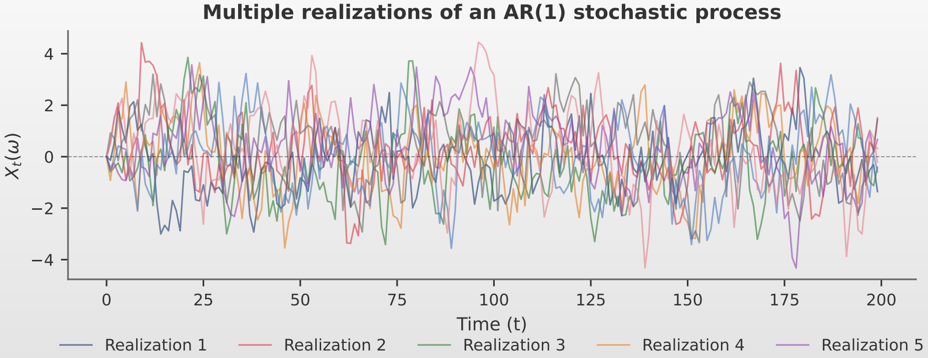
Two Perspectives

- **Fixed** ω: A *realization* $\{X_t(\omega)\}_{t\in\mathcal{T}}$
- **Fixed** t: A *random variable* X_t

Key Insight

A time series we observe is **one realization** of the underlying stochastic process.

Stochastic Process: Visual Illustration



Interpretation

Each line is a different realization from the same stochastic process. We observe only one realization but want to understand the process.

Moments of a Stochastic Process

First Two Moments Characterize Weak Properties

Mean Function: $\mu_t = \mathbb{E}[X_t]$

Autocovariance Function (ACVF):

$$\gamma(t, s) = \mathrm{Cov}(X_t, X_s) = \mathbb{E}[(X_t - \mu_t)(X_s - \mu_s)]$$

Autocorrelation Function (ACF):

$$\rho(t, s) = \frac{\gamma(t, s)}{\sqrt{\mathrm{Var}(X_t) \cdot \mathrm{Var}(X_s)}}$$

Properties

$\rho(t, s) \in [-1, 1]$ and $\rho(t, t) = 1$

Key Point

In general, μ_t and $\gamma(t, s)$ may depend on t.

Why Stationarity Matters

Stationarity is a fundamental assumption for time series analysis:

Without Stationarity:

- Mean, variance change over time
- Past may not predict future
- Standard methods fail
- Spurious correlations

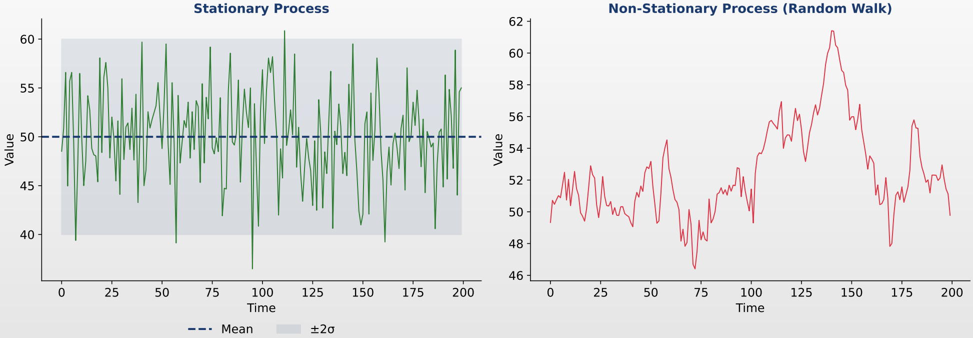
With Stationarity:

- Statistical properties constant
- Can estimate from one realization
- Valid inference possible
- Models are meaningful

Key Principle

Most time series models (ARMA, ARIMA, etc.) require stationarity. Non-stationary series must be transformed (e.g., differencing) before modeling.

Stationary vs Non-Stationary: Visual Comparison



- **Stationary**: Constant mean and variance – fluctuates around a fixed level
- **Non-stationary**: Mean and/or variance change over time
- Visual inspection is the first step; formal tests (ADF, KPSS) confirm

TSA_ch1_stationarity

Strict Stationarity

Definition 2 (Strict (Strong) Stationarity)

A process $\{X_t\}$ is **strictly stationary** if for all k, all $t_1, \ldots, t_k$, and all h:

$$(X_{t_1}, X_{t_2}, \ldots, X_{t_k}) \stackrel{d}{=} (X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_k+h})$$

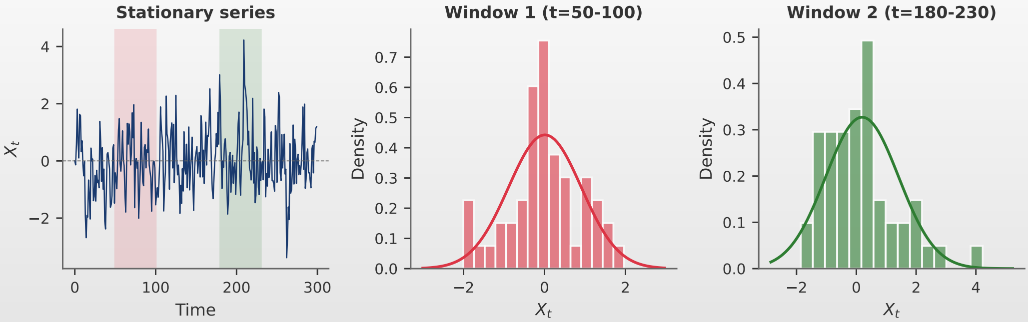
Implications

- All marginal distributions $F_{X_t}(x)$ identical
- $\mathbb{E}[X_t] = \mu$ (constant mean)
- $\text{Var}(X_t) = \sigma^2$ (constant variance)
- Joint distributions depend only on lag

Note

Strict stationarity is a strong condition, often impractical to verify.

