



## Time Series Analysis and Forecasting

### Chapter 3: ARIMA Models for Non-Stationary Data

Bachelor program, Faculty of Cybernetics, Statistics and Economic Informatics  
Bucharest University of Economic Studies, Romania



Prof. dr. Daniel Traian Pele  
danpele@ese.ro



Academic Year 2025–2026



# Outline

- 1 Non-Stationarity in Time Series
- 2 Differencing and the Difference Operator
- 3 ARIMA(p,d,q) Models
- 4 Unit Root Tests
- 5 ARIMA Model Identification
- 6 ARIMA Estimation
- 7 Diagnostic Checking
- 8 Forecasting with ARIMA
- 9 Practical Example
- 10 Summary

# Why Non-Stationarity Matters

## The Problem

Many economic and financial time series are **non-stationary**:

- GDP, stock prices, exchange rates, inflation indices
- They exhibit trends, changing means, or growing variance

## Consequences of Non-Stationarity

- Standard ARMA models assume stationarity
- OLS regression with non-stationary data leads to **spurious regression**
- Sample moments (mean, variance, ACF) are not consistent estimators
- Statistical inference becomes invalid

# Types of Non-Stationarity

## Deterministic Trend

$$Y_t = \alpha + \beta t + \varepsilon_t$$

- Trend is a deterministic function of time
- Can be removed by **detrending**
- Shocks have temporary effects

## Stochastic Trend (Unit Root)

$$Y_t = Y_{t-1} + \varepsilon_t$$

- Random walk process
- Must be removed by **differencing**
- Shocks have permanent effects

## Key Distinction

Correct identification is crucial: detrending a unit root process or differencing a trend-stationary process both lead to misspecification!

# The Random Walk Process

## Definition 1 (Random Walk)

A **random walk** is defined as:

$$Y_t = Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2)$$

With initial condition  $Y_0 = 0$ , we have:

$$Y_t = \sum_{i=1}^t \varepsilon_i$$

## Properties of Random Walk

- $\mathbb{E}[Y_t] = 0$  (constant mean)
- $\text{Var}(Y_t) = t\sigma^2$  (variance grows with time!)
- $\text{Cov}(Y_t, Y_{t-k}) = (t-k)\sigma^2$  for  $k \leq t$
- ACF:  $\rho_k = \sqrt{\frac{t-k}{t}} \rightarrow 1$  as  $t \rightarrow \infty$

## Definition 2 (Random Walk with Drift)

A random walk with drift includes a constant term:

$$Y_t = \mu + Y_{t-1} + \varepsilon_t$$

Equivalently:

$$Y_t = Y_0 + \mu t + \sum_{i=1}^t \varepsilon_i$$

## Properties

- $\mathbb{E}[Y_t] = Y_0 + \mu t$  (mean grows linearly)
- $\text{Var}(Y_t) = t\sigma^2$  (variance still grows)
- The drift  $\mu$  creates an upward or downward trend
- Still non-stationary despite having a “trend”

## Definition 3 (Integrated Process of Order $d$ )

A time series  $\{Y_t\}$  is **integrated of order  $d$** , written  $Y_t \sim I(d)$ , if:

- $Y_t$  is non-stationary
- $(1 - L)^d Y_t = \Delta^d Y_t$  is stationary
- $(1 - L)^{d-1} Y_t$  is still non-stationary

## Common Cases

- $I(0)$ : Stationary process (e.g., ARMA)
- $I(1)$ : First difference is stationary (most common for economic data)
- $I(2)$ : Second difference is stationary (less common)

# The Difference Operator

## Definition 4 (First Difference)

The **first difference operator**  $\Delta$  is defined as:

$$\Delta Y_t = Y_t - Y_{t-1} = (1 - L)Y_t$$

where  $L$  is the lag operator ( $LY_t = Y_{t-1}$ ).

