



Time Series Analysis and Forecasting

## Chapter 4: SARIMA Models

Seasonal Time Series

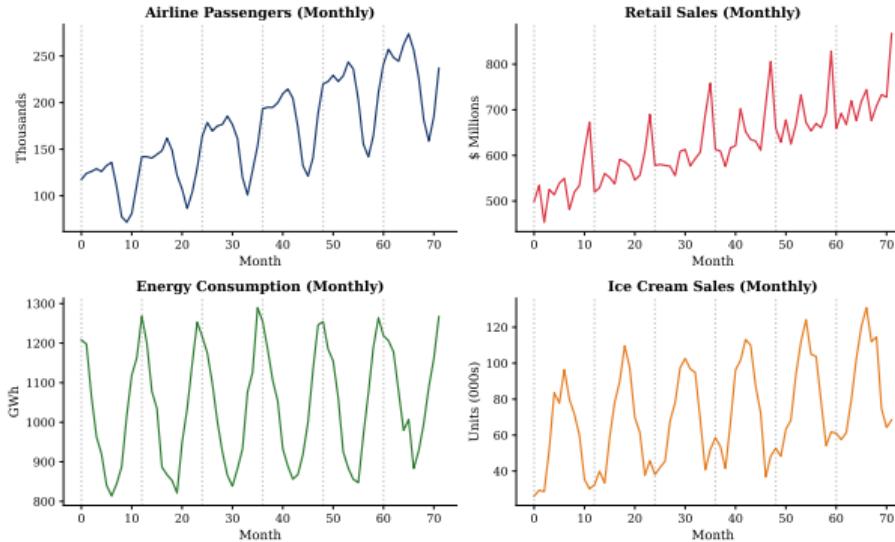


# Outline

- 1 Seasonality in Time Series
- 2 Seasonal Differencing
- 3 The SARIMA Model
- 4 Seasonal ACF and PACF Patterns
- 5 Estimation and Diagnostics
- 6 Forecasting with SARIMA
- 7 Real Data Application: Airline Passengers
- 8 Case Study: Airline Passengers
- 9 Summary

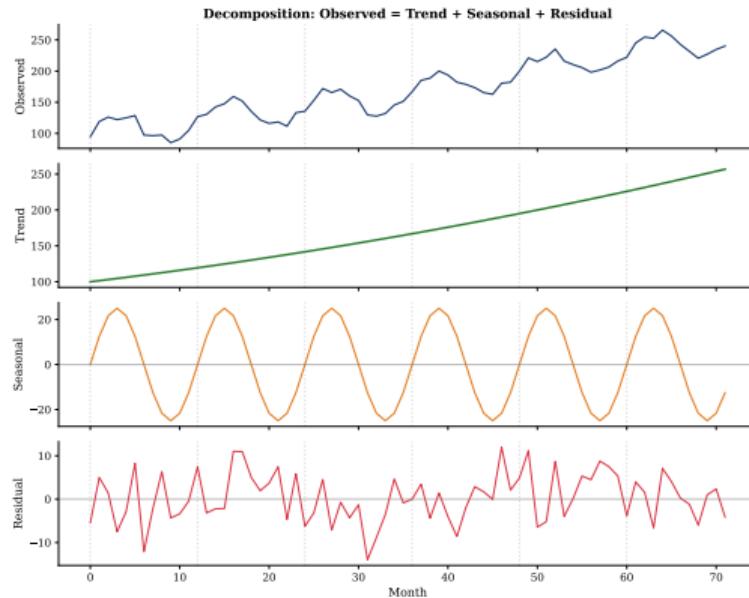
# Motivating Example: Seasonality Is Everywhere

Seasonal Patterns Are Everywhere!



- Retail sales exhibit clear **annual patterns**: December peaks, January troughs
- Standard ARIMA models cannot capture these **repeating seasonal cycles**
- Ignoring seasonality leads to systematic forecast errors

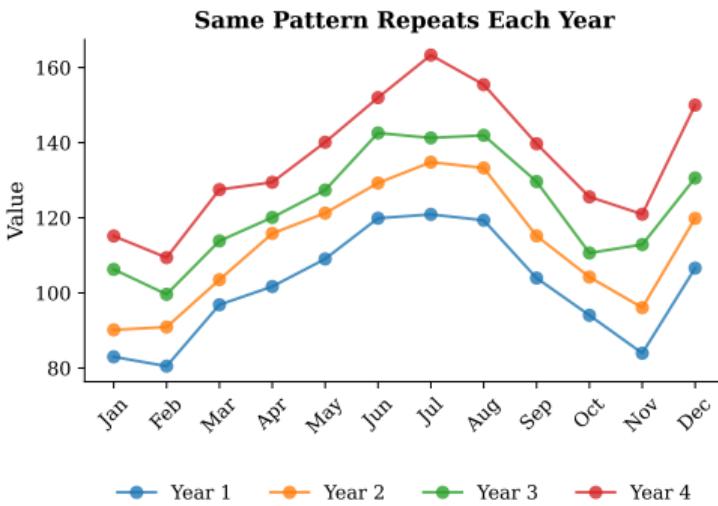
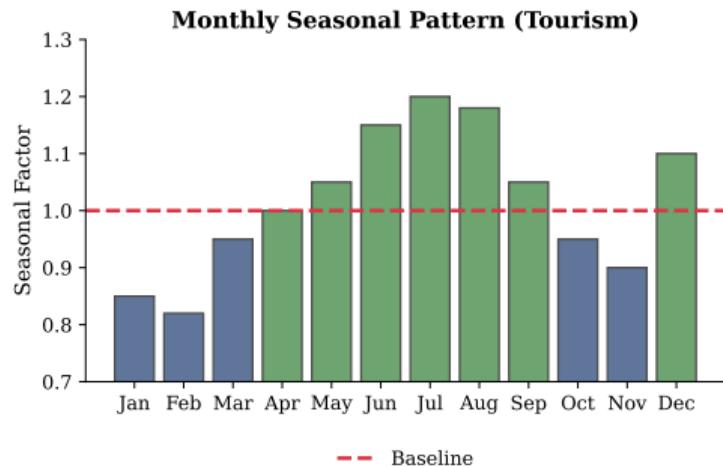
# Understanding Seasonal Components



- Seasonal time series = **Trend + Seasonal Pattern + Residuals**
- Decomposition helps visualize each component separately
- SARIMA models capture both trend dynamics and seasonal behavior

# Real-World Application: Monthly Patterns

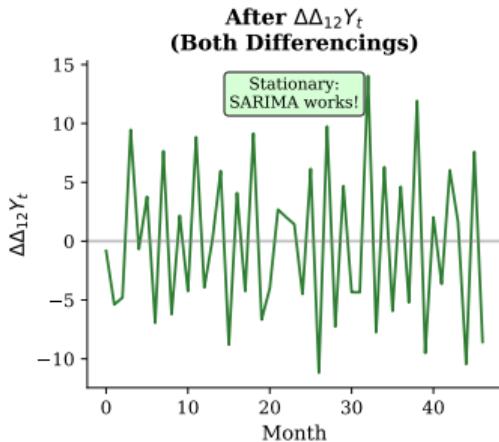
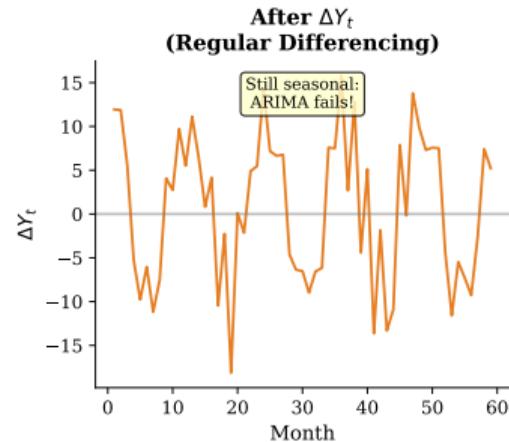
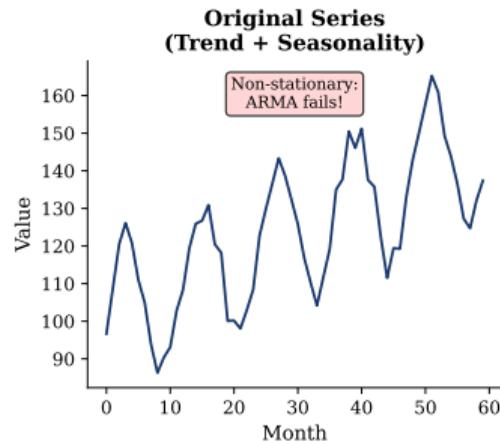
## Understanding Seasonal Patterns



- Energy demand shows strong **monthly seasonality** (heating/cooling cycles)
- Pattern repeats predictably each year with slight variations
- Utility companies use SARIMA forecasts for capacity planning

# Why Do We Need SARIMA?

## Why ARIMA Is Not Enough for Seasonal Data



- Left: Seasonal ACF shows spikes at lags 12, 24, 36... (annual pattern)
- Right: ARIMA residuals still show seasonal autocorrelation — model is incomplete
- SARIMA adds **seasonal AR and MA terms** to capture these patterns

# What We'll Learn Today

## Concepts

- Identifying seasonal patterns
- Seasonal differencing operator
- SARIMA( $p, d, q)(P, D, Q)_s$  notation
- The famous “Airline Model”
- Model selection for seasonal data

## Skills

- Diagnose seasonality from ACF/PACF
- Determine seasonal period  $s$
- Choose  $(P, D, Q)$  seasonal orders
- Implement SARIMA in Python/R
- Forecast seasonal time series

## Key Insight

SARIMA = ARIMA applied at **two frequencies**: the regular (short-term) and seasonal (long-term) levels