Strict Stationarity: Visual Illustration



Interpretation

Stationary: any two windows have the same joint distribution. Non-stationary: distribution changes over time.

Weak (Covariance) Stationarity

Definition 3 (Weak Stationarity)

A process $\{X_t\}$ is **weakly stationary** (or covariance stationary) if:

1. $\mathbb{E}[X_t] = \mu$ (constant mean)
2. $\text{Var}(X_t) = \sigma^2 < \infty$ (constant, finite variance)
3. $\text{Cov}(X_t, X_{t+h}) = \gamma(h)$ (covariance depends only on lag h)

Key property: Autocovariance is a function of lag only:

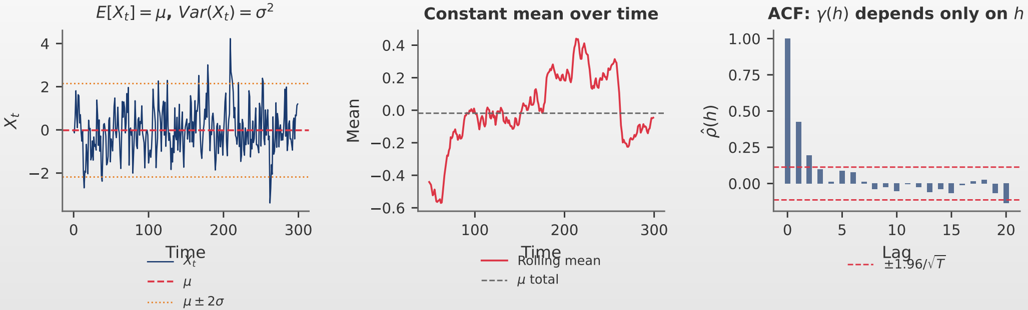
$$\gamma(h) = \text{Cov}(X_t, X_{t+h}) = \mathbb{E}[(X_t - \mu)(X_{t+h} - \mu)] \tag{1}$$

Autocorrelation function:

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{\text{Cov}(X_t, X_{t+h})}{\text{Var}(X_t)} \tag{2}$$

Note: $\rho(0) = 1$ and $\rho(h) = \rho(-h)$ (symmetry)

Weak Stationarity: Visual Illustration



Interpretation

Left: constant mean and variance. Right: autocovariance depends only on lag h, not time t.

Properties of the Autocovariance Function

ACVF Properties for Weakly Stationary Process

The ACVF $\gamma(h)$ satisfies:

1. **Symmetry:** $\gamma(h) = \gamma(-h)$
2. **Maximum at zero:** $|\gamma(h)| \leq \gamma(0)$
3. **Non-negative definiteness**

Implication

Not every function can be an autocovariance function.

White Noise Process

Definition 4 (White Noise)

A process $\{\varepsilon_t\}$ is **white noise**, denoted $\varepsilon_t \sim WN(0, \sigma^2)$, if:

1. $\mathbb{E}[\varepsilon_t] = 0$ for all t
2. $\text{Var}(\varepsilon_t) = \sigma^2$ for all t
3. $\text{Cov}(\varepsilon_t, \varepsilon_s) = 0$ for $t \neq s$

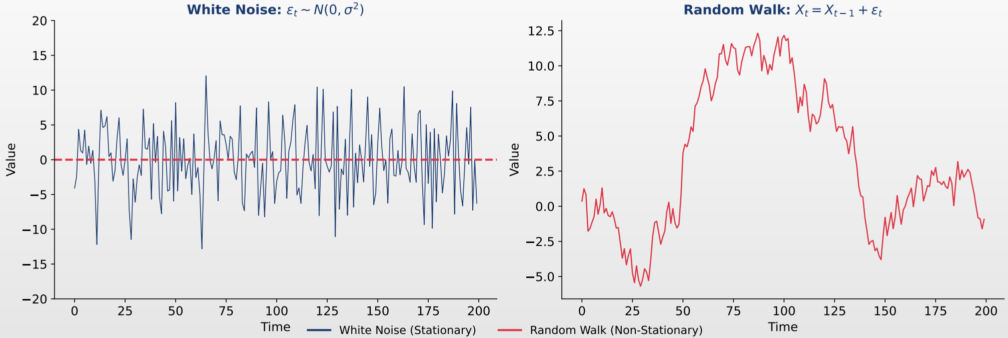
ACF of White Noise

$$\rho(h) = \begin{cases} 1 & \text{if } h = 0 \\ 0 & \text{if } h \neq 0 \end{cases}$$

Types

- **Weak**: Uncorrelated
- **Strong**: i.i.d.
- **Gaussian**: $\varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$

