Time Series Analysis and Forecasting

Chapter 3: ARIMA Models

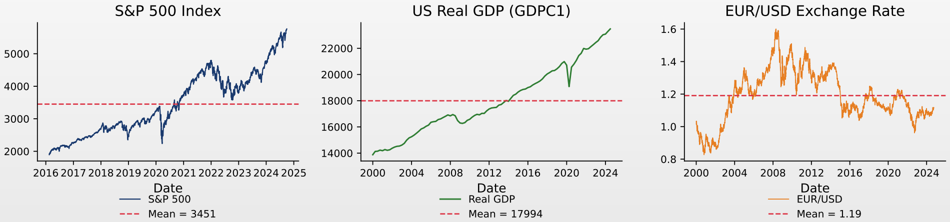
Daniel Traian PELE

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

Outline

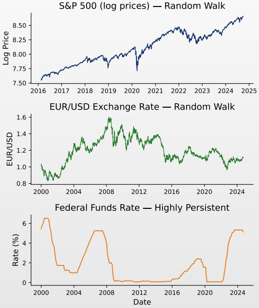
Motivating Example: Non-Stationary Data Is Everywhere

Non-stationary data: sample mean is meaningless



- Stock prices, GDP, exchange rates all exhibit **trends** or **wandering behavior**
- The sample mean (red line) is meaningless for a random walk
- Standard ARMA models **cannot** handle these series directly

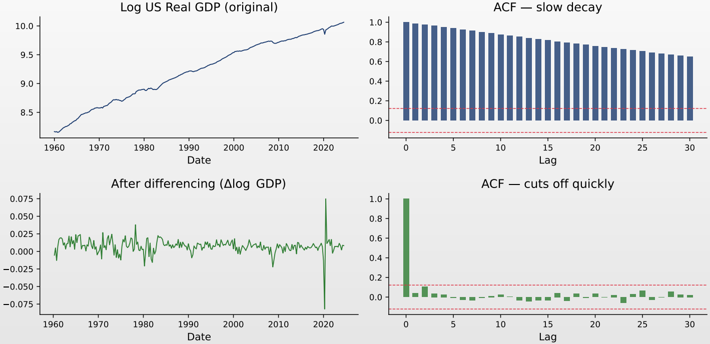
Real-World Applications



The Challenge

Financial/economic data are typically I(1): stock prices, exchange rates, interest rates.

The Solution: Differencing



Key Insight

Differencing transforms a non-stationary series into a stationary one: $\Delta Y_t = Y_t - Y_{t-1}$. The ACF changes from slow decay to quick decay!

What We'll Learn Today

Core Concepts

1. **Non-Stationarity**: Why it matters and how to detect it
2. **Unit Root Tests**: ADF, PP, KPSS tests
3. **Differencing**: The key transformation
4. **ARIMA Models**: Combining differencing with ARMA
5. **Box-Jenkins Methodology**: Identify $\rightarrow$ Estimate $\rightarrow$ Diagnose

By the End of This Lecture

You will be able to model and forecast non-stationary time series like stock prices, GDP, and exchange rates using ARIMA models.

Why Non-Stationarity Matters

The Problem

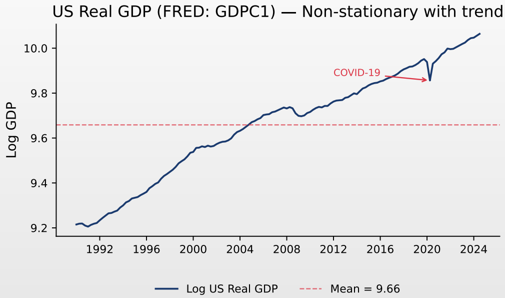
Many economic and financial time series are **non-stationary**:

- 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

Example: US Real GDP



- Clear upward **trend** – mean is not constant
- This is a classic example of a **non-stationary** time series
- We cannot apply ARMA models directly to this data

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)

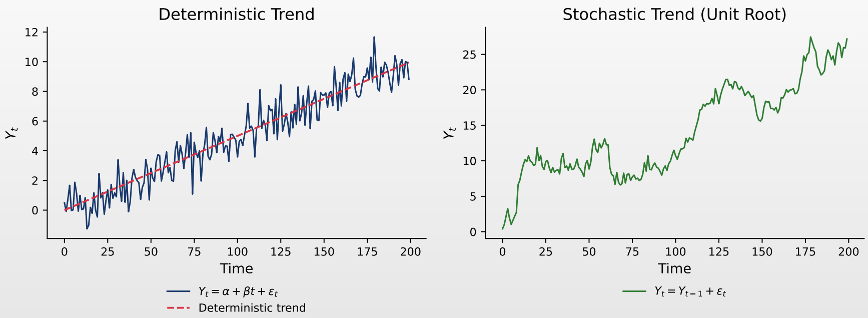
$$Y_t = Y_{t-1} + \varepsilon_t$$

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

Key Distinction

Correct identification is crucial: detrending a unit root $\rightarrow$ misspecification; differencing trend-stationary $\rightarrow$ misspecification.

Visualizing the Difference



- **Left**: Deterministic trend – deviations from trend are temporary
- **Right**: Stochastic trend – shocks accumulate permanently
- Both look similar, but require **different** treatments!