# What is Seasonality?

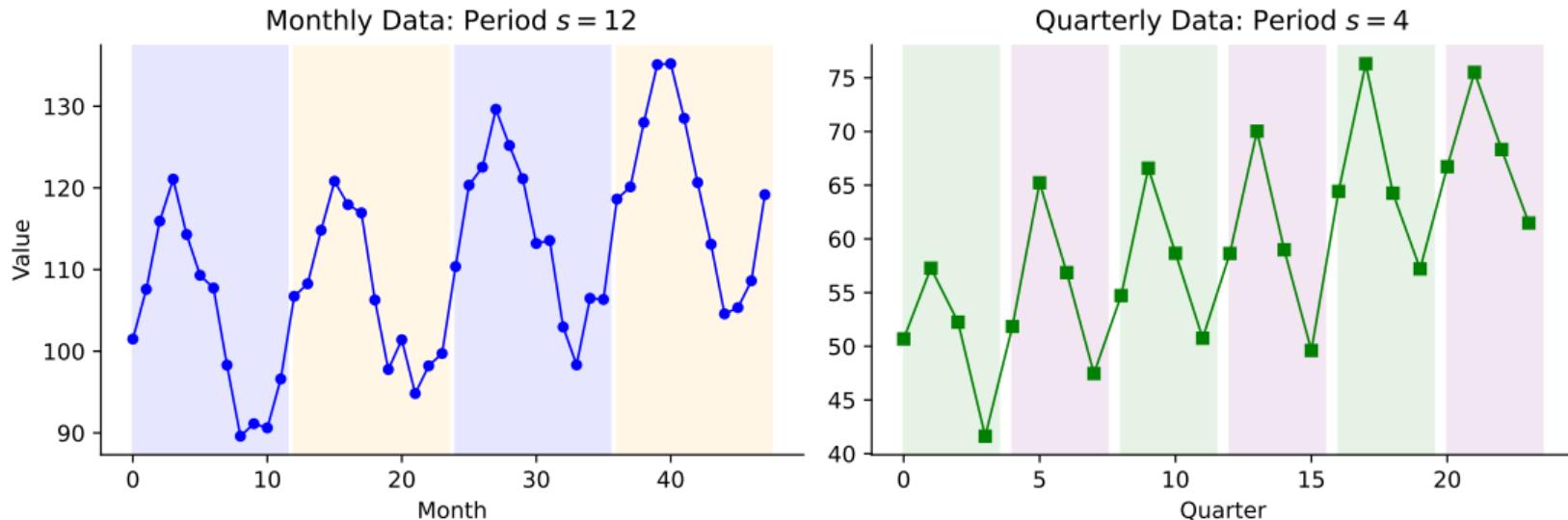
## Definition 1 (Seasonality)

A time series exhibits **seasonality** when it shows regular, periodic fluctuations that repeat over a fixed period  $s$  (the seasonal period).

## Common Seasonal Periods

- Monthly data:  $s = 12$  (annual cycle)
- Quarterly data:  $s = 4$  (annual cycle)
- Weekly data:  $s = 52$  (annual) or  $s = 7$  (weekly pattern)
- Daily data:  $s = 7$  (weekly pattern)

## Seasonality: Visual Illustration



### Seasonal Periods

Left: Monthly data with  $s = 12$  (annual cycle). Right: Quarterly data with  $s = 4$ . The pattern repeats every  $s$  periods — this regularity is exploited by SARIMA models.

## Examples of Seasonal Data

### Economic Series

- Retail sales (holiday peaks)
- Tourism (summer/winter)
- Agricultural production
- Energy consumption
- Employment (hiring cycles)

### Other Domains

- Weather/temperature
- Website traffic
- Hospital admissions
- Transportation usage
- Electricity demand

## Why It Matters

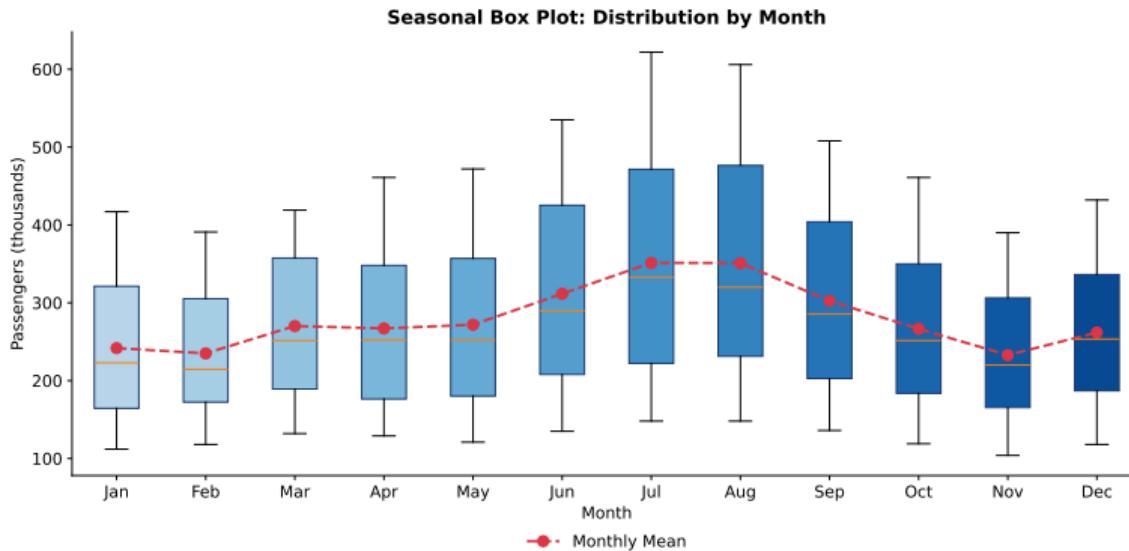
Ignoring seasonality leads to biased forecasts and invalid inference!

## Example: Airline Passengers Data



- Monthly international airline passengers (1949–1960)
- Clear **upward trend** and **growing seasonal amplitude**
- Summer peaks reflect vacation travel patterns

# Visualizing Seasonal Patterns



- Box plot reveals consistent seasonal pattern across years
- July–August show highest passenger counts (summer travel)
- November–February show lowest counts (winter months)

# Deterministic vs Stochastic Seasonality

## Deterministic Seasonality

Fixed seasonal pattern:  $Y_t = \sum_{j=1}^s \gamma_j D_{jt} + \varepsilon_t$  where  $D_{jt}$  are seasonal dummies.

### Properties:

- Pattern constant over time
- Removed by regression

## Stochastic Seasonality

Evolving seasonal pattern:  $\Delta_s Y_t = Y_t - Y_{t-s}$  exhibits dependence structure.

### Properties:

- Pattern evolves over time
- Requires seasonal differencing

# Detecting Seasonality

## Visual Methods

- Time series plot – look for repeating patterns
- Seasonal subseries plot – compare same seasons across years
- ACF plot – spikes at seasonal lags ( $s, 2s, 3s, \dots$ )

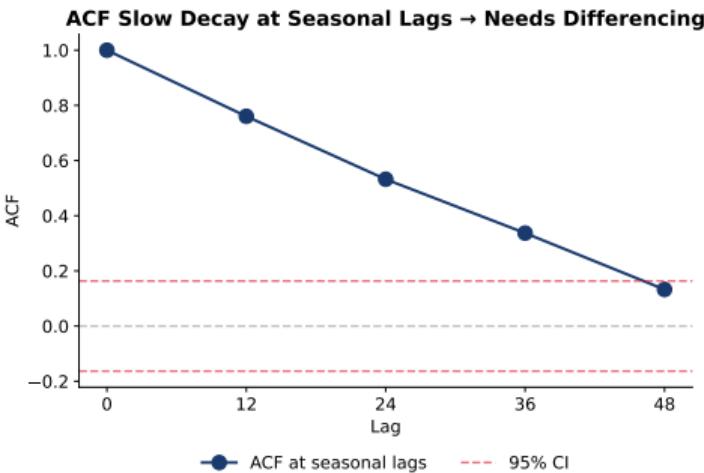
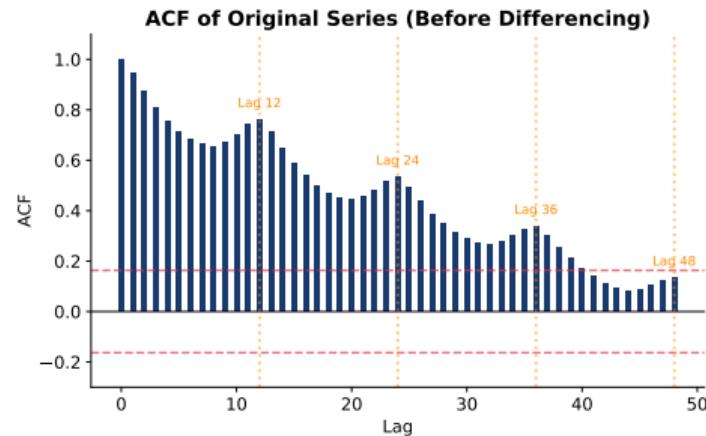
## Statistical Tests

- Seasonal unit root tests (HEGY, CH, OCSB)
- F-test for seasonal dummies
- Kruskal-Wallis test (nonparametric)

## ACF Signature

Strong seasonality: ACF shows significant spikes at lags  $s, 2s, 3s, \dots$

# ACF Reveals Seasonal Structure



- Slow decay at all lags indicates non-stationarity (trend)
- Spikes at lags 12, 24, 36 confirm seasonal pattern ( $s = 12$ )
- ACF at seasonal lags shows slow decay  $\Rightarrow$  needs seasonal differencing

## F-test for Seasonal Dummies: Intuition

### What Does This Test Do?

Tests whether the **mean values differ significantly across seasons.**

- If January mean  $\neq$  February mean  $\neq \dots \neq$  December mean  $\Rightarrow$  seasonality
- Compares a model WITH seasonal dummies vs. a model WITHOUT

### The Models Being Compared

**Restricted:**  $Y_t = \alpha + \varepsilon_t$     **Unrestricted:**  $Y_t = \alpha + \sum_{j=1}^{s-1} \gamma_j D_{jt} + \varepsilon_t$   
where  $D_{jt} = 1$  if observation  $t$  is in season  $j$ , 0 otherwise.

### Key Idea

If adding seasonal dummies **significantly reduces** prediction errors, then seasonality is present.

## F-test for Seasonal Dummies: Formula and Example

### F-statistic Formula

$$F = \frac{(SSR_R - SSR_U)/(s - 1)}{SSR_U/(n - s)} \sim F_{s-1, n-s}$$

- $SSR_R$  = Sum of Squared Residuals from restricted model (no dummies)
- $SSR_U$  = Sum of Squared Residuals from unrestricted model (with dummies)
- $s - 1$  = number of restrictions (monthly: 11, quarterly: 3)

### Numerical Example (Monthly Data, $n=120$ )

$SSR_R = 15000$ ,  $SSR_U = 8500$ ,  $s = 12$

$$F = \frac{(15000 - 8500)/11}{8500/108} = \frac{590.9}{78.7} = 7.51$$

Critical value  $F_{0.05, 11, 108} \approx 1.87$ . Since  $7.51 > 1.87$ : **Reject  $H_0$**   $\Rightarrow$  Seasonality present!

# Kruskal-Wallis Test: Intuition

## What Does This Test Do?

A **nonparametric** test that checks if observations from different seasons come from the same distribution.

- Works by **ranking** all observations from smallest to largest
- Checks if ranks are evenly distributed across seasons
- If one season consistently has higher/lower ranks  $\Rightarrow$  seasonality

## Why Use It Instead of F-test?

- **No normality assumption** – works with any distribution
- **Robust to outliers** – extreme values don't distort results

## Limitation

Less powerful than F-test when data IS normally distributed.