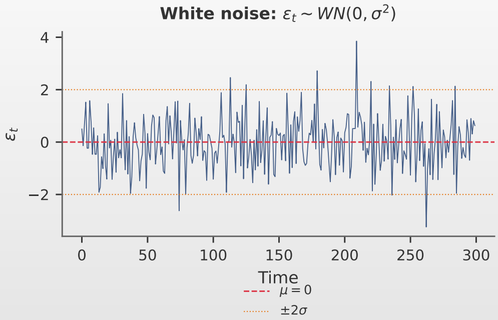
White Noise vs Random Walk: Comparison

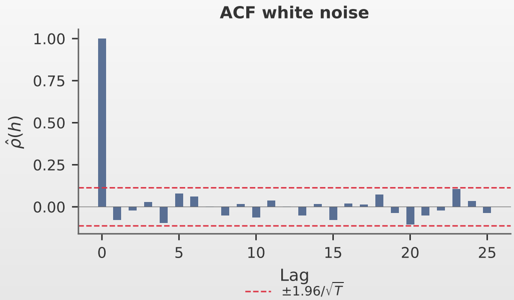


- **White noise**: Fluctuates around zero – stationary, constant variance
- **Random walk**: Cumulative sum of white noise – wanders away, non-stationary
- Random walk is the simplest non-stationary process (unit root)

TSA_ch1_wn_rw

White Noise: Visual Illustration





Interpretation

Left: white noise fluctuates around zero with constant variance. Right: ACF shows no autocorrelation (all zero after lag 0).

Random Walk Process

Definition

$X_t = X_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim WN(0, \sigma^2)$, $X_0 = 0$
Explicit form: $X_t = \sum_{i=1}^{t} \varepsilon_i$

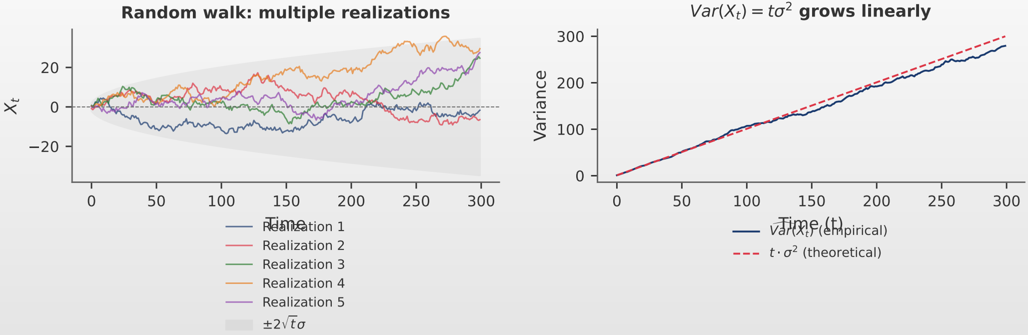
Properties

- $\mathbb{E}[X_t] = 0$ (constant mean)
- $\text{Var}(X_t) = t\sigma^2$ (grows with time!)
- $\text{Cov}(X_t, X_s) = \min(t, s) \cdot \sigma^2$

Non-Stationary!

Random walk is **not stationary** because variance depends on t.

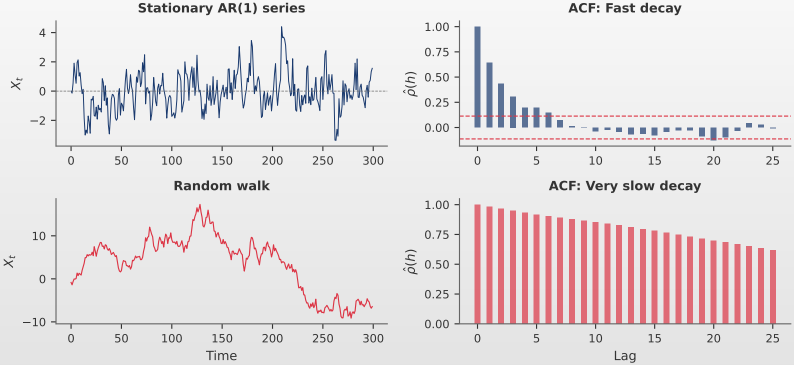
Random Walk: Visualization



Interpretation

Left: Multiple paths diverge over time. **Right:** Variance grows linearly: $\text{Var}(X_t) = t\sigma^2$.

ACF Comparison: Stationary vs Random Walk



Key Diagnostic

ACF of stationary process decays quickly; ACF of random walk decays very slowly.

Sample Autocorrelation Function

Sample ACF at Lag h

$$\hat{\rho}(h) = \frac{\sum_{t=1}^{T-h}(x_t - \bar{x})(x_{t+h} - \bar{x})}{\sum_{t=1}^{T}(x_t - \bar{x})^2} \tag{3}$$

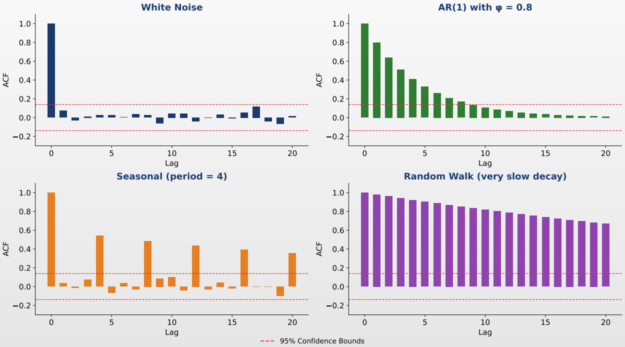
Properties

- $\hat{\rho}(0) = 1$ always
- $|\hat{\rho}(h)| \leq 1$

Significance Test

Under white noise: $\hat{\rho}(h) \approx N(0, 1/T)$
95% bounds: $\pm 1.96/\sqrt{T}$

ACF Patterns for Different Processes



- **White noise**: ACF drops to zero immediately (no dependence)
- **AR(1)**: ACF decays exponentially – indicates autoregressive structure
- **Seasonal**: ACF shows spikes at seasonal lags (e.g., 12, 24 for monthly)
- **Random walk**: ACF decays very slowly – sign of non-stationarity

TSA_ch1_acf_examples

Partial Autocorrelation Function (PACF)

Definition

PACF ϕ_{hh}: Correlation between X_t and X_{t+h} after removing the linear effect of $X_{t+1}, \ldots, X_{t+h-1}$.