The Random Walk Process

Definition 1 (Random Walk)

A **random walk** is defined as:

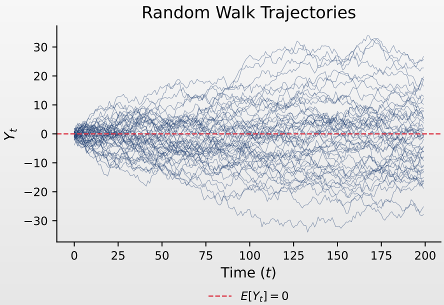
$$Y_t = Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2)$$

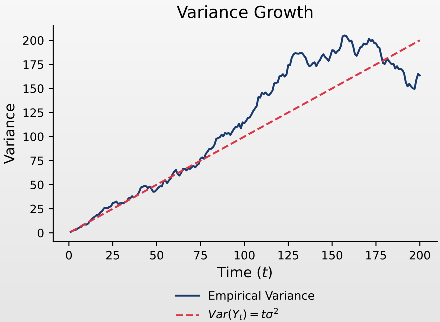
With initial condition $Y_0 = 0$, we have: $Y_t = \sum_{i=1}^{t} \varepsilon_i$

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: $\rho_k = \sqrt{\frac{t-k}{t}} \to 1$ as $t \to \infty$

Random Walk: Visual Illustration





Key Properties

Left: Paths wander unpredictably, no mean reversion. **Right**: $\mathrm{Var}(Y_t) = t\sigma^2$ grows linearly $\Rightarrow$ non-stationary.

Proof: Random Walk Variance

Claim: For $Y_t = Y_{t-1} + \varepsilon_t$ with $Y_0 = 0$: $\text{Var}(Y_t) = t\sigma^2$

Proof: By recursive substitution: $Y_t = \sum_{i=1}^{t} \varepsilon_i$
Taking variance:

$$\text{Var}(Y_t) = \text{Var}\left(\sum_{i=1}^{t} \varepsilon_i\right) = \sum_{i=1}^{t} \text{Var}(\varepsilon_i) + 2\sum_{i<j} \text{Cov}(\varepsilon_i, \varepsilon_j)$$

Since ε_t independent (white noise): $\text{Var}(Y_t) = \sum_{i=1}^{t} \sigma^2 = \boxed{t\sigma^2}$

Variance depends on $t \Rightarrow$ non-stationary

Proof: Random Walk Autocovariance

Claim: $\text{Cov}(Y_t, Y_{t-k}) = (t-k)\sigma^2$ for $k \leq t$

Proof: Using $Y_t = \sum_{i=1}^{t} \varepsilon_i$ and $Y_{t-k} = \sum_{i=1}^{t-k} \varepsilon_i$:

$$\text{Cov}(Y_t, Y_{t-k}) = \text{Cov}\left(\sum_{i=1}^{t} \varepsilon_i, \sum_{j=1}^{t-k} \varepsilon_j\right)$$

$$= \sum_{i=1}^{t} \sum_{j=1}^{t-k} \text{Cov}(\varepsilon_i, \varepsilon_j) = \sum_{i=1}^{t-k} \text{Var}(\varepsilon_i) = \boxed{(t-k)\sigma^2}$$

Only terms with $i = j$ survive (when $i \leq t-k$).

ACF:

$$\rho(k) = \frac{\text{Cov}(Y_t, Y_{t-k})}{\sqrt{\text{Var}(Y_t)\text{Var}(Y_{t-k})}} = \frac{(t-k)\sigma^2}{\sqrt{t\sigma^2 \cdot (t-k)\sigma^2}} = \sqrt{\frac{t-k}{t}}$$

Random Walk with Drift

Definition 2 (Random Walk with Drift)

A random walk with drift includes a constant term:

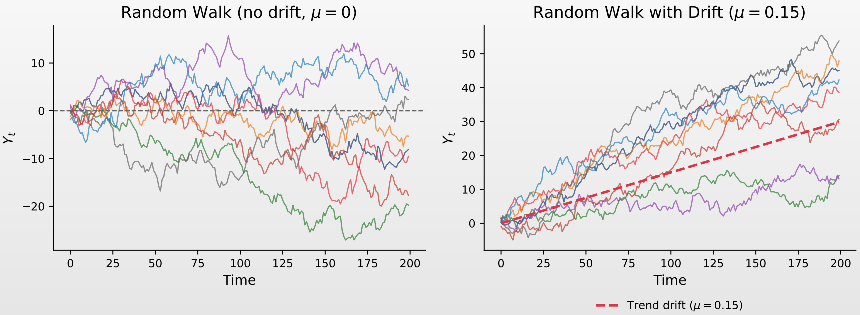
$$Y_t = \mu + Y_{t-1} + \varepsilon_t$$

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 μ creates an upward or downward trend
- Still non-stationary despite having a “trend”

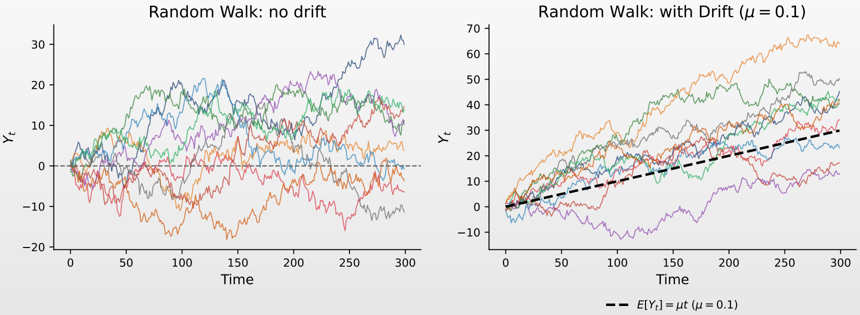
Random Walk with Drift: Visual Illustration