# Kruskal-Wallis Test: Formula and Example

## Test Statistic

$$H = \frac{12}{N(N+1)} \sum_{j=1}^s \frac{R_j^2}{n_j} - 3(N+1) \quad \text{where } N = \text{total obs.}, n_j = \text{obs. in season } j, R_j = \text{sum of ranks.}$$

## Example: Quarterly Sales ( $n=20$ , $s=4$ )

Data ranked 1-20. Rank sums: Q1:  $R_1 = 15$ , Q2:  $R_2 = 35$ , Q3:  $R_3 = 70$ , Q4:  $R_4 = 90$

$$H = \frac{12}{20 \times 21} \left( \frac{15^2}{5} + \frac{35^2}{5} + \frac{70^2}{5} + \frac{90^2}{5} \right) - 3(21) = 12.6$$

Critical value  $\chi^2_{0.05,3} = 7.81$ . Since  $12.6 > 7.81$ : **Reject  $H_0$**   $\Rightarrow$  Seasonality!

## In Python

```
scipy.stats.kruskal(q1, q2, q3, q4)
```

# HEGY Test: What Problem Does It Solve?

## The Key Question

Given a seasonal time series, we need to know:

- ① Does it need **regular differencing** ( $(1 - L)$ )?  $\Rightarrow$  set  $d = 1$
- ② Does it need **seasonal differencing** ( $(1 - L^s)$ )?  $\Rightarrow$  set  $D = 1$

HEGY tests for **both** types of unit roots simultaneously!

## Why Not Just Use ADF?

ADF only tests for a **regular** unit root at frequency zero. Seasonal data can have unit roots at **seasonal frequencies** that ADF misses!

## HEGY Tests Multiple Frequencies

Quarterly: tests at  $0, \pi, \pm\pi/2$ . Monthly: tests at  $0, \pi, \pm\pi/6, \pm\pi/3, \pm\pi/2, \pm2\pi/3, \pm5\pi/6$ .

# HEGY Test: The Regression Formula (Quarterly)

## HEGY Auxiliary Regression

For quarterly data ( $s = 4$ ), estimate:

$$\Delta_4 y_t = \pi_1 z_{1,t-1} + \pi_2 z_{2,t-1} + \pi_3 z_{3,t-2} + \pi_4 z_{4,t-2} + \sum_{j=1}^k \phi_j \Delta_4 y_{t-j} + \varepsilon_t$$

## Transformed Variables

$$z_{1t} = (1 + L + L^2 + L^3)y_t = y_t + y_{t-1} + y_{t-2} + y_{t-3}$$

$$z_{2t} = -(1 - L + L^2 - L^3)y_t = -y_t + y_{t-1} - y_{t-2} + y_{t-3}$$

$$z_{3t} = -(1 - L^2)y_t = -y_t + y_{t-2} ; \quad z_{4t} = -(L - L^3)y_t = -y_{t-1} + y_{t-3}$$

## Hypotheses

$$H_0 : \pi_1 = 0 \text{ (freq. 0)}, H_0 : \pi_2 = 0 \text{ (freq. } \pi\text{)}, H_0 : \pi_3 = \pi_4 = 0 \text{ (freq. } \pm\pi/2\text{)}$$

## HEGY Test: Decision Rules with Examples

### HEGY Critical Values (5%, n=100, with constant)

Test	Statistic	Critical Value	If NOT rejected...
$t_1 (\pi_1 = 0)$	t-stat	-2.88	Need $d = 1$
$t_2 (\pi_2 = 0)$	t-stat	-2.88	Need $D = 1$
$F_{34} (\pi_3 = \pi_4 = 0)$	F-stat	6.57	Need $D = 1$

### Example: Quarterly GDP

Suppose HEGY gives:  $t_1 = -1.52$ ,  $t_2 = -4.21$ ,  $F_{34} = 2.15$

- $t_1 = -1.52 > -2.88$ : Cannot reject  $\Rightarrow$  need  $d = 1$
- $t_2 = -4.21 < -2.88$ : Reject  $\Rightarrow$  no unit root at  $\pi$
- $F_{34} = 2.15 < 6.57$ : Cannot reject  $\Rightarrow$  need  $D = 1$

**Conclusion:** Use SARIMA with  $d = 1, D = 1$

## Canova-Hansen Test: The Opposite of HEGY

### HEGY vs Canova-Hansen: Different Null Hypotheses!

	HEGY	Canova-Hansen
$H_0$	Seasonal unit root	<b>No</b> seasonal unit root
$H_1$	No seasonal unit root	Seasonal unit root
Reject $H_0$	Use seasonal dummies	Use $(1 - L^s)$ differencing
Don't reject	Use $(1 - L^s)$ differencing	Use seasonal dummies

### Why Does This Matter?

- HEGY: “Prove there’s NO unit root” (conservative toward differencing)
- CH: “Prove there IS a unit root” (conservative toward dummies)
- Use **both** tests for robust conclusions!

# Canova-Hansen Test: Formula

## Test Procedure

1. Regress  $y_t$  on seasonal dummies:  $y_t = \sum_{j=1}^s \gamma_j D_{jt} + u_t$
2. Compute partial sums at seasonal frequency  $\lambda_i$ :  $S_{it}^{(c)} = \sum_{j=1}^t \hat{u}_j \cos(\lambda_i j)$ ,  $S_{it}^{(s)} = \sum_{j=1}^t \hat{u}_j \sin(\lambda_i j)$

## LM Test Statistic

$$LM_i = \frac{1}{T^2 \hat{\omega}_i} \left[ \sum_{t=1}^T (S_{it}^{(c)})^2 + \sum_{t=1}^T (S_{it}^{(s)})^2 \right]$$

where  $\hat{\omega}_i$  = consistent estimate of spectral density at frequency  $\lambda_i$ .

## Decision

Reject  $H_0$  (stationarity) if  $LM >$  critical value  $\Rightarrow$  seasonal differencing needed.

## Summary: Choosing the Right Seasonality Test

Test	$H_0$	If Reject	Best For
F-test	No seasonality	Seasonality exists	Normal data
Kruskal-Wallis	No seasonal diff.	Seasonality exists	Non-normal, outliers
HEGY	Unit root exists	Use dummies	Determining $d$ , $D$
Canova-Hansen	No unit root	Use $(1 - L^s)$	Confirming stability

### Key Insight

F-test/Kruskal-Wallis: “*Is there seasonality?*”

HEGY/Canova-Hansen: “*What type?*” (deterministic vs stochastic)

# The Seasonal Difference Operator

## Definition 2 (Seasonal Difference)

The **seasonal difference operator**  $\Delta_s$  is defined as:

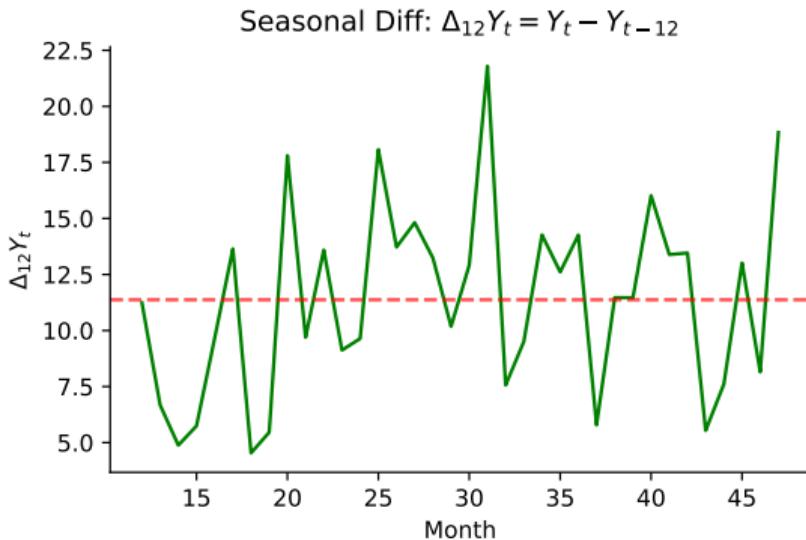
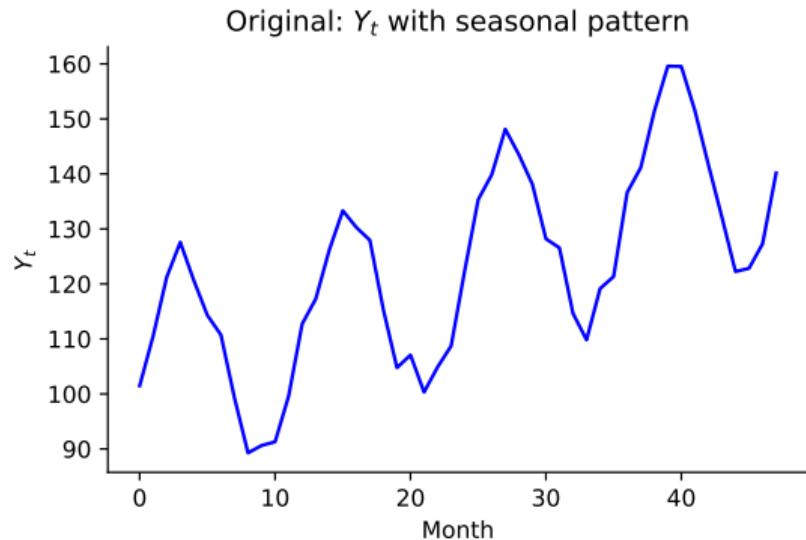
$$\Delta_s Y_t = (1 - L^s) Y_t = Y_t - Y_{t-s}$$

where  $L^s Y_t = Y_{t-s}$  is the seasonal lag operator.

## Examples

- Monthly data ( $s = 12$ ):  $\Delta_{12} Y_t = Y_t - Y_{t-12}$   
Compares each month to the same month last year
- Quarterly data ( $s = 4$ ):  $\Delta_4 Y_t = Y_t - Y_{t-4}$   
Compares each quarter to the same quarter last year

## Seasonal Difference: Visual Illustration



### Effect of Seasonal Differencing

Left: Original series with clear seasonal pattern. Right: After  $\Delta_{12} = (1 - L^{12})$ , seasonal pattern is removed.  
Year-over-year comparison eliminates seasonal effects.

## Proof: Seasonal Differencing Removes Deterministic Seasonality

**Claim:** If  $Y_t = \mu_t + \varepsilon_t$  where  $\mu_t = \mu_{t-s}$  (periodic mean), then  $\Delta_s Y_t$  removes the seasonal mean.