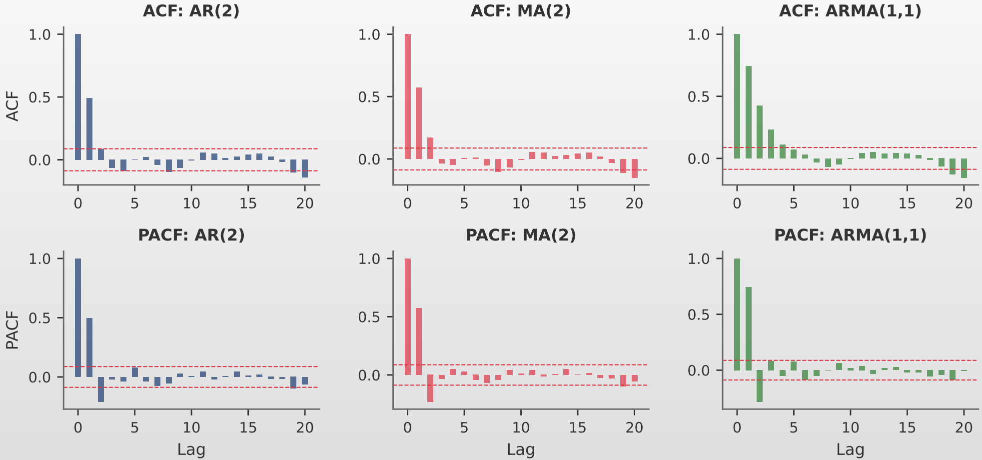
Interpretation

- $\phi_{11} = \rho(1)$ (same as ACF at lag 1)
- ϕ_{22}: correlation controlling for X_{t+1}
- Measures *direct* dependence

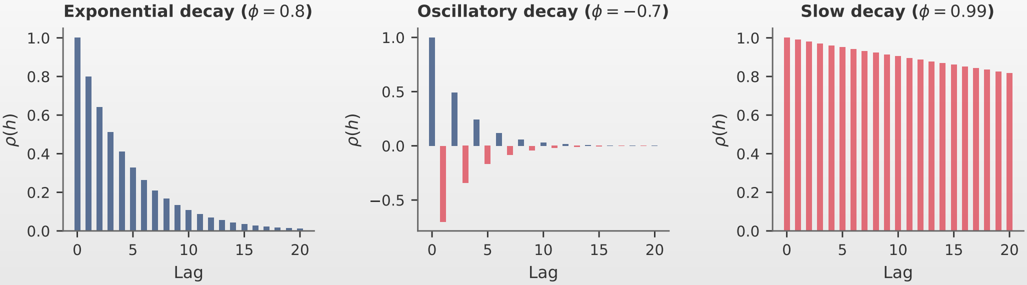
Key Application

- AR(p): PACF **cuts off** after lag p
- MA(q): ACF **cuts off** after lag q

ACF and PACF Patterns



Theoretical ACF for AR(1)



Interpretation

For AR(1): $X_t = \phi X_{t-1} + \varepsilon_t$, the theoretical ACF is $\rho(h) = \phi^h$.

The Lag Operator

Definition 5 (Lag Operator)

The **lag operator** (or backshift operator) L is defined by:

$$LX_t = X_{t-1}$$

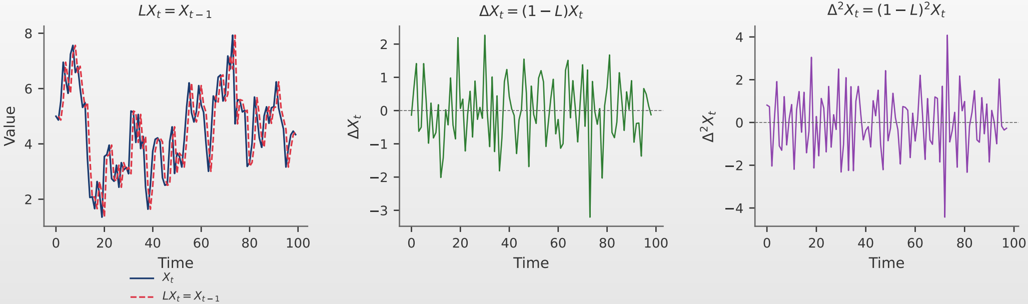
Properties

- $L^k X_t = X_{t-k}$ (lag by k periods)
- $L^0 = I$ (identity)
- $(1 - \phi L)X_t = X_t - \phi X_{t-1}$

Examples

- AR(1): $(1 - \phi L)X_t = \varepsilon_t$
- MA(1): $X_t = (1 + \theta L)\varepsilon_t$
- AR(p): $\phi(L)X_t = \varepsilon_t$

Lag Operator: Visual Illustration



Interpretation

The lag operator L shifts every observation back by one time period: $LX_t = X_{t-1}$.

Differencing

First Difference

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

Why Difference?