Comparison

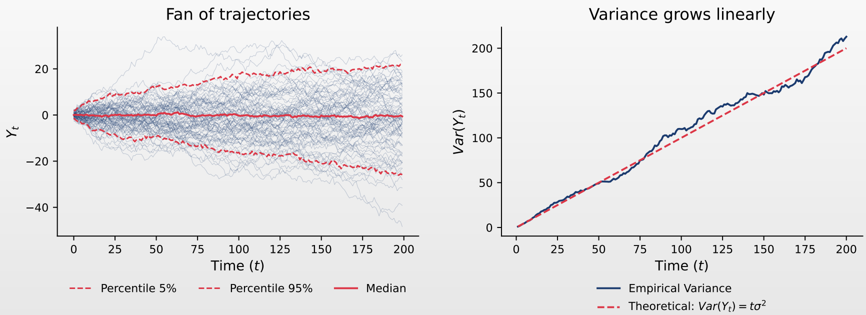
Without drift (blue): wanders around zero with no direction. With drift $\mu > 0$ (red): systematic upward trend. Both are non-stationary — drift adds deterministic trend to stochastic wandering.

Simulating Random Walks



- **Left**: Pure random walks – no drift, wander unpredictably
- **Right**: Random walks with drift – upward trend on average
- Each path is unique; uncertainty grows over time

Variance Growth: Why Random Walks Are Non-Stationary



- **Left**: Fan of paths shows uncertainty growing over time
- **Right**: Variance grows linearly: $\text{Var}(Y_t) = t\sigma^2$
- This violates stationarity (variance should be constant)

Integrated Processes

Definition 3 (Integrated Process of Order d)

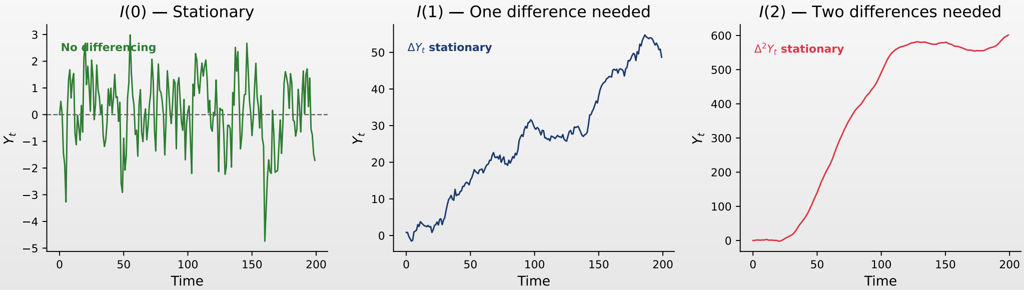
A time series $\{Y_t\}$ is **integrated of order** d, written $Y_t \sim I(d)$, if:

- 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)

Integrated Process: Visual Illustration



$I(0)$: stationary. $I(1)$: one difference needed. $I(2)$: two differences needed.

TSA_ch3_def_integrated

The Difference Operator

Definition 4 (First Difference)

The **first difference operator** Δ is defined as: $\Delta Y_t = Y_t - Y_{t-1} = (1 - L)Y_t$, where L is the lag operator ($LY_t = Y_{t-1}$).

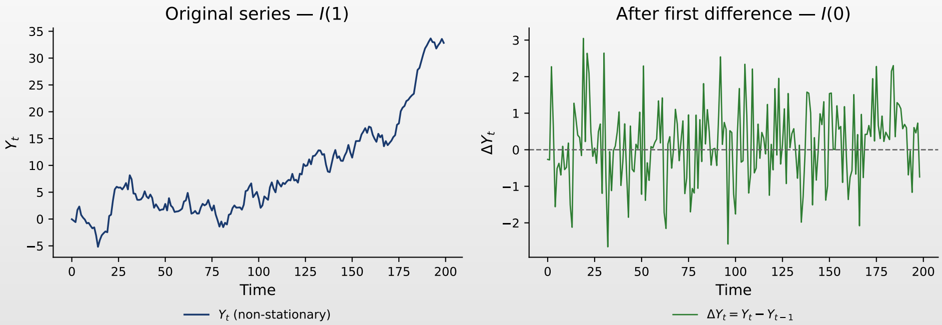
Higher-Order Differences

- 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).

First Difference: Visual Illustration



Left: non-stationary series. Right: after first difference, the series becomes stationary.

TSA_ch3_def_difference

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

Proof: Differencing Induces Stationarity

Claim: If $Y_t \sim I(1)$, then $\Delta Y_t = Y_t - Y_{t-1}$ is stationary.

Proof for Random Walk with Drift: $Y_t = \mu + Y_{t-1} + \varepsilon_t$
The first difference is:

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

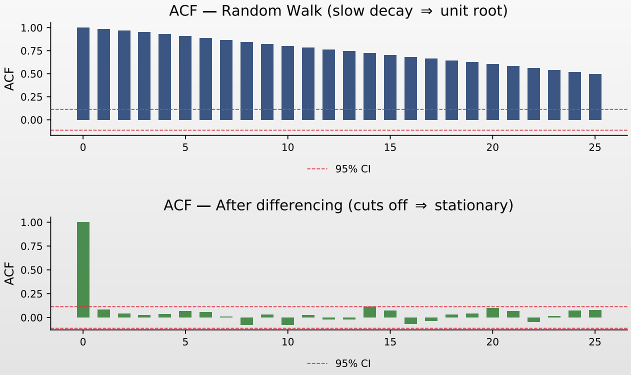
Check stationarity conditions:

1. **Mean:** $\mathbb{E}[\Delta Y_t] = \mu$ (constant, does not depend on t) ✓
2. **Variance:** $\text{Var}(\Delta Y_t) = \text{Var}(\varepsilon_t) = \sigma^2$ (constant) ✓
3. **Autocovariance:** $\text{Cov}(\Delta Y_t, \Delta Y_{t-k}) = \text{Cov}(\varepsilon_t, \varepsilon_{t-k}) = 0$ for $k \neq 0$ ✓

General Principle

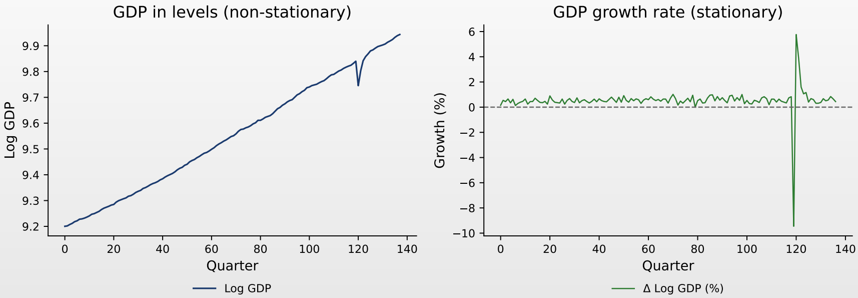
Differencing removes the "memory" that causes variance to accumulate. For $I(d)$ processes, d differences are needed.

ACF Diagnostic: Detecting Non-Stationarity



- **Top**: Random walk ACF decays very slowly $\Rightarrow$ unit root
- **Bottom**: After differencing, ACF cuts off $\Rightarrow$ stationary

Differencing in Practice: GDP Example



Transformation

Left: GDP in levels with clear upward trend (non-stationary). **Right**: GDP growth rate $\Delta \log(GDP_t)$ fluctuates around constant mean (stationary). One difference removes the stochastic trend.

Overdifferencing

Warning: Overdifferencing

Differencing more than necessary introduces problems:

- Creates artificial negative autocorrelation
  - ACF shows spurious patterns
- Inflates variance
  - Reduces forecast accuracy
- Loses information
  - Cannot recover original level

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))

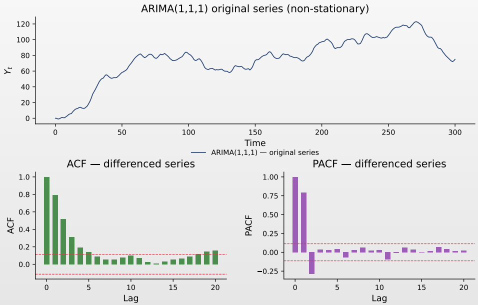
A time series $\{Y_t\}$ follows an **ARIMA(p,d,q)** process if:

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

where:

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

ARIMA: Visual Illustration



Interpretation

Top: original ARIMA series (non-stationary). Bottom: after differencing d times, ACF/PACF reveal the AR and MA orders for the stationary component.

ARIMA Components

AR(p)	I(d)	MA(q)
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

ARI(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: $\mathbb{E}[\Delta Y_t] = \frac{c}{1-\phi_1-\cdots-\phi_p}$
  - Linear trend in levels
- For $d = 2$: c affects the **curvature**
  - Quadratic trend in levels
- Often $c = 0$ is assumed when $d \geq 1$
  - No deterministic trend component

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

Dickey-Fuller Test Variants

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 AR dynamics beyond AR(1) exist, DF residuals will be autocorrelated.

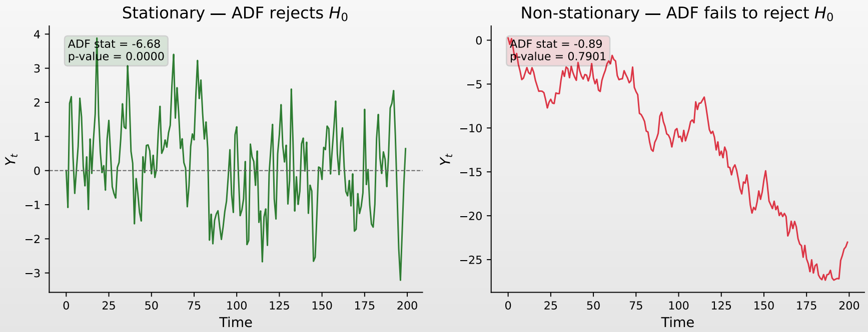
Definition 6 (ADF Test)

Add lagged differences: $\Delta Y_t = \alpha + \beta t + \gamma Y_{t-1} + \sum_{j=1}^{k} \delta_j \Delta Y_{t-j} + \varepsilon_t$
Test $H_0 : \gamma = 0$ using ADF critical values.

Choosing Lag Length k

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

ADF Test: Visual Illustration



Left: stationary series – ADF rejects unit root. Right: non-stationary – ADF fails to reject.

TSA_ch3_def_adf

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

Decision Rule

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

The Phillips-Perron (PP) Test

Motivation

Like ADF, tests H_0: Unit root vs H_1: Stationary, but uses a **non-parametric correction** for serial correlation instead of adding lagged differences.

Test Statistic

The PP test modifies the DF t-statistic:

$$Z_t = t_{\hat{\gamma}} \cdot \sqrt{\frac{\hat{\sigma}^2}{\hat{\lambda}^2}} - \frac{T(\hat{\lambda}^2 - \hat{\sigma}^2)(se(\hat{\gamma}))}{2\hat{\lambda}^2 \cdot s}$$

where $\hat{\lambda}^2$ is a consistent estimate of the long-run variance using Newey-West.