**Proof:** Let  $Y_t = \mu_t + \varepsilon_t$  where  $\mu_t$  has period  $s$ . Apply seasonal difference:

$$\begin{aligned}\Delta_s Y_t &= Y_t - Y_{t-s} = (\mu_t + \varepsilon_t) - (\mu_{t-s} + \varepsilon_{t-s}) \\ &= \mu_t - \mu_{t-s} + \varepsilon_t - \varepsilon_{t-s} \\ &= 0 + \varepsilon_t - \varepsilon_{t-s} \quad (\text{since } \mu_t = \mu_{t-s})\end{aligned}$$

**Properties of  $\Delta_s Y_t = \varepsilon_t - \varepsilon_{t-s}$ :**

- $\mathbb{E}[\Delta_s Y_t] = 0$  (constant mean)
- $\text{Var}(\Delta_s Y_t) = 2\sigma^2$  (constant variance)
- Autocovariance:  $\gamma(s) = -\sigma^2$ ,  $\gamma(k) = 0$  for  $k \neq 0, s$

### Result

Seasonal differencing transforms periodic seasonal pattern into MA(1) at seasonal lag.

# Combining Regular and Seasonal Differencing

## Full Differencing

For series with both trend and seasonality:

$$\Delta\Delta_s Y_t = (1 - L)(1 - L^s)Y_t$$

## Expansion

$$(1 - L)(1 - L^s)Y_t = Y_t - Y_{t-1} - Y_{t-s} + Y_{t-s-1}$$

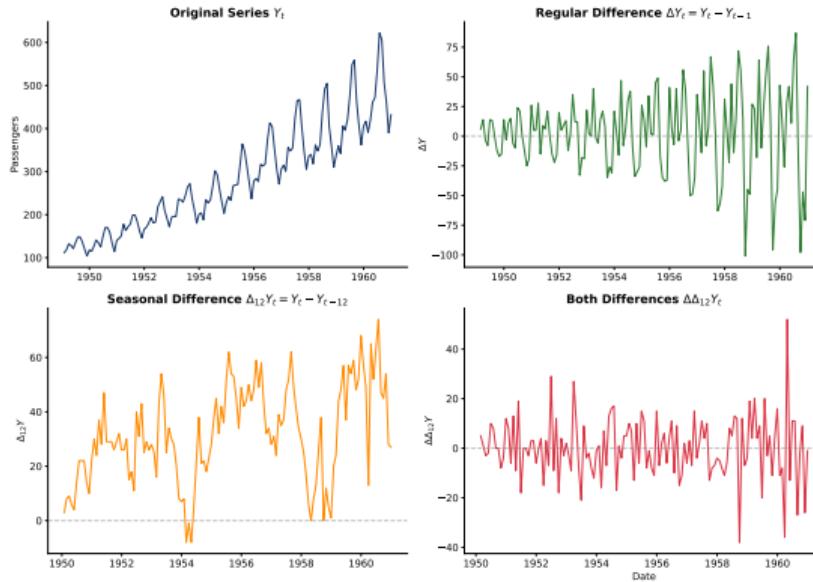
For monthly data ( $s = 12$ ):

$$\Delta\Delta_{12} Y_t = Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}$$

## Order of Differencing

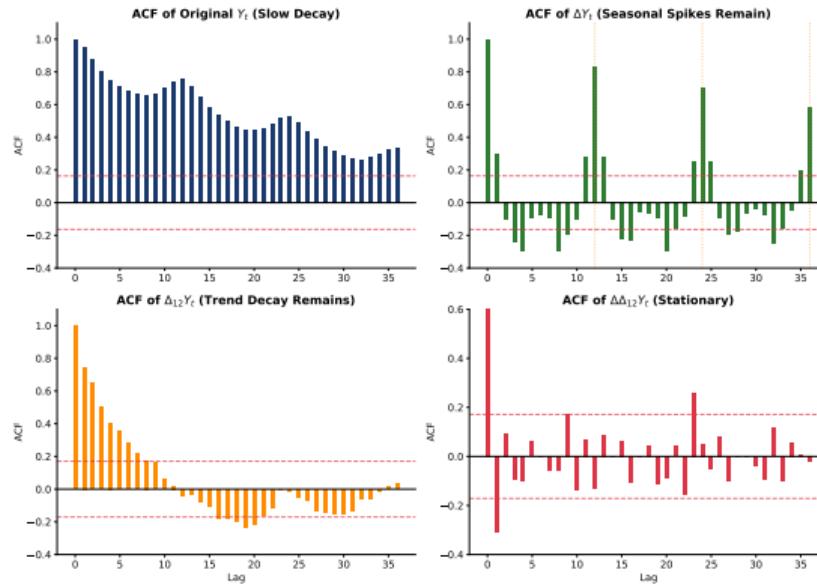
- $d$ : number of regular differences (trend removal)
- $D$ : number of seasonal differences (seasonal trend removal)

# Effect of Differencing Operations



- Regular differencing removes trend but seasonal pattern remains
- Seasonal differencing removes seasonality but trend pattern remains
- **Both differences** needed to achieve stationarity

# ACF Before and After Differencing



- Original ACF: slow decay indicates non-stationarity
- After  $\Delta$ : seasonal spikes remain at lags 12, 24, 36
- After  $\Delta_{12}$ : trend decay remains at early lags
- After  $\Delta\Delta_{12}$ : ACF cuts off  $\Rightarrow$  stationary

## Definition 3 (Seasonally Integrated Process)

A series  $Y_t$  is **seasonally integrated** of order  $(d, D)_s$ , written  $Y_t \sim I(d, D)_s$ , if:

$$(1 - L)^d (1 - L^s)^D Y_t$$

is stationary.

## Common Cases

- $I(1, 0)_{12}$ : Regular unit root only (monthly)
- $I(0, 1)_{12}$ : Seasonal unit root only
- $I(1, 1)_{12}$ : Both regular and seasonal unit roots

# SARIMA Model Definition

Definition 4 (SARIMA( $p, d, q$ )  $\times$  ( $P, D, Q$ ) $_s$ )

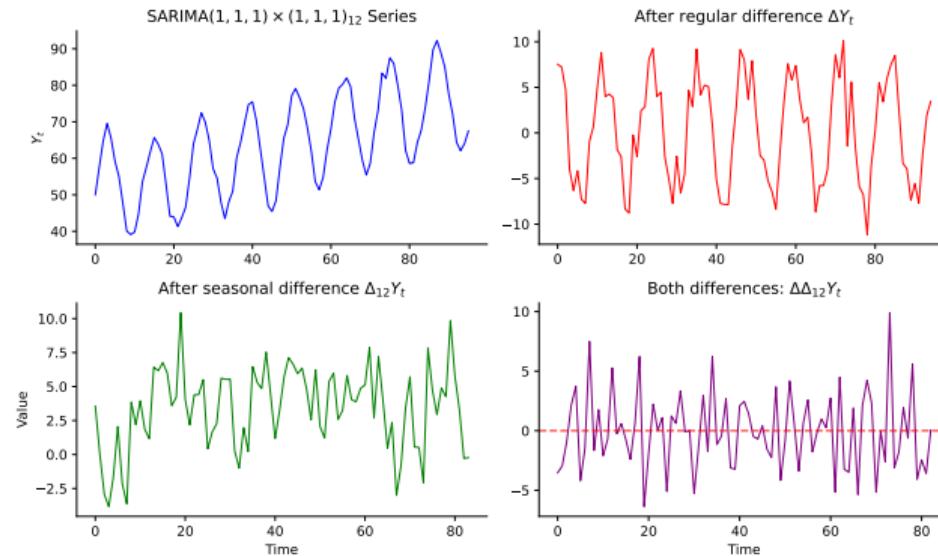
The **Seasonal ARIMA** model is:

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

## Components

- $\phi(L) = 1 - \phi_1L - \dots - \phi_pL^p$ : Non-seasonal AR
- $\Phi(L^s) = 1 - \Phi_1L^s - \dots - \Phi_PL^{Ps}$ : Seasonal AR
- $\theta(L) = 1 + \theta_1L + \dots + \theta_qL^q$ : Non-seasonal MA
- $\Theta(L^s) = 1 + \Theta_1L^s + \dots + \Theta_QL^{Qs}$ : Seasonal MA
- $(1 - L)^d$ : Regular differencing;  $(1 - L^s)^D$ : Seasonal differencing

# SARIMA: Visual Illustration



## Differencing Strategy

Progressive transformation: Original → regular difference (removes trend) → seasonal difference (removes seasonality) → both. Apply minimum differencing needed to achieve stationarity.

## Proof: Multiplicative Seasonal Structure

**Why multiplicative?** Consider SARIMA(1, 0, 0)  $\times$  (1, 0, 0)<sub>s</sub>:

$$(1 - \phi L)(1 - \Phi L^s) Y_t = \varepsilon_t$$

**Expand the product:**

$$\begin{aligned}(1 - \phi L)(1 - \Phi L^s) Y_t &= Y_t - \phi LY_t - \Phi L^s Y_t + \phi\Phi L^{s+1} Y_t \\ &= Y_t - \phi Y_{t-1} - \Phi Y_{t-s} + \phi\Phi Y_{t-s-1}\end{aligned}$$

**Result:** The model includes an **interaction term**  $\phi\Phi Y_{t-s-1}$

### Interpretation (Monthly, $s = 12$ )

$Y_t$  depends on:

- $Y_{t-1}$ : Last month (short-run dynamics)
- $Y_{t-12}$ : Same month last year (seasonal effect)
- $Y_{t-13}$ : Interaction of both effects

### Parsimony

Multiplicative form: 2 parameters ( $\phi, \Phi$ ). Additive would need 3+ parameters.

# SARIMA Notation

## Full Specification

SARIMA( $p, d, q) \times (P, D, Q)_s$  has 7 parameters to specify:

Parameter	Meaning
$p$	Non-seasonal AR order
$d$	Non-seasonal differencing order
$q$	Non-seasonal MA order
$P$	Seasonal AR order
$D$	Seasonal differencing order
$Q$	Seasonal MA order
$s$	Seasonal period

## Example

SARIMA(1, 1, 1)  $\times$  (1, 1, 1)<sub>12</sub>: Monthly data with AR(1), MA(1), seasonal AR(1), seasonal MA(1), and both regular and seasonal differencing.

## Common SARIMA Models

Airline Model:  $\text{SARIMA}(0, 1, 1) \times (0, 1, 1)_s$

$(1 - L)(1 - L^s)Y_t = (1 + \theta L)(1 + \Theta L^s)\varepsilon_t$  – Classic model (Box & Jenkins, 1970)

SARIMA(1, 0, 0)  $\times$  (1, 0, 0)<sub>s</sub>

$(1 - \phi L)(1 - \Phi L^s)Y_t = \varepsilon_t$  – Pure seasonal and non-seasonal AR

SARIMA(0, 1, 1)  $\times$  (0, 1, 0)<sub>s</sub>

$(1 - L)(1 - L^s)Y_t = (1 + \theta L)\varepsilon_t$  – Random walk + seasonal diff + MA(1)

## Why Multiplicative?

The seasonal and non-seasonal parts **multiply**:

$$\phi(L)\Phi(L^s) \quad \text{and} \quad \theta(L)\Theta(L^s)$$

Example: SARIMA(1, 0, 0)  $\times$  (1, 0, 0)<sub>12</sub>

$$(1 - \phi L)(1 - \Phi L^{12}) Y_t = \varepsilon_t$$

$$\text{Expanding: } Y_t - \phi Y_{t-1} - \Phi Y_{t-12} + \phi\Phi Y_{t-13} = \varepsilon_t$$

The cross-term  $\phi\Phi Y_{t-13}$  captures interaction!