## Higher-Order Differences

- Second difference:  $\Delta^2 Y_t = \Delta(\Delta Y_t) = (1 - L)^2 Y_t$
- $\Delta^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2}$
- $d$ -th difference:  $\Delta^d Y_t = (1 - L)^d Y_t$

## Key Result

If  $Y_t \sim I(d)$ , then  $\Delta^d Y_t \sim I(0)$  (stationary).

## Example: Differencing a Random Walk

### Random Walk to White Noise

Let  $Y_t = Y_{t-1} + \varepsilon_t$  (random walk).

Taking the first difference:

$$\Delta Y_t = Y_t - Y_{t-1} = \varepsilon_t$$

The first difference is white noise – a stationary process!

### Interpretation

- A random walk is  $I(1)$
- One difference transforms it to  $I(0)$
- The “changes” in a random walk are stationary

# Overdifferencing

## Warning: Overdifferencing

Differencing more than necessary introduces problems:

- Creates artificial negative autocorrelation
- Inflates variance
- Loses information

## Example

If  $Y_t \sim I(1)$ , then  $\Delta Y_t \sim I(0)$ .

But if we difference again:

$$\Delta^2 Y_t = \Delta Y_t - \Delta Y_{t-1} = \varepsilon_t - \varepsilon_{t-1}$$

This is an MA(1) with  $\theta = 1$  (non-invertible boundary)!

# Definition of ARIMA

## Definition 5 (ARIMA(p,d,q))

A time series  $\{Y_t\}$  follows an **ARIMA(p,d,q)** process if:

$$\phi(L)(1-L)^d Y_t = c + \theta(L)\varepsilon_t$$

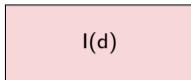
where:

- $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$  (AR polynomial)
- $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q$  (MA polynomial)
- $d$  is the order of integration (number of differences)
- $\varepsilon_t \sim WN(0, \sigma^2)$

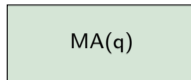
# ARIMA Components



Autoregressive  
Memory



Integration  
Differencing



Moving Average  
Shocks

## Special Cases

- $ARIMA(p,0,q) = ARMA(p,q)$  – stationary
- $ARIMA(0,1,0) =$  Random walk
- $ARIMA(0,1,1) = IMA(1,1)$  – exponential smoothing
- $ARIMA(1,1,0) = ARI(1,1)$  – differenced  $AR(1)$

## ARIMA(1,1,0) Example

### AR(1,1) Model

$$\Delta Y_t = c + \phi_1 \Delta Y_{t-1} + \varepsilon_t$$

Equivalently:

$$(1 - \phi_1 L)(1 - L)Y_t = c + \varepsilon_t$$

### Interpretation

- The **changes** in  $Y_t$  follow an AR(1) process
- If  $|\phi_1| < 1$ , the changes are stationary
- $Y_t$  itself has a stochastic trend
- Common model for many economic time series

## ARIMA(0,1,1) Example

### IMA(1,1) Model

$$\Delta Y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Equivalently:

$$(1 - L)Y_t = c + (1 + \theta_1 L)\varepsilon_t$$

### Connection to Exponential Smoothing

The IMA(1,1) model is equivalent to **simple exponential smoothing**:

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha) \hat{Y}_t$$

where  $\alpha = 1 + \theta_1$  (for  $-1 < \theta_1 < 0$ ).

# The Role of the Constant in ARIMA

## Constant Term in ARIMA(p,d,q)

When  $d > 0$ , the constant  $c$  has a different interpretation:

$$\phi(L)(1 - L)^d Y_t = c + \theta(L)\varepsilon_t$$

## Important Implications

- For  $d = 1$ :  $c$  represents the **drift** (average change)

$$\mathbb{E}[\Delta Y_t] = \frac{c}{1 - \phi_1 - \dots - \phi_p}$$

- For  $d = 2$ :  $c$  affects the **curvature** of the trend
- Often  $c = 0$  is assumed when  $d \geq 1$

# Testing for Unit Roots

## Why Test?

Before fitting an ARIMA model, we need to determine:

- 1 Is the series stationary? (Is  $d = 0$ ?)
- 2 If not, how many differences are needed? (What is  $d$ ?)