Advantages over ADF

- Robust to heteroskedasticity and serial correlation
- No need to select lag length (uses bandwidth instead)

The KPSS Test

Reversed Hypotheses

Unlike ADF: H_0: Stationary vs H_1: Unit root

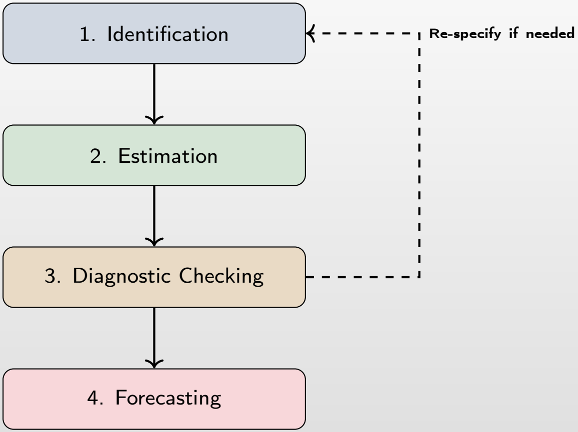
KPSS Procedure

Decompose: $Y_t = \xi t + r_t + \varepsilon_t$ where $r_t = r_{t-1} + u_t$. Test whether $\text{Var}(u_t) = 0$.

Complementary Use with ADF

- ADF rejects, KPSS doesn't $\Rightarrow$ 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 ΔY_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:

- $\text{AIC} = -2\ln(L) + 2k; \quad \text{BIC} = -2\ln(L) + k\ln(n)$

Lower is better. BIC penalizes complexity more.

Auto-ARIMA Algorithms

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!

Estimation Methods

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

Parameter Constraints

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

Residual Analysis

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.

Point Forecasts

Minimum MSE Forecast

The optimal h-step ahead forecast is the conditional expectation: $\hat{Y}_{T+h|T} = \mathbb{E}[Y_{T+h}|Y_T, Y_{T-1}, \ldots]$

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 ε_{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 ψ_j are MA(∞) 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 $h \to \infty$. Intervals widen over time!

Long-Run Forecasts for ARIMA

Behavior as $h \to \infty$

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.

Rolling Forecasting: Concept

What is Rolling Forecasting?

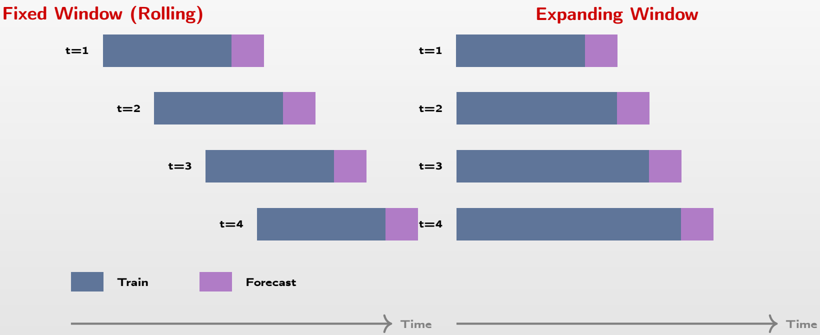
A technique to evaluate forecast accuracy out-of-sample:

1. Fix a **training window** of size w
2. Estimate model on observations $t = 1, \ldots, w$
3. Forecast h steps ahead: $\hat{Y}_{w+h|w}$
4. **Roll** the window forward by one period
5. Repeat until end of sample

Why Rolling Forecasts?

- Mimics real-time forecasting scenario
- Provides multiple forecast errors for evaluation
- Avoids overfitting to full sample

Fixed vs Expanding Window



Comparison

- **Fixed**: Window slides forward, constant size — adapts to regime changes
- **Expanding**: Window grows over time — uses all historical data

1-Step vs Multi-Step Forecasting

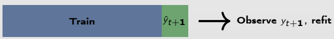
1-Step Ahead (Recursive)

- Forecast only next period
  - Refit model after each step
  - Use actual value once revealed
- Most accurate for short horizons

Multi-Step (Direct)

- Forecast multiple periods ahead
  - No refit between steps
  - Uses forecasted values as inputs
- Uncertainty compounds over horizon

1-Step Ahead



Multi-Step (h=3)

Rolling Forecast: Step-by-Step Example

Setup: ARIMA(1,1,0) with $\phi_1 = 0.6$

Model: $\Delta Y_t = \phi_1 \Delta Y_{t-1} + \varepsilon_t$ where $\Delta Y_t = Y_t - Y_{t-1}$

Given Data at Time T

$Y_{T-2} = 100, \quad Y_{T-1} = 103, \quad Y_T = 108 \quad \Rightarrow \quad \Delta Y_{T-1} = 3, \quad \Delta Y_T = 5$

1-Step Ahead Point Forecast

$$\hat{\Delta Y}_{T+1|T} = \phi_1 \cdot \Delta Y_T = 0.6 \times 5 = 3$$
$$\hat{Y}_{T+1|T} = Y_T + \hat{\Delta Y}_{T+1|T} = 108 + 3 = \boxed{111}$$

Multi-Step Point Forecasts

2-Step Ahead Forecast

$$\Delta\hat{Y}_{T+2|T} = \phi_1 \cdot \Delta\hat{Y}_{T+1|T} = 0.6 \times 3 = 1.8$$
$$\hat{Y}_{T+2|T} = \hat{Y}_{T+1|T} + \Delta\hat{Y}_{T+2|T} = 111 + 1.8 = \boxed{112.8}$$

General Formula for *h*-Step Forecast (ARIMA(1,1,0))

$$\Delta\hat{Y}_{T+h|T} = \phi_1^h \cdot \Delta Y_T$$
$$\hat{Y}_{T+h|T} = Y_T + \Delta Y_T \cdot \frac{\phi_1(1-\phi_1^h)}{1-\phi_1}$$