## Interpretation

Multiplicative structure allows parsimonious modeling of complex seasonal patterns with few parameters.

## Key Insight

Seasonal models show patterns at both:

- Non-seasonal lags:  $1, 2, 3, \dots$
- Seasonal lags:  $s, 2s, 3s, \dots$

Model	ACF	PACF
SAR( $P$ )	Decays at $s, 2s, \dots$	Cuts off after $Ps$
SMA( $Q$ )	Cuts off after $Qs$	Decays at $s, 2s, \dots$
SARMA	Decays at seasonal lags	Decays at seasonal lags

## Example: Airline Model ACF/PACF

SARIMA(0, 1, 1)  $\times$  (0, 1, 1)<sub>12</sub>

After differencing  $W_t = (1 - L)(1 - L^{12})Y_t$ :

$$W_t = (1 + \theta L)(1 + \Theta L^{12})\varepsilon_t$$

### Expected ACF Pattern

- Spike at lag 1 (from  $\theta$ )
- Spike at lag 12 (from  $\Theta$ )
- Spike at lag 13 (from  $\theta \cdot \Theta$  interaction)
- All other lags near zero

### Expected PACF Pattern

- Exponential decay at lags 1, 2, 3, ...
- Exponential decay at lags 12, 24, 36, ...

## Step-by-Step Process

- ① Examine ACF for slow decay at seasonal lags  $\Rightarrow$  seasonal differencing
- ② After differencing, check ACF/PACF patterns
- ③ Non-seasonal behavior at lags  $1, 2, \dots, s - 1$
- ④ Seasonal behavior at lags  $s, 2s, 3s, \dots$

## Practical Tips

- Start with  $d \leq 1$  and  $D \leq 1$
- Usually  $P, Q \leq 2$  is sufficient
- Use information criteria (AIC, BIC) for final selection
- Auto-SARIMA algorithms can help

## Maximum Likelihood Estimation

Standard approach for SARIMA:

- Conditional MLE (conditional on initial values)
- Exact MLE (via Kalman filter)

## Computational Considerations

- More parameters than ARIMA  $\Rightarrow$  more data needed
- Seasonal parameters estimated from lags  $s, 2s, \dots$
- Need sufficient seasonal cycles (at least 3-4 years of monthly data)

## Stationarity Conditions

Both non-seasonal and seasonal AR polynomials must have roots outside the unit circle:

- $\phi(z) = 0 \Rightarrow |z| > 1$
- $\Phi(z^s) = 0 \Rightarrow |z| > 1$

## Invertibility Conditions

Both non-seasonal and seasonal MA polynomials must have roots outside the unit circle:

- $\theta(z) = 0 \Rightarrow |z| > 1$
- $\Theta(z^s) = 0 \Rightarrow |z| > 1$

## Residual Analysis

After fitting SARIMA, check that residuals are white noise:

- ① Plot residuals over time (no patterns)
- ② ACF of residuals (no significant spikes)
- ③ Ljung-Box test at multiple lags including seasonal
- ④ Normality tests (Q-Q plot, Jarque-Bera)

## Important

Check ACF at **both** non-seasonal and seasonal lags!

Significant ACF at lag 12 suggests inadequate seasonal modeling.

## Information Criteria

Compare competing SARIMA models using:

- $AIC = -2 \ln(L) + 2k$
- $BIC = -2 \ln(L) + k \ln(n)$
- $AICc = AIC + \frac{2k(k+1)}{n-k-1}$  (corrected for small samples)

where  $k = p + q + P + Q + 1$  (plus 1 for variance).

## Auto-SARIMA

Python's `pmdarima.auto_arima()` with `seasonal=True` automatically searches for optimal  $(p, d, q) \times (P, D, Q)_s$ .

## Forecast Computation

SARIMA forecasts are computed recursively:

- Replace future  $\varepsilon_{T+h}$  with 0
- Replace future  $Y_{T+h}$  with forecasts  $\hat{Y}_{T+h|T}$
- Use known past values  $Y_T, Y_{T-1}, \dots$

## Seasonal Pattern in Forecasts

SARIMA forecasts naturally capture seasonality:

- Short-term: influenced by recent values
- Long-term: revert to seasonal pattern

## Uncertainty Quantification

$(1 - \alpha)\%$  prediction interval:

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

Variance computed from  $\text{MA}(\infty)$  representation.

## Key Properties

- Intervals widen with forecast horizon
- For  $I(1, 1)_s$  series: intervals grow without bound
- Seasonal pattern visible in point forecasts
- Uncertainty captures both trend and seasonal variation

## Behavior as $h \rightarrow \infty$

- Point forecasts converge to deterministic seasonal pattern
- If drift present: linear trend + seasonal pattern
- Forecast intervals continue to widen

## Practical Implication

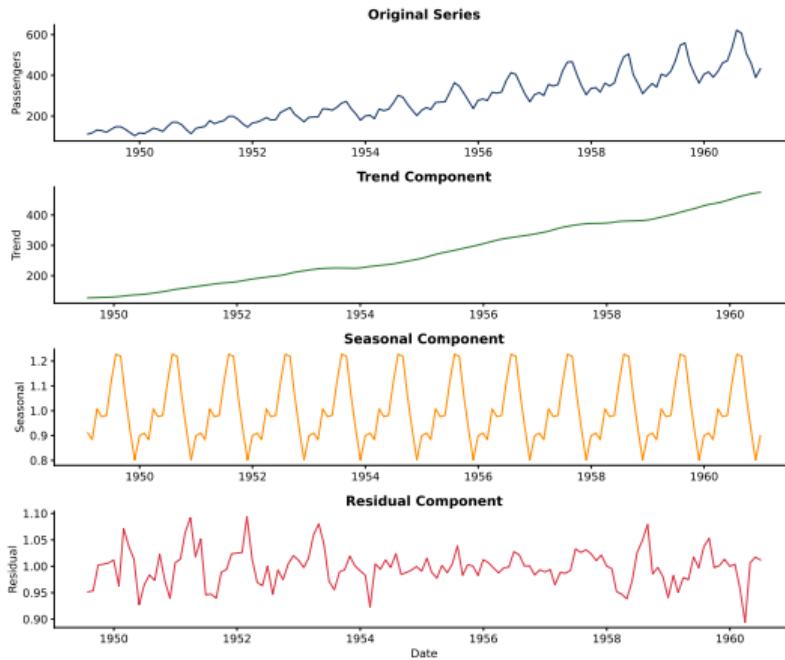
- Short-term: SARIMA captures both level and season
- Medium-term: Good seasonal forecasts, growing uncertainty
- Long-term: Mainly reflects seasonal pattern, wide intervals

## Airline Passengers Data



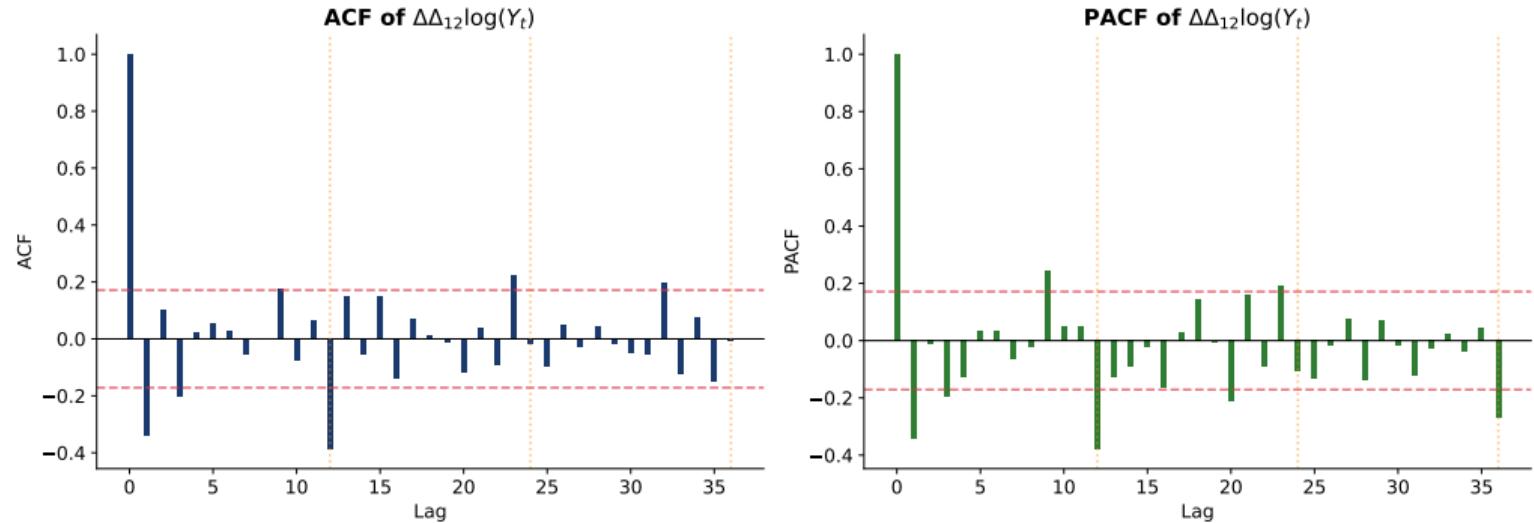
- Classic dataset: Monthly international airline passengers (1949-1960)
- Clear upward trend and growing seasonal amplitude