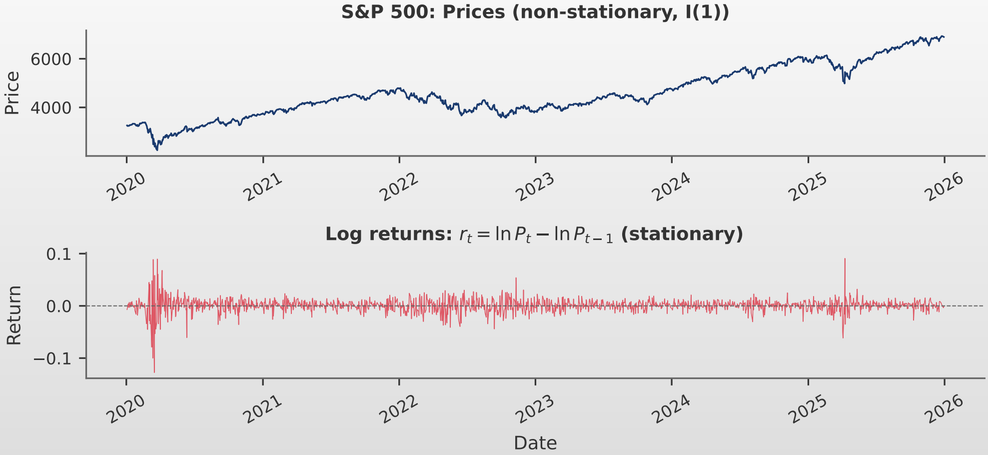
- Removes trend and unit root
- Random walk: $\Delta X_t = \varepsilon_t$ (white noise)

Integrated Process

$X_t \sim I(d)$ if $\Delta^d X_t$ is stationary

- $I(0)$: Stationary (no differencing)
- $I(1)$: One difference needed
- $I(2)$: Two differences needed

Effect of Differencing: S&P 500



Augmented Dickey-Fuller (ADF) Test

Model

$\Delta X_t = \alpha + \gamma X_{t-1} + \sum_{i=1}^{p} \delta_i \Delta X_{t-i} + \varepsilon_t$

Hypotheses

- H_0: $\gamma = 0$ (unit root)
- H_1: $\gamma < 0$ (stationary)

Test Statistic

$$\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})}$$

Decision Rule

$\tau <$ critical value $\Rightarrow$ Reject H_0 $\Rightarrow$ Stationary
$\tau \geq$ critical value $\Rightarrow$ Non-stationary

KPSS Test

Model

$X_t = \xi t + r_t + \varepsilon_t$ where $r_t = r_{t-1} + u_t$

Hypotheses (opposite of ADF)

- H_0: $\sigma_u^2 = 0$ (stationary)
- H_1: $\sigma_u^2 > 0$ (unit root)

Test Statistic

$$LM = \frac{\sum_{t=1}^{T} S_t^2}{T^2 \hat{\sigma}^2}$$

where $S_t = \sum_{i=1}^{t} \hat{e}_i$

Decision Rule

$LM >$ critical value $\Rightarrow$ Reject H_0 $\Rightarrow$ Non-stationary
$LM \leq$ critical value $\Rightarrow$ Stationary

Using ADF and KPSS Together

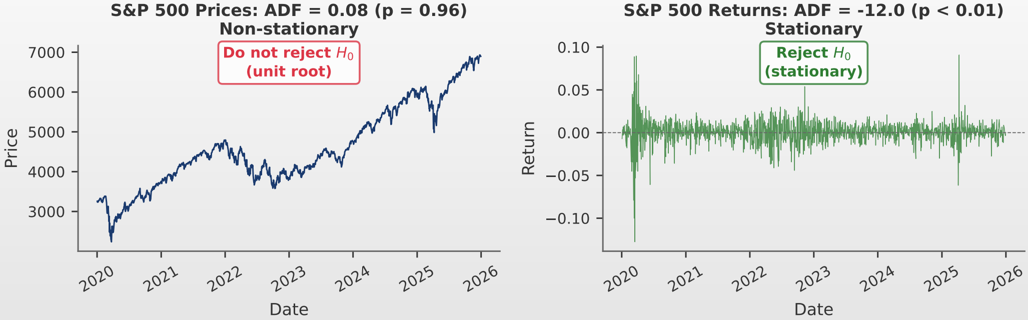
Confirmatory Testing for Robust Conclusions

ADF	KPSS	Conclusion
Reject H_0	Fail to reject H_0	Stationary
Fail to reject H_0	Reject H_0	Unit Root
Reject H_0	Reject H_0	Inconclusive
Fail to reject H_0	Fail to reject H_0	Inconclusive

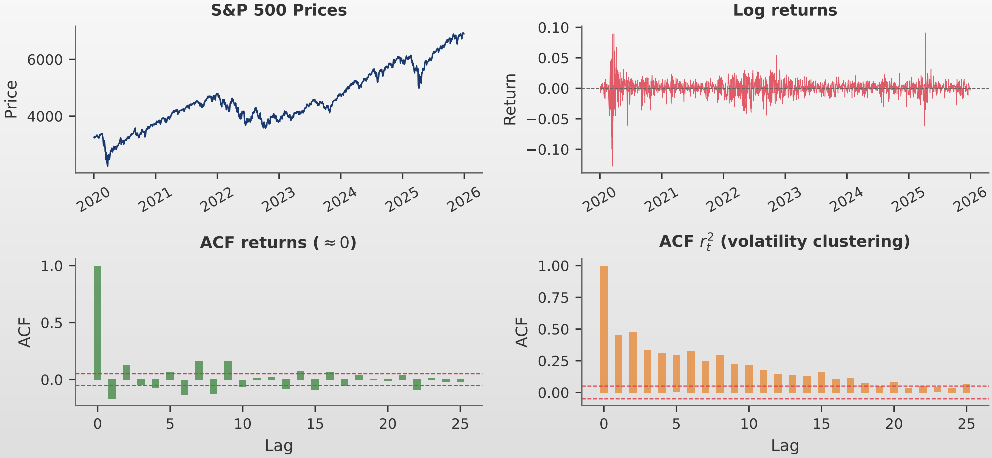
Recommended Workflow

1. Run ADF test (null = unit root)
2. Run KPSS test (null = stationary)
3. If results agree, proceed with confidence
4. If inconclusive, consider alternative tests (PP, DF-GLS)