## Common Unit Root Tests

- **Dickey-Fuller (DF)** and **Augmented Dickey-Fuller (ADF)**
- **Phillips-Perron (PP)**
- **KPSS** (stationarity test – reversed null hypothesis)

# The Dickey-Fuller Test

## Setup

Consider the AR(1) model:  $Y_t = \phi Y_{t-1} + \varepsilon_t$

Subtract  $Y_{t-1}$  from both sides:

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

where  $\gamma = \phi - 1$ .

## Hypotheses

- $H_0$ :  $\gamma = 0$  (unit root,  $\phi = 1$ , non-stationary)
- $H_1$ :  $\gamma < 0$  (stationary,  $|\phi| < 1$ )

## Key Issue

Under  $H_0$ , the  $t$ -statistic does **not** follow a standard  $t$ -distribution! Must use Dickey-Fuller critical values.

## Three Specifications

- ① No constant, no trend:

$$\Delta Y_t = \gamma Y_{t-1} + \varepsilon_t$$

- ② With constant (drift):

$$\Delta Y_t = \alpha + \gamma Y_{t-1} + \varepsilon_t$$

- ③ With constant and trend:

$$\Delta Y_t = \alpha + \beta t + \gamma Y_{t-1} + \varepsilon_t$$

## Choosing the Right Specification

- Examine the data: does it have a visible trend?
- Including unnecessary terms reduces power
- Excluding necessary terms leads to incorrect inference

# Augmented Dickey-Fuller (ADF) Test

## The Problem with Simple DF

If the true model has AR dynamics beyond AR(1), the DF test residuals will be autocorrelated, invalidating the test.

## Definition 6 (ADF Test)

Add lagged differences to absorb serial correlation:

$$\Delta Y_t = \alpha + \beta t + \gamma Y_{t-1} + \sum_{j=1}^k \delta_j \Delta Y_{t-j} + \varepsilon_t$$

Test  $H_0 : \gamma = 0$  using ADF critical values.

## Choosing Lag Length $k$

- Use information criteria (AIC, BIC)
- Start with  $k_{max}$  and reduce until last lag is significant

## ADF Test Critical Values

Model	1%	5%	10%
No constant, no trend	-2.58	-1.95	-1.62
With constant	-3.43	-2.86	-2.57
With constant and trend	-3.96	-3.41	-3.13

Table: Asymptotic critical values for ADF test (approximate)

### Decision Rule

- If test statistic  $<$  critical value  $\Rightarrow$  Reject  $H_0$  (stationary)
- If test statistic  $\geq$  critical value  $\Rightarrow$  Fail to reject  $H_0$  (unit root)

# The KPSS Test

## Reversed Hypotheses

Unlike ADF, the KPSS test has:

- $H_0$ : Series is stationary (possibly around a trend)
- $H_1$ : Series has a unit root

## KPSS Procedure

Decompose:  $Y_t = \xi t + r_t + \varepsilon_t$

where  $r_t = r_{t-1} + u_t$  is a random walk.

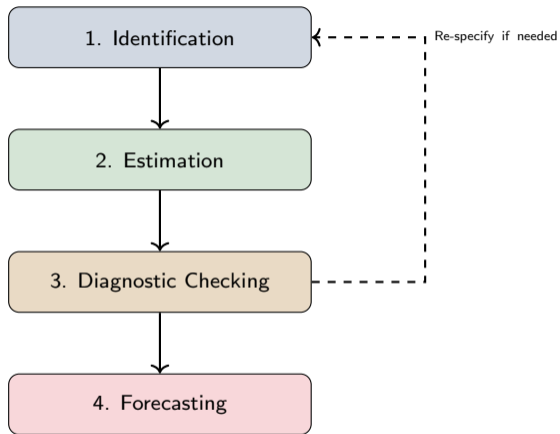
Test whether  $\text{Var}(u_t) = 0$  (stationarity).