# Seasonal Decomposition



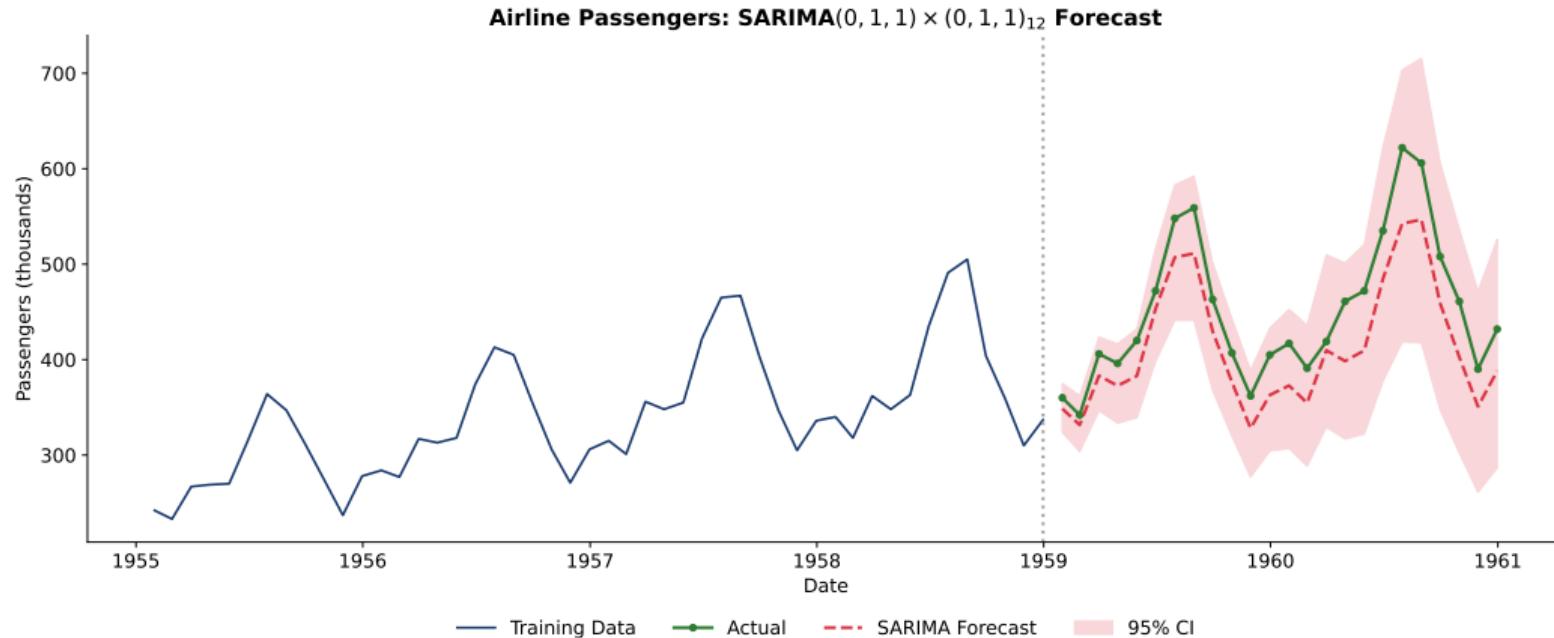
- Trend: Strong upward growth
- Seasonality: Summer peaks (vacation travel)
- Residual: Random variation after removing trend and season

# ACF/PACF Analysis



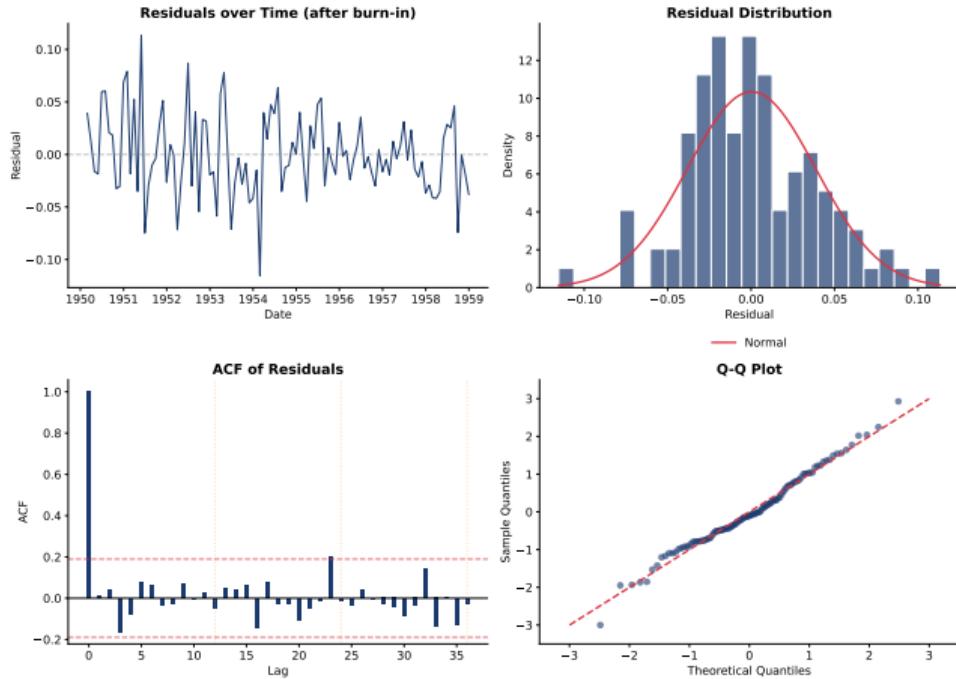
- After  $\Delta\Delta_{12}$  differencing: spikes at lags 1 and 12
- Suggests SARIMA(0, 1, 1)  $\times$  (0, 1, 1)<sub>12</sub> (Airline model)

# SARIMA Forecast Results



- SARIMA captures both trend and seasonal pattern
- Forecasts show appropriate seasonal peaks and troughs

# Model Diagnostics



- Residuals appear random; ACF within bounds at all lags
- Model adequately captures seasonal structure

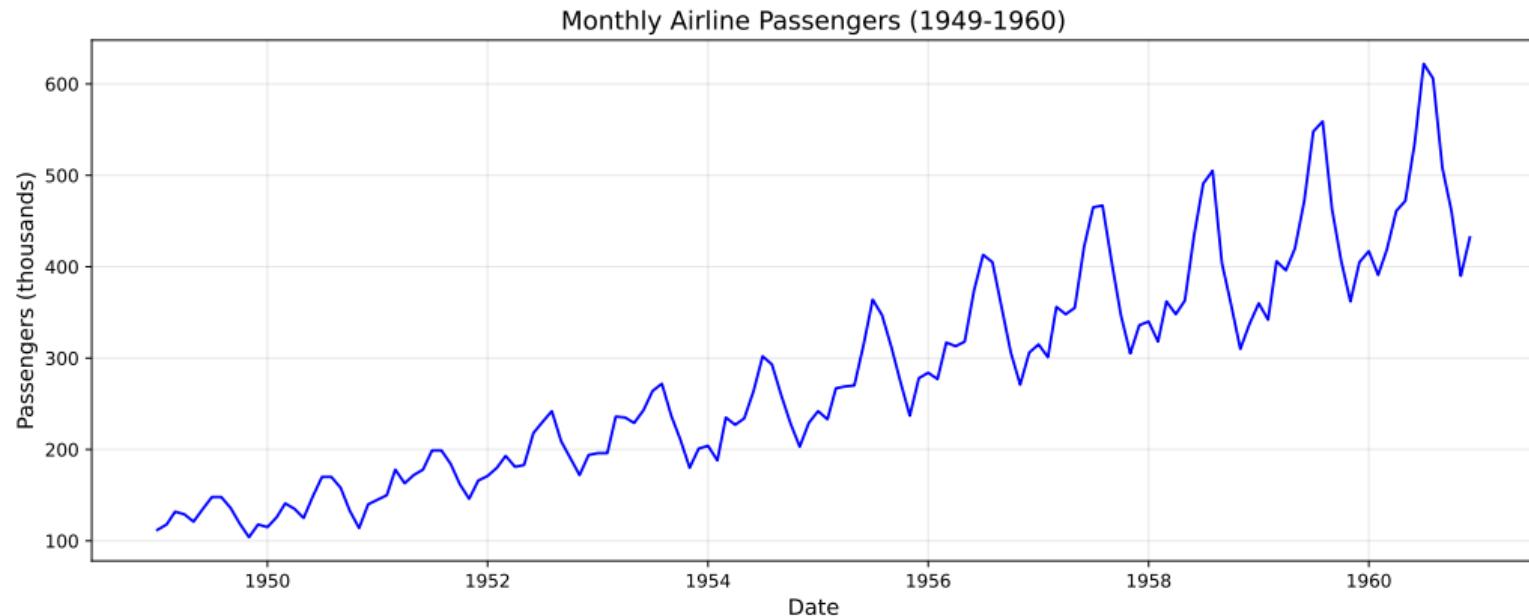
## Fitting SARIMA in Python

```
from statsmodels.tsa.statespace.sarimax import SARIMAX  
model = SARIMAX(y, order=(0,1,1), seasonal_order=(0,1,1,12))  
results = model.fit()  
forecast = results.get_forecast(steps=24)
```

## Note

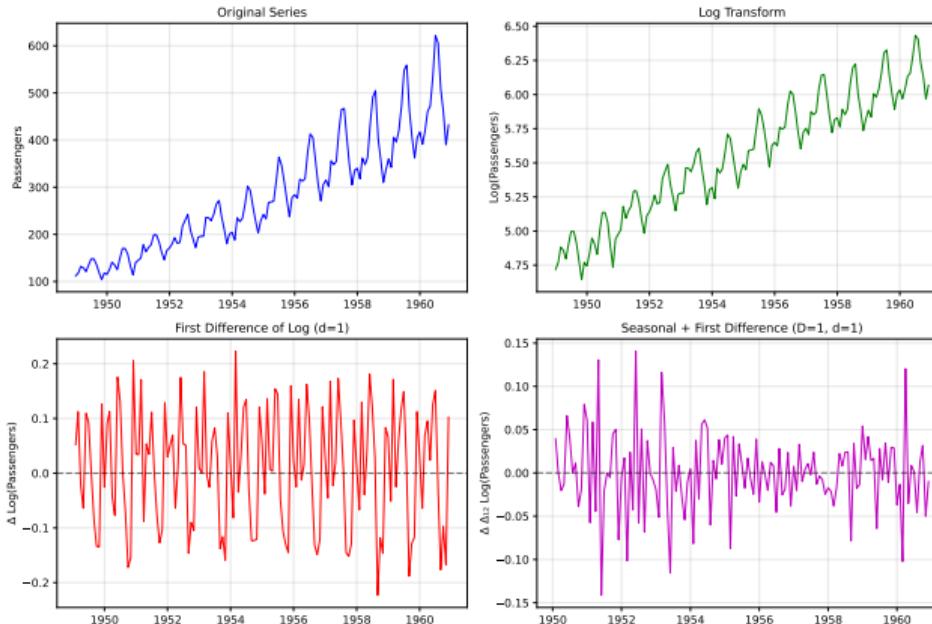
Complete Python examples with comments are provided in the Jupyter notebooks.