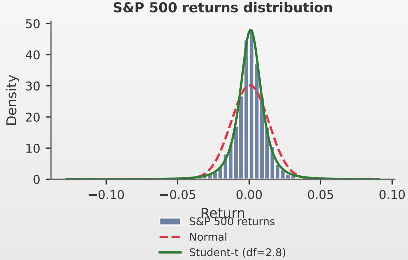
ADF Test: Visualization with S&P 500

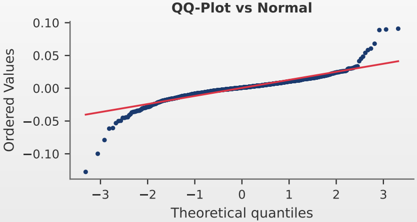


S&P 500 Analysis: Overview



Stylized Facts of Financial Returns





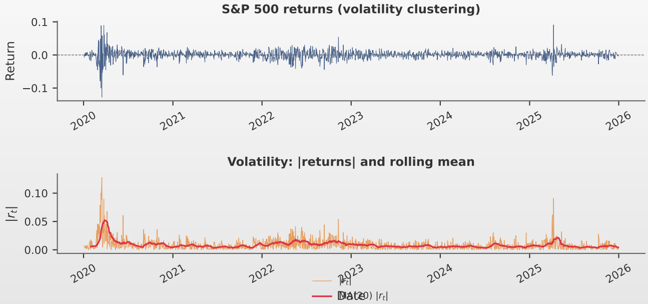
Observed properties:

- Negative skewness (left tail)
- Excess kurtosis ($\gg 3$)
- Heavy tails (fat tails)

Implications:

- Normal distribution inadequate
- Extreme events more likely
- Need Student-t or similar

Volatility Clustering



Stylized Fact

Large returns (positive or negative) tend to be followed by large returns. This **volatility clustering** motivates ARCH/GARCH models (future chapters).

Key Takeaways

Summary

1. **Time series** = observations indexed by time with temporal dependence
2. **Decomposition**: Additive $X_t = T_t + S_t + \varepsilon_t$ or Multiplicative
3. **Exponential Smoothing**: SES (level), Holt (trend), Holt-Winters (seasonal)
4. **Forecast Evaluation**: MAE, RMSE, MAPE; use train/validation/test splits
5. **Seasonality Modeling**: Dummy variables (any pattern) or Fourier terms (smooth)
6. **Trend Handling**: Differencing (stochastic) or regression (deterministic)
7. **Stationarity**: Mean, variance, autocovariance constant over time
8. **ACF/PACF**: Essential for identifying dependence structure
9. **Unit root tests**: ADF (H_0: unit root) vs KPSS (H_0: stationary)

Important Formulas I

Decomposition

Additive: $X_t = T_t + S_t + \varepsilon_t$ Multiplicative: $X_t = T_t \times S_t \times \varepsilon_t$

Simple Exponential Smoothing (SES)

$\hat{X}_{t+1|t} = \alpha X_t + (1-\alpha)\hat{X}_{t|t-1}$ where $\alpha \in (0,1)$

Holt's Linear Trend

$\ell_t = \alpha X_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$

Holt-Winters Additive

$\ell_t = \alpha(X_t - S_{t-s}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $S_t = \gamma(X_t - \ell_t) + (1-\gamma)S_{t-s}$

Important Formulas II

Moving Average (Trend Estimation)

$\hat{T}_t = \frac{1}{2q+1} \sum_{j=-q}^{q} X_{t+j}$

Autocovariance and Autocorrelation

$\gamma(h) = \mathrm{Cov}(X_t, X_{t+h}) \qquad \rho(h) = \frac{\gamma(h)}{\gamma(0)}$

Random Walk

$X_t = X_{t-1} + \varepsilon_t \quad \Rightarrow \quad \mathrm{Var}(X_t) = t\sigma^2$ (non-stationary)

Differencing

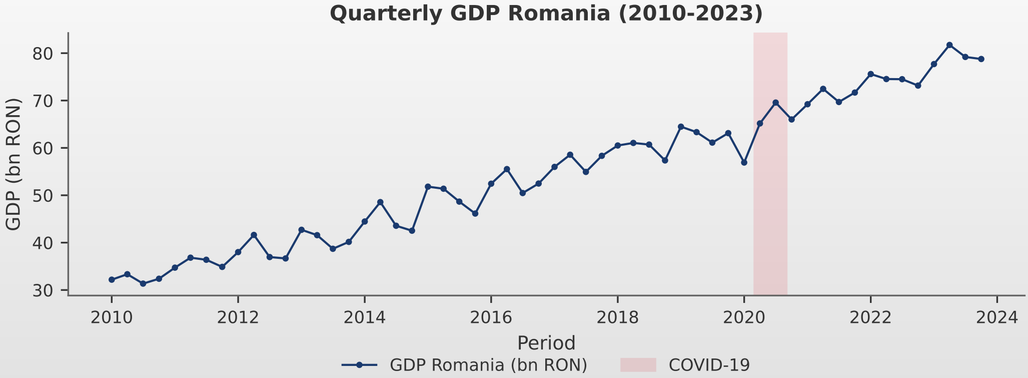
$\Delta X_t = (1 - L)X_t = X_t - X_{t-1}$