Numerical: 3-Step Forecast

$$\hat{Y}_{T+3|T} = 108 + 5 \times \frac{0.6(1-0.6^3)}{1-0.6} = 108 + 5 \times 1.092 = \boxed{113.46}$$

Confidence Intervals: Formulas

Forecast Error Variance

For ARIMA(1,1,0), the *h*-step forecast error variance:

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

where $\psi_j = \phi_1^{j-1}(1 + \phi_1 + \cdots + \phi_1^{j-1}) = \phi_1^{j-1} \cdot \frac{1-\phi_1^j}{1-\phi_1}$

$(1-\alpha)$% Confidence Interval

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

For 95% CI: $z_{0.025} = 1.96$

Confidence Interval: Numerical Example

Given: $\sigma^2 = 4$, $\phi_1 = 0.6$, $\hat{Y}_{T+1|T} = 111$

1-Step Ahead CI

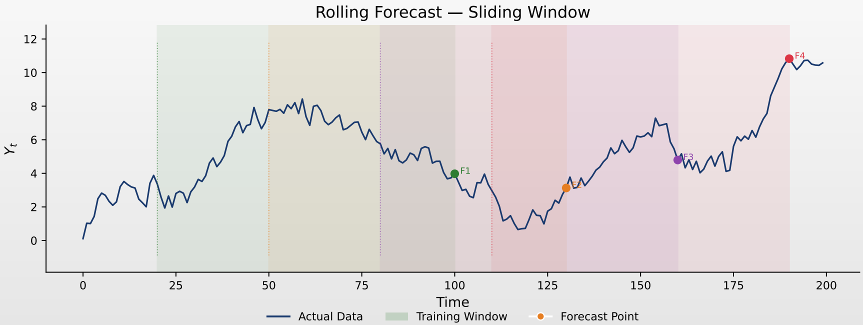
$$\text{Var}(e_{T+\mathbf{1}|T}) = \sigma^{\mathbf{2}} = 4$$
$$95\%\ \text{CI} = 111 \pm 1.96 \times \sqrt{4} = 111 \pm 3.92 = \boxed{[107.08,\ 114.92]}$$

2-Step Ahead CI (for $\hat{Y}_{T+2|T} = 112.8$)

$$\psi_{\mathbf{1}} = 1 + \phi_{\mathbf{1}} = 1.6, \quad \text{Var}(e_{T+\mathbf{2}|T}) = 4(1 + 1.6^{\mathbf{2}}) = 14.24$$
$$95\%\ \text{CI} = 112.8 \pm 1.96 \times \sqrt{14.24} = 112.8 \pm 7.40 = \boxed{[105.40,\ 120.20]}$$

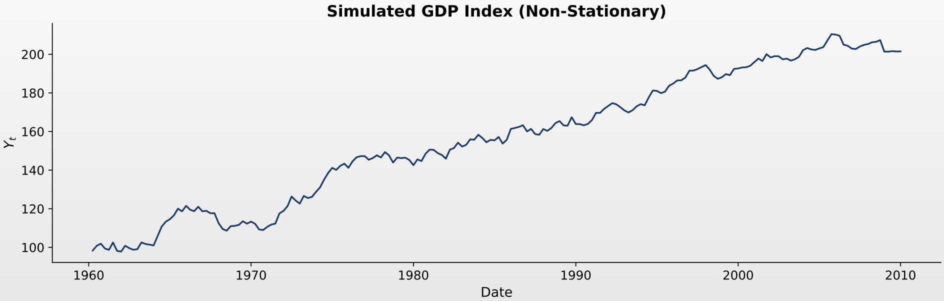
Note: CI widens as horizon increases!

Rolling Window Illustration



- Each window produces a 1-step ahead forecast
- Compare forecasts to actuals to compute RMSE, MAE
- Rolling window keeps model estimation up-to-date

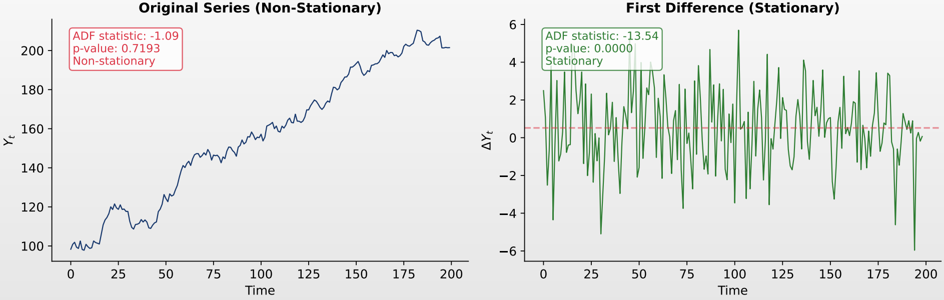
Case Study: US Real GDP (FRED)



Data: FRED GDPC1 (1960Q1–2024Q3)

Quarterly Real GDP, seasonally adjusted, billions of chained 2017 dollars. Non-stationary series with upward trend $\Rightarrow$ requires differencing.