## Complementary Use

Use ADF and KPSS together:

- ADF rejects, KPSS doesn't reject  $\Rightarrow$  Stationary
- ADF doesn't reject, KPSS rejects  $\Rightarrow$  Unit root
- Both reject or neither  $\Rightarrow$  Inconclusive

# The Box-Jenkins Methodology



## Step 1: Determining $d$

### Procedure

- 1 Plot the time series – look for trends, changing variance
- 2 Examine ACF – slow decay suggests non-stationarity
- 3 Apply unit root tests (ADF, KPSS)
- 4 If non-stationary, difference and repeat

### Practical Guidelines

- Most economic series:  $d = 1$  is sufficient
- Rarely need  $d > 2$
- If ACF of  $\Delta Y_t$  still decays slowly, try  $d = 2$
- Watch for overdifferencing (ACF with  $\rho_1 \approx -0.5$ )

## Step 2: Determining $p$ and $q$

### After Differencing

Once  $W_t = \Delta^d Y_t$  is stationary, use ACF/PACF to identify ARMA( $p, q$ ):

Model	ACF	PACF
AR( $p$ )	Decays exponentially	Cuts off after lag $p$
MA( $q$ )	Cuts off after lag $q$	Decays exponentially
ARMA( $p, q$ )	Decays	Decays

### Information Criteria

When patterns are unclear, compare models using:

- $AIC = -2 \ln(L) + 2k$
- $BIC = -2 \ln(L) + k \ln(n)$

Lower is better. BIC penalizes complexity more.

## Automated Model Selection

Modern software can automatically select  $(p, d, q)$ :

- Python: `pmdarima.auto_arima()`
- R: `forecast::auto.arima()`

## How Auto-ARIMA Works

- 1 Use unit root tests to determine  $d$
- 2 Fit models for various  $(p, q)$  combinations
- 3 Select model with lowest AIC/BIC
- 4 Optionally use stepwise search for efficiency

## Caution

Automated selection is helpful but not infallible. Always check diagnostics!

## Maximum Likelihood Estimation (MLE)

The standard approach for ARIMA:

- Assumes  $\varepsilon_t \sim N(0, \sigma^2)$
- Maximizes the likelihood function
- Provides consistent, efficient estimators
- Yields standard errors for inference

## Conditional vs Exact MLE

- **Conditional MLE:** Conditions on initial values
- **Exact MLE:** Treats initial values as unknown
- Difference diminishes as sample size grows

## Stationarity and Invertibility

The estimated ARIMA model should satisfy:

- **AR stationarity:** Roots of  $\phi(z) = 0$  outside unit circle
- **MA invertibility:** Roots of  $\theta(z) = 0$  outside unit circle

## Checking in Practice

Most software reports:

- Estimated coefficients with standard errors
- Roots of AR and MA polynomials
- Warning if near-unit-root detected

## What to Check

If the model is correct, residuals  $\hat{\varepsilon}_t$  should be white noise:

- 1 Zero mean
- 2 Constant variance
- 3 No autocorrelation
- 4 (Optional) Normality

## Diagnostic Tools

- **Residual ACF/PACF:** Should show no significant spikes
- **Ljung-Box test:** Tests for autocorrelation at multiple lags
- **Q-Q plot:** Checks normality assumption
- **Residual vs fitted:** Checks for heteroskedasticity

# The Ljung-Box Test

## Definition 7 (Ljung-Box Q Statistic)

$$Q(m) = n(n+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{n-k}$$

Under  $H_0$  (no autocorrelation):  $Q(m) \sim \chi^2(m - p - q)$

## Usage

- Choose  $m \approx \ln(n)$  or  $m = 10$  for quarterly,  $m = 20$  for monthly
- Degrees of freedom adjusted for estimated parameters
- Reject if  $Q(m)$  exceeds critical value

## If Test Fails

Consider adding AR or MA terms, or check for structural breaks.