## Case Study: Airline Passengers Data



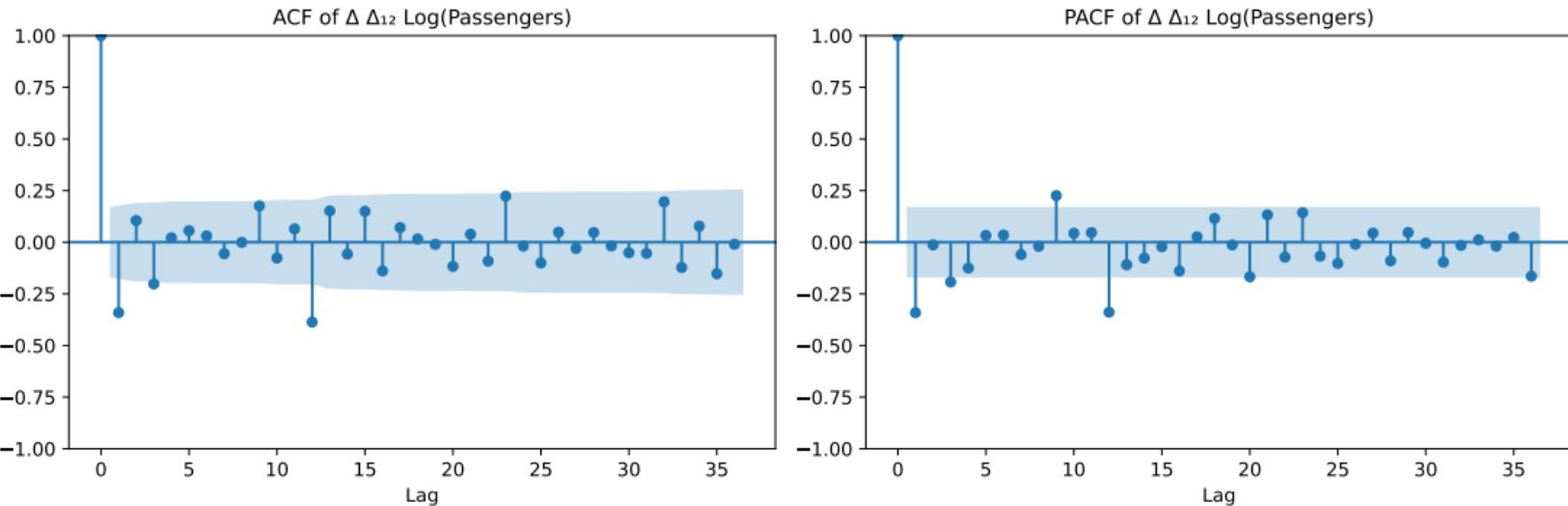
- Classic Box-Jenkins dataset: monthly airline passengers (1949-1960)
- Clear upward trend and increasing seasonal amplitude
- Multiplicative seasonality suggests log transformation

## Step 1: Transformations



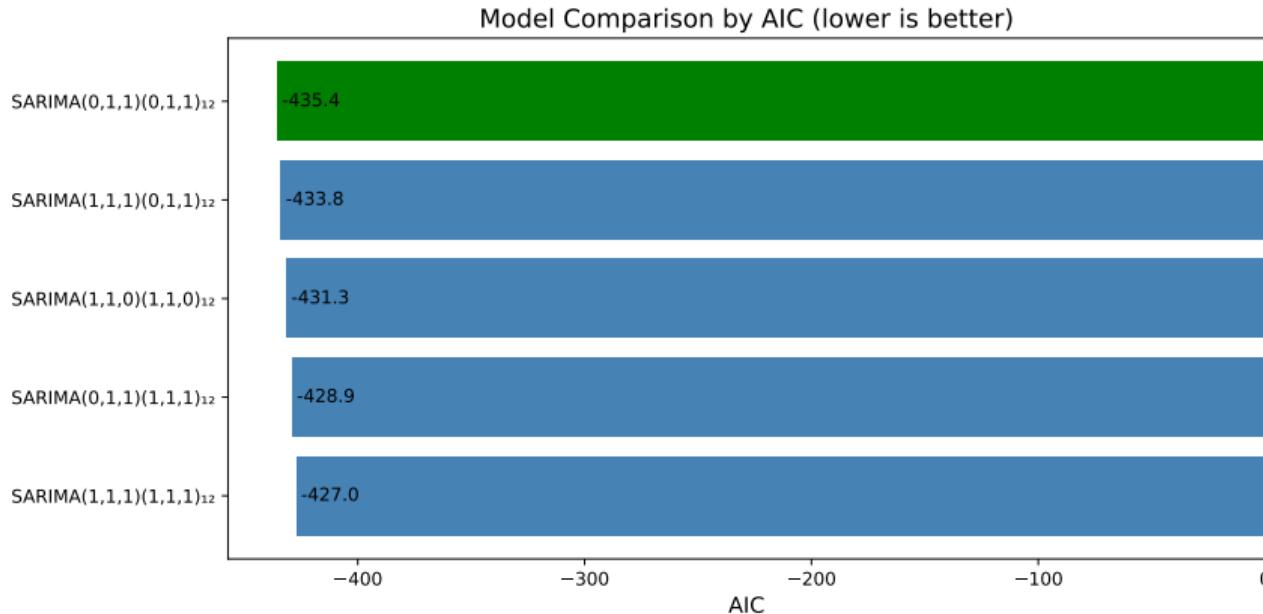
- Log transform stabilizes variance (multiplicative → additive)
- First difference removes trend; seasonal difference removes seasonality
- Double-differenced series appears stationary

## Step 2: ACF/PACF Analysis



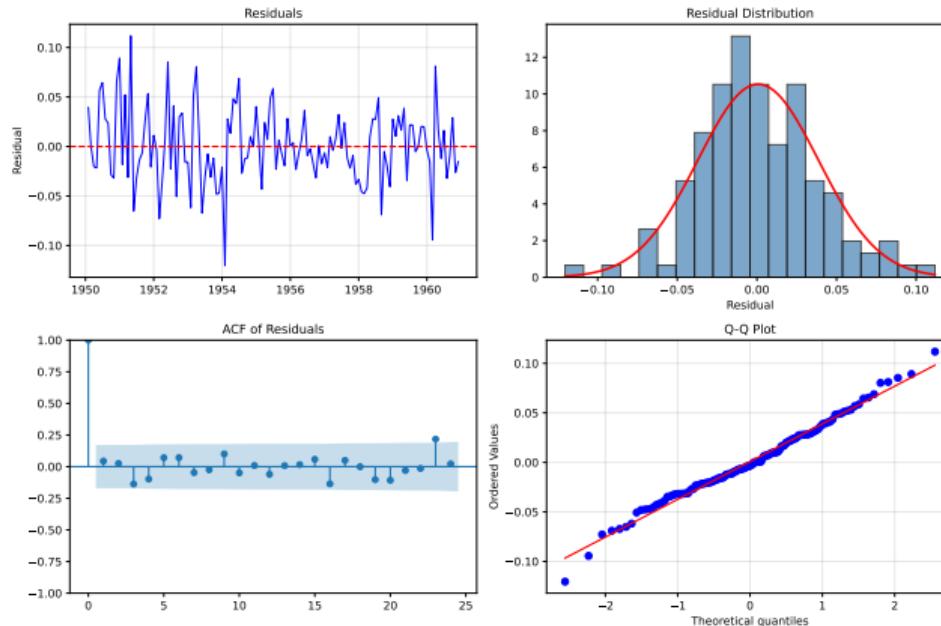
- ACF: Significant spike at lag 1 and lag 12  $\Rightarrow$  MA(1), SMA(1)
- PACF: Exponential decay pattern confirms MA structure
- Suggests SARIMA(0, 1, 1)  $\times$  (0, 1, 1)<sub>12</sub> (airline model)

## Step 3: Model Comparison



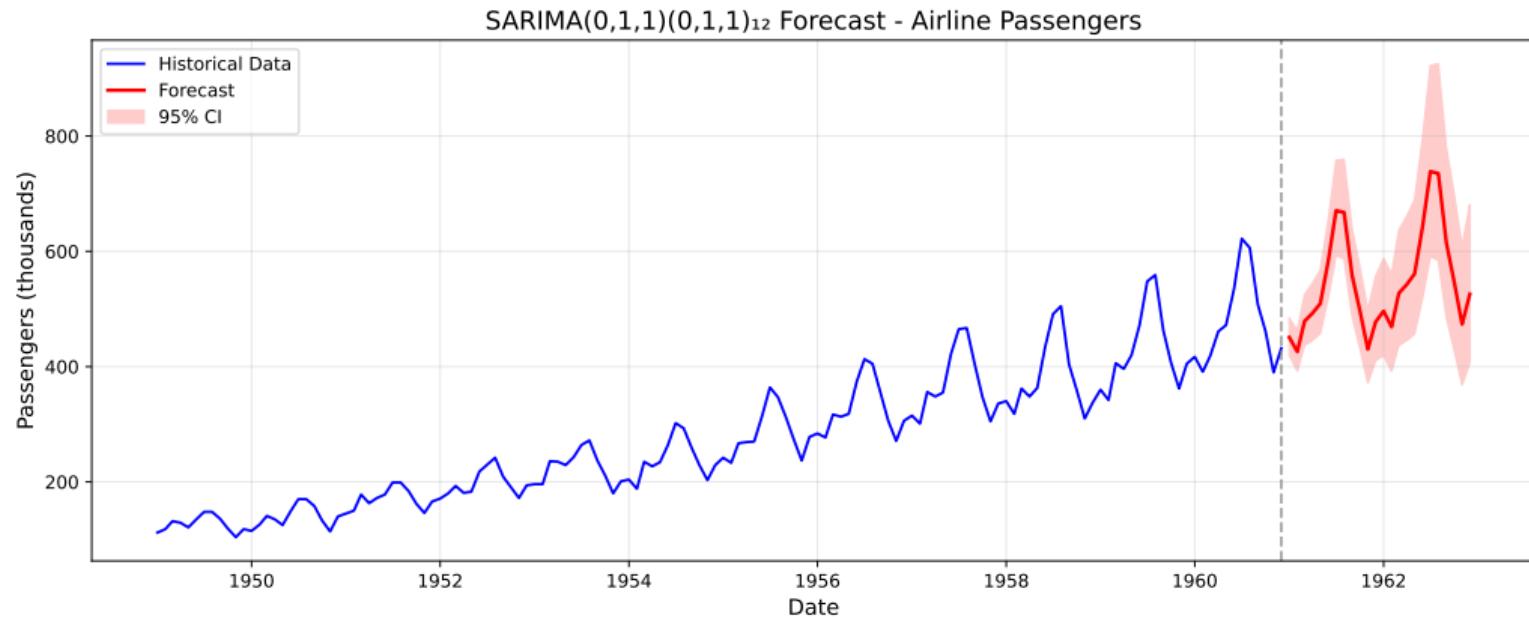
- Compare candidate SARIMA models using AIC criterion
- SARIMA(0, 1, 1)  $\times$  (0, 1, 1)<sub>12</sub> provides best fit (lowest AIC)
- This is the famous “airline model” identified by Box & Jenkins

## Step 4: Residual Diagnostics



- Residuals appear random with no remaining autocorrelation
- Q-Q plot shows approximate normality
- Model adequately captures both trend and seasonal structure

## Step 5: Forecasting



- 24-month forecast with 95% confidence interval
- Model captures seasonal pattern and upward trend
- Prediction intervals widen appropriately with forecast horizon

# Key Takeaways

## Main Points

- ① **Seasonality** is common in economic and business data
- ② **Seasonal differencing**  $(1 - L^s)$  removes stochastic seasonality
- ③ **SARIMA** $(p, d, q) \times (P, D, Q)_s$  extends ARIMA for seasonal data
- ④ **Multiplicative structure** captures seasonal-trend interactions
- ⑤ **ACF/PACF** show patterns at both regular and seasonal lags
- ⑥ **Model selection:** Use AIC/BIC or auto-SARIMA algorithms

## Next Steps

Chapter 5 will cover multivariate time series: VAR models, Granger causality, and cointegration.