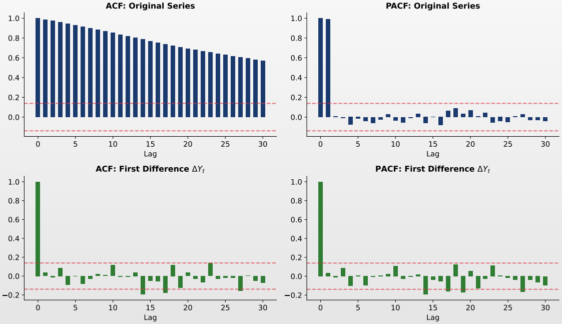
Step 1: ADF Test for Stationarity



ADF Test Results

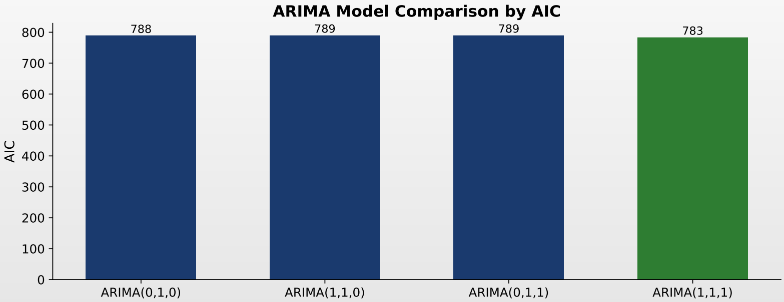
Original series: Large p-value $\Rightarrow$ fail to reject H_0 (unit root present). **First difference**: p-value $< 0.01 \Rightarrow$ reject $H_0 \Rightarrow d = 1$ is sufficient.

Step 2: ACF/PACF Before and After Differencing



Top: Slow ACF decay (non-stationary) | Bottom: After differencing, low-order ARMA

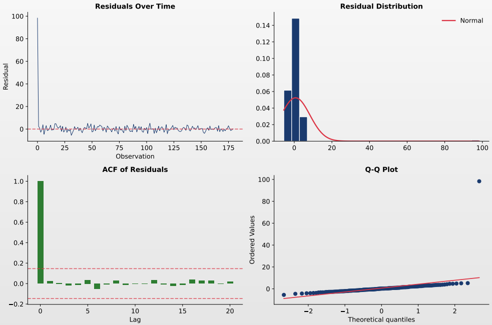
Step 3: ARIMA Model Comparison



Model Selection

Compare ARIMA(0,1,0), ARIMA(1,1,0), ARIMA(0,1,1), ARIMA(1,1,1). The model with lowest AIC is selected.

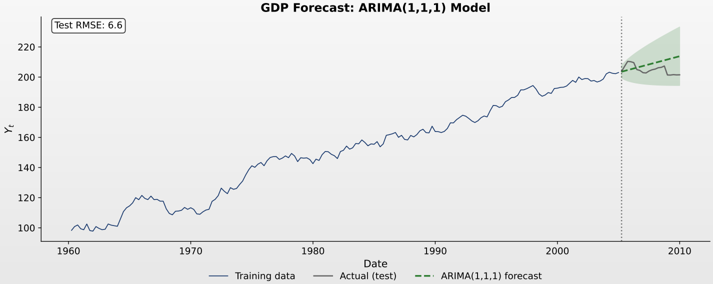
Step 4: Diagnostic Checking



ARIMA(1,1,1) Diagnostics

ACF: no autocorrelation ✓ Q-Q: **non-normal** (COVID-19 outlier) JB test: $p < 0.001$

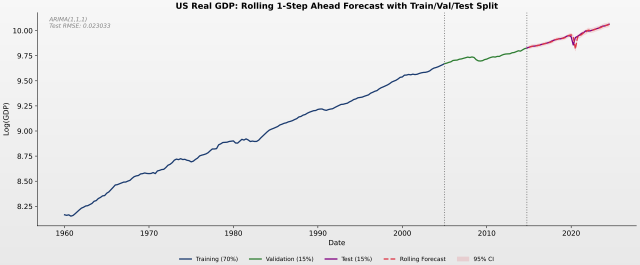
Step 5: Out-of-Sample Forecasting



Train/Val/Test Split (70%/15%/15%)

Train 70% (blue): Estimation | **Val 15%** (green): Tuning | **Test 15%** (purple): Evaluation with 95% CI

Step 6: Rolling Forecast with Train/Val/Test



Rolling 1-Step Ahead Forecast (Expanding Window, 95% CI)

Train 70% $\rightarrow$ Val 15% $\rightarrow$ Test 15% | Expanding window refits model at each step

Key Takeaways

Main Points

1. **Non-stationarity** common in economic data – use differencing to transform I(d) to I(0)
2. **ARIMA(p,d,q)** combines differencing with ARMA modeling
3. **Unit root tests** (ADF, KPSS) help determine integration order d
4. **Box-Jenkins**: Identify $\rightarrow$ Estimate $\rightarrow$ Diagnose $\rightarrow$ Forecast
5. **Forecasts** for I(1) series have growing uncertainty over horizon

Next: Chapter 4 extends ARIMA to handle seasonality (SARIMA)

Quiz Question 1

Question

A time series Y_t follows a random walk: $Y_t = Y_{t-1} + \varepsilon_t$. What is $\text{Var}(Y_t)$?

(A) σ^2 (constant)

(B) $t \cdot \sigma^2$ (grows linearly with time)

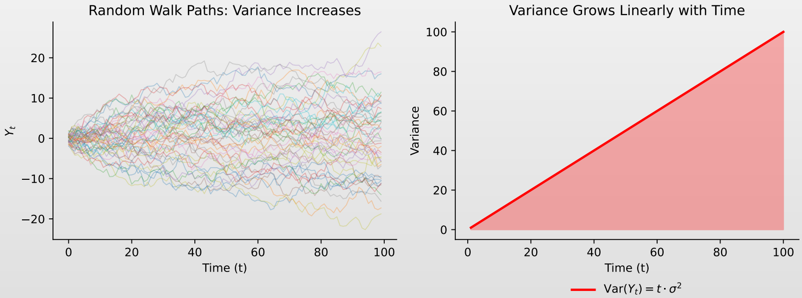
(C) σ^2/t (decreases with time)

(D) σ^{2t} (grows exponentially)

Quiz Question 1: Answer

Correct Answer: (B) $\mathrm{Var}(Y_t) = t \cdot \sigma^2$

Random walk variance grows linearly with time — this is why random walks are non-stationary.