### Minimum MSE Forecast

The optimal  $h$ -step ahead forecast is the conditional expectation:

$$\hat{Y}_{T+h|T} = \mathbb{E}[Y_{T+h}|Y_T, Y_{T-1}, \dots]$$

### ARIMA(1,1,1) Forecasting

Model:  $(1 - \phi_1 L)(1 - L)Y_t = c + (1 + \theta_1 L)\varepsilon_t$

One-step forecast:

$$\hat{Y}_{T+1|T} = c + Y_T + \phi_1(Y_T - Y_{T-1}) + \theta_1 \hat{\varepsilon}_T$$

For  $h > 1$ : replace unknown  $\varepsilon_{T+j}$  with 0, unknown  $Y_{T+j}$  with  $\hat{Y}_{T+j|T}$

# Forecast Intervals

## Forecast Uncertainty

The  $h$ -step forecast error variance:

$$\text{Var}(e_{T+h}) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2$$

where  $\psi_j$  are  $\text{MA}(\infty)$  coefficients.

## Confidence Intervals

Under normality,  $(1 - \alpha)\%$  interval:

$$\hat{Y}_{T+h|T} \pm z_{\alpha/2} \sqrt{\text{Var}(e_{T+h})}$$

## Key Property for $I(1)$ Series

For integrated processes, forecast variance grows without bound as  $h \rightarrow \infty$ . Intervals widen over time!

# Long-Run Forecasts for ARIMA

## Behavior as $h \rightarrow \infty$

For ARIMA(p,1,q) with drift  $c$ :

- Point forecasts: Linear trend with slope = drift
- Forecast intervals: Width grows with  $\sqrt{h}$

For ARIMA(p,1,q) without drift:

- Point forecasts: Converge to last level
- Forecast intervals: Still grow unboundedly

## Practical Implication

ARIMA forecasts are most reliable for short horizons. Long-term forecasts have very wide uncertainty bands.

## Example: Modeling GDP Growth

### Typical Workflow

- 1 **Plot:** Observe trend in  $\log(\text{GDP})$
- 2 **Test:** ADF suggests  $I(1)$
- 3 **Difference:** Work with GDP growth rate
- 4 **Identify:** ACF/PACF suggest  $\text{ARIMA}(1,1,0)$  or  $\text{ARIMA}(0,1,1)$
- 5 **Estimate:** Fit both, compare AIC
- 6 **Diagnose:** Check residuals
- 7 **Forecast:** Generate predictions with intervals

## Key Libraries

```
import pandas as pd
import numpy as np
from statsmodels.tsa.arima.model import ARIMA
from statsmodels.tsa.stattools import adfuller, kpss
import pmdarima as pm
```

## Auto-ARIMA Example

```
# Automatic model selection
model = pm.auto_arima(y, start_p=0, start_q=0,
                     max_p=3, max_q=3, d=None,
                     seasonal=False, trace=True,
                     error_action='ignore')
print(model.summary())
```

# Key Takeaways




## Main Points

- 1 **Non-stationarity** is common in economic data – must be addressed
- 2 **Differencing** transforms  $I(d)$  to  $I(0)$
- 3 **ARIMA(p,d,q)** combines differencing with ARMA modeling
- 4 **Unit root tests** (ADF, KPSS) help determine  $d$
- 5 **Box-Jenkins methodology**: Identify → Estimate → Diagnose
- 6 **Forecasts** for  $I(1)$  series have growing uncertainty

## Next Steps

Chapter 4 will extend ARIMA to handle seasonality: SARIMA models.

## References

-  Box, G.E.P., Jenkins, G.M., Reinsel, G.C., & Ljung, G.M. (2015). *Time Series Analysis: Forecasting and Control*. 5th ed. Wiley.
-  Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press.
-  Enders, W. (2014). *Applied Econometric Time Series*. 4th ed. Wiley.
-  Hyndman, R.J. & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*. 3rd ed. OTexts.