Next Chapter Preview

Chapter 2: ARMA Models

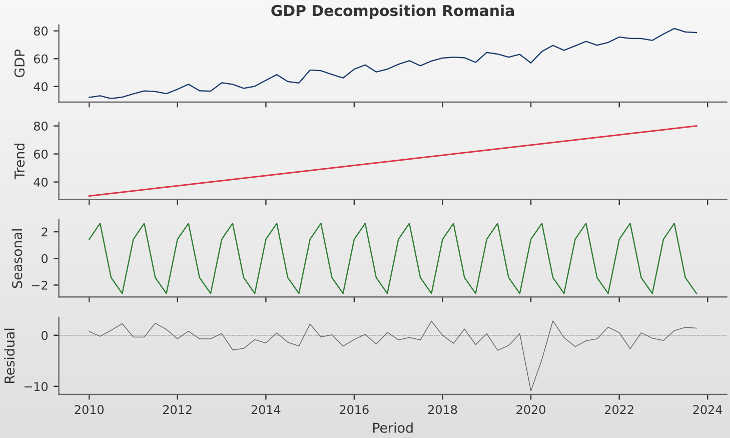
- Autoregressive (AR) models
- Moving Average (MA) models
- Combined ARMA models
- Model identification using ACF/PACF
- Parameter estimation
- Model diagnostics
- Forecasting

Case Study: Romanian Quarterly GDP



- **Data**: Romanian quarterly GDP, 2010–2023 (source: INS/Eurostat)
- **Observations**: Upward trend, quarterly seasonality, COVID-19 shock in 2020

GDP Series Decomposition



Interpreting the Components

Identified Components

- **Trend**: Sustained economic growth
- **Seasonality**: Regular quarterly pattern (Q4 > Q1)
- **Residual**: Includes COVID-19 shock from 2020

Lessons Learned

- Decomposition helps understand data structure
- External shocks (COVID) appear in residual component
- Seasonality must be modeled explicitly

Next Steps

In the following chapters we will learn to model each component: ARIMA for trend, SARIMA for seasonality.

Quiz Question 1

Question

A time series Y_t shows upward movement over years plus repeating patterns each quarter. Which components are present?

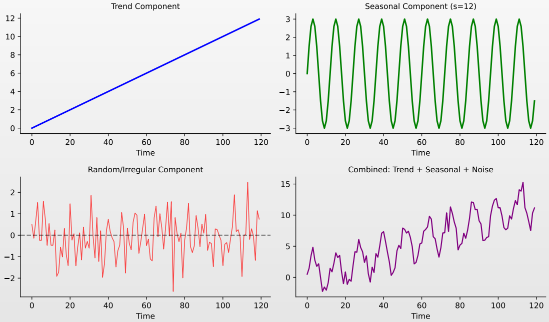
(A) Trend only

(B) Seasonality only

(C) Trend and Seasonality

(D) Random noise only

Quiz Question 1: Answer



Correct Answer: (C) Trend and Seasonality

Upward movement = Trend; Quarterly patterns = Seasonality (s=4)

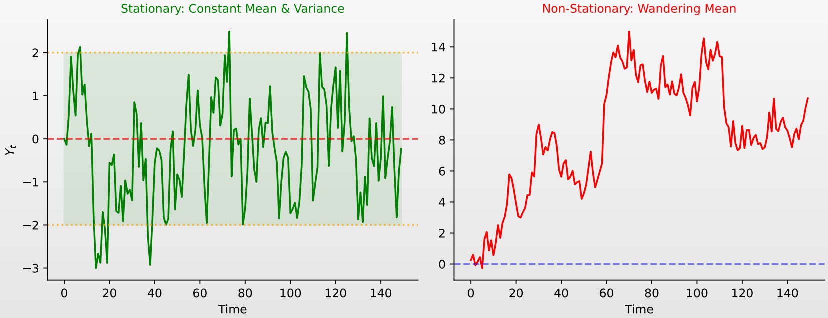
Quiz Question 2

Question

Which of the following is a characteristic of a stationary time series?

(A) Mean changes over time
(B) Variance increases with time
(C) Constant mean and variance over time
(D) Contains a trend component

Quiz Question 2: Answer



Correct Answer: (C) Constant mean and variance over time

Stationarity requires: constant mean, constant variance, and autocovariance depends only on lag.

Quiz Question 3

Question

For a white noise process, what does the ACF look like at lags $k > 0$?

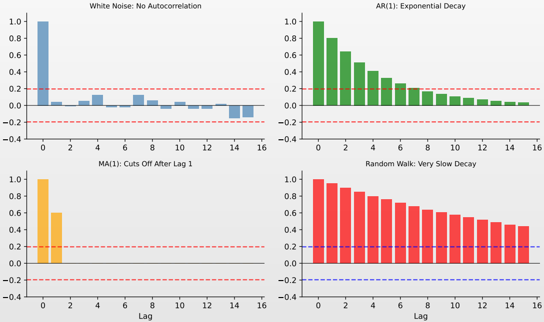
(A) Exponential decay

(B) All values significant and positive

(C) All values approximately zero (within confidence bands)

(D) Alternating positive and negative

Quiz Question 3: Answer



Correct Answer: (C) Approximately zero within confidence bands

White noise has no autocorrelation: $\rho_k = 0$ for all $k \neq 0$.