Quiz Question 2

Question

If a series Y_t is I(2), how many times must you difference it to achieve stationarity?

(A) 0 times (already stationary)

(B) 1 time

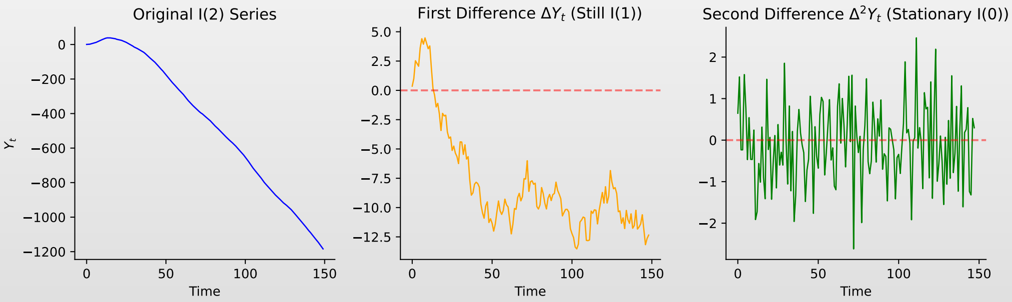
(C) 2 times

(D) Cannot be made stationary by differencing

Quiz Question 2: Answer

Correct Answer: (C) 2 times

I(d) means "integrated of order d" — requires d differences for stationarity.



Quiz Question 3

Question

You run an ADF test and get a test statistic of -2.1 with critical values: -3.43 (1%), -2.86 (5%), -2.57 (10%). What do you conclude?

(A) Reject H_0: series is stationary at all levels

(B) Reject H_0: series is stationary at 10% level only

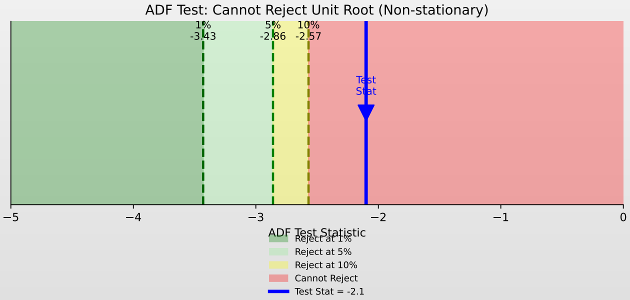
(C) Fail to reject H_0: series likely has a unit root

(D) The test is inconclusive

Quiz Question 3: Answer

Correct Answer: (C) Fail to reject H_0: series has unit root

Test stat $-2.1 > -2.57$ (10% CV) $\Rightarrow$ Cannot reject at any level. Consider differencing.



Quiz Question 4

Question

For an ARIMA(1,1,0) model, what is the ACF pattern of the **differenced** series ΔY_t?

(A) Cuts off after lag 1

(B) Decays exponentially

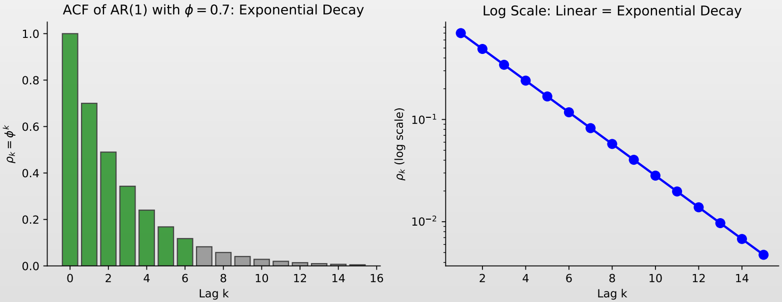
(C) Alternates in sign

(D) Is zero at all lags

Quiz Question 4: Answer

Correct Answer: (B) Decays exponentially

ARIMA(1,1,0) $\Rightarrow \Delta Y_t$ follows AR(1) with ACF $\rho_k = \phi_1^k$ (geometric decay).



Quiz Question 5

Question

What happens to ARIMA forecast confidence intervals as the horizon h increases for an I(1) series?

(A) They stay constant

(B) They narrow (more precision)

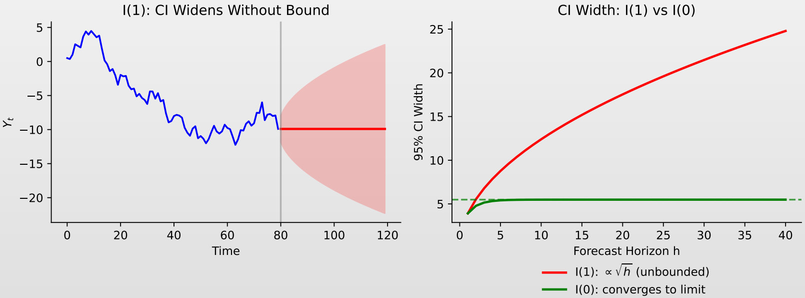
(C) They widen without bound

(D) They widen but converge to a limit

Quiz Question 5: Answer

Correct Answer: (C) They widen without bound

For I(1): CI width $\propto \sqrt{h}$ (unbounded). For I(0): CIs converge to a limit.



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.