Quiz Question 4

Question

What is the key difference between white noise and a random walk?

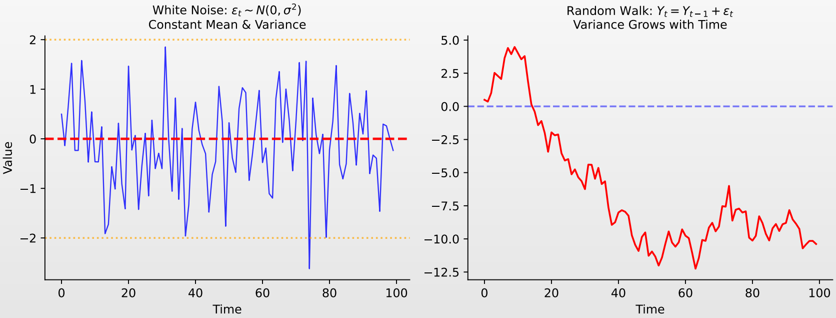
(A) White noise has a trend, random walk doesn't

(B) Random walk is the cumulative sum of white noise

(C) Both are stationary processes

(D) White noise has higher variance

Quiz Question 4: Answer



Correct Answer: (B) Random walk = cumulative sum of white noise

$Y_t = Y_{t-1} + \varepsilon_t = \sum_{i=1}^{t} \varepsilon_i$ where ε_t is white noise.

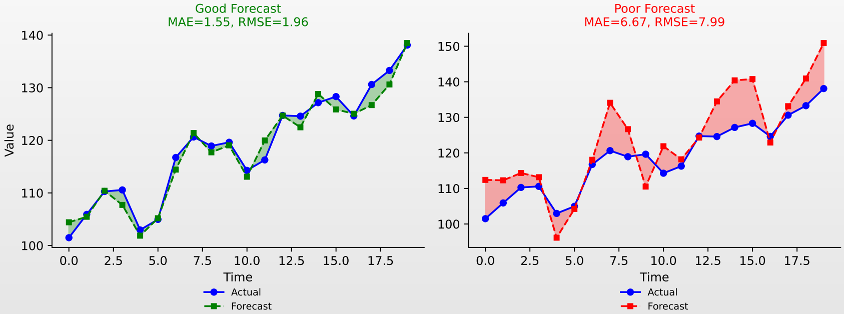
Quiz Question 5

Question

Which forecast error metric is most sensitive to large errors (outliers)?

(A) MAE (Mean Absolute Error)
(B) RMSE (Root Mean Squared Error)
(C) MAPE (Mean Absolute Percentage Error)
(D) All are equally sensitive

Quiz Question 5: Answer



Correct Answer: (B) RMSE

RMSE squares errors, so large errors have disproportionate impact: $\sqrt{\frac{1}{n}\sum e_t^2}$

Quiz Question 6

Question

When should you use multiplicative decomposition instead of additive?

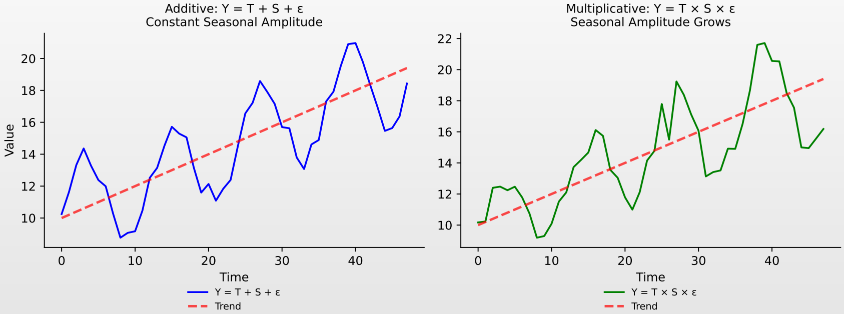
(A) When the series has no trend

(B) When seasonal amplitude is constant

(C) When seasonal amplitude grows with the level of the series

(D) When the series is stationary

Quiz Question 6: Answer



Correct Answer: (C) Seasonal amplitude grows with level

Multiplicative: $Y_t = T_t \times S_t \times \varepsilon_t$ — seasonal swings proportional to trend.

References

- Hyndman, R.J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*. 3rd ed., OTexts.
- Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press.
- Box, G.E.P., Jenkins, G.M., Reinsel, G.C., & Ljung, G.M. (2015). *Time Series Analysis: Forecasting and Control*. 5th ed., Wiley.
- Tsay, R.S. (2010). *Analysis of Financial Time Series*. 3rd ed., Wiley.
- Cleveland, R.B., Cleveland, W.S., McRae, J.E., & Terpenning, I. (1990). STL: A Seasonal-Trend Decomposition. *Journal of Official Statistics*, 6(1), 3-73.

Data Sources

Real Data Used in This Chapter

- **Airline Passengers**: Box-Jenkins classic dataset, 1949–1960
- **S&P 500**: Yahoo Finance (SPY), historical data
- **Sunspots**: Statsmodels dataset, monthly observations

Software & Tools

- **Python**: statsmodels, pandas, matplotlib, yfinance
- **R**: forecast, tseries packages
- **Data Sources**: Yahoo Finance, FRED Economic Data

Thank You!

Questions?

Charts generated using Python (statsmodels, matplotlib)

Course materials available at: `https://github.com/danpele/Time-Series-Analysis`