## Quiz Question 1

### Question

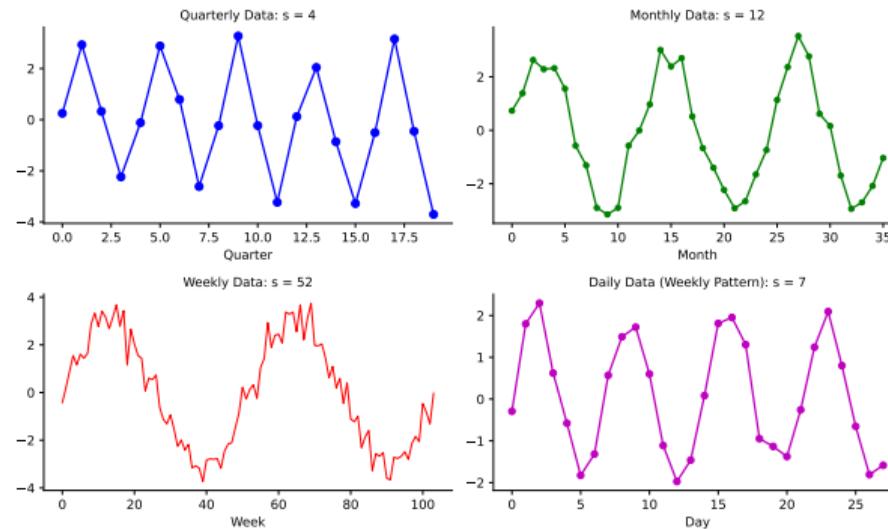
For monthly data with annual seasonality, what is the seasonal period  $s$ ?

- A  $s = 4$
- B  $s = 7$
- C  $s = 12$
- D  $s = 52$

## Quiz Question 1: Answer

Correct Answer: (C)  $s = 12$  (12 months per year)

Common periods: Quarterly=4, Monthly=12, Weekly=52, Daily=7, Hourly=24



## Quiz Question 2

### Question

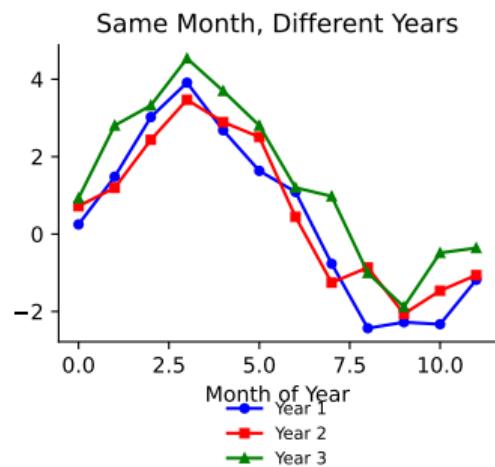
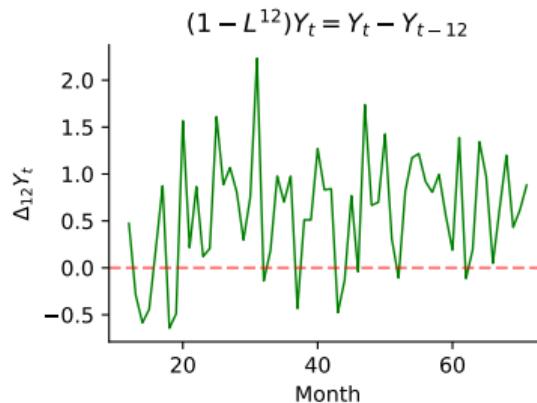
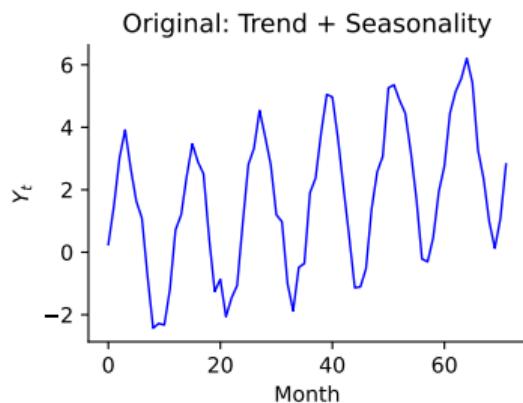
What does the seasonal difference operator  $(1 - L^{12})$  do to a monthly series?

- A Computes  $Y_t - Y_{t-1}$  (month-to-month change)
- B Computes  $Y_t - Y_{t-12}$  (year-over-year change)
- C Computes the 12-month moving average
- D Removes the trend component only

## Quiz Question 2: Answer

Correct Answer: (B) Year-over-year change

$(1 - L^{12})Y_t = Y_t - Y_{t-12}$  removes the seasonal pattern by comparing same months.



## Quiz Question 3

### Question

In SARIMA(1, 1, 1)  $\times$  (1, 1, 1)<sub>12</sub> notation, what does the (1, 1, 1)<sub>12</sub> part represent?

- A AR(1), differencing once, MA(1) at the regular level
- B Seasonal AR(1), seasonal differencing once, seasonal MA(1)
- C 12 AR terms, 12 differences, 12 MA terms
- D The model has 12 parameters in total

## Quiz Question 3: Answer

Correct Answer: (B)

Seasonal AR(1), seasonal differencing once, seasonal MA(1)

### SARIMA Notation Breakdown

SARIMA( $p, d, q$ )  $\times$  ( $P, D, Q$ )<sub>s</sub>:

- ( $p, d, q$ ) Non-seasonal: AR( $p$ ),  $d$  differences, MA( $q$ )
- ( $P, D, Q$ )<sub>s</sub> Seasonal: SAR( $P$ ),  $D$  seasonal diffs, SMA( $Q$ )

For  $(1, 1, 1) \times (1, 1, 1)_{12}$ :

- Non-seasonal: AR(1), one regular difference, MA(1)
- Seasonal: SAR(1) at lag 12, one  $\Delta_{12}$ , SMA(1) at lag 12

## Quiz Question 4

### Question

The “Airline Model” is  $\text{SARIMA}(0, 1, 1) \times (0, 1, 1)_{12}$ . How many parameters need to be estimated (excluding variance)?

- A 1
- B 2
- C 4
- D 12

## Quiz Question 4: Answer

Correct Answer: (B)

2 parameters

### Model Structure

SARIMA(0, 1, 1)  $\times$  (0, 1, 1)<sub>12</sub>:

$$(1 - L)(1 - L^{12})Y_t = (1 + \theta_1 L)(1 + \Theta_1 L^{12})\varepsilon_t$$

Parameters:

- $\theta_1$ : non-seasonal MA coefficient
- $\Theta_1$ : seasonal MA coefficient

Total: **2 parameters** (plus  $\sigma^2$ )

### Why “Airline Model”?

Box & Jenkins (1970) used this model to forecast international airline passengers. It's remarkably effective for many seasonal economic series!

## Quiz Question 5

### Question

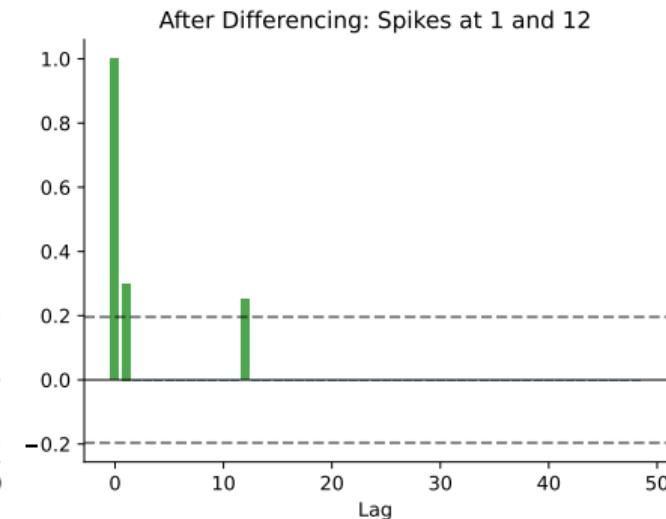
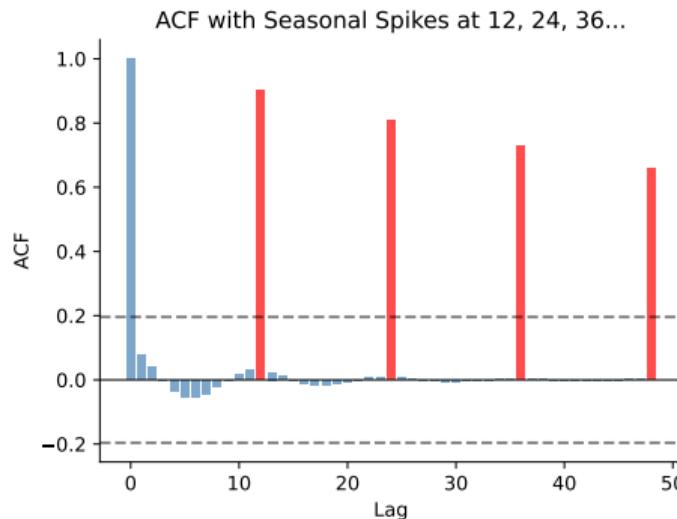
You observe significant ACF spikes at lags 12, 24, and 36 in a monthly series. What does this suggest?

- A The series has a unit root
- B The series has annual seasonality that needs seasonal differencing
- C The series follows an AR(36) process
- D The series is already stationary

## Quiz Question 5: Answer

Correct Answer: (B) Needs seasonal differencing

ACF spikes at 12, 24, 36 = stochastic seasonality. Apply  $(1 - L^{12})$  to remove it.



## Quiz Question 6

### Question

After applying  $(1 - L)(1 - L^{12})$  to a monthly series, the ACF shows a significant spike only at lag 1 and lag 12. What SARIMA model is suggested?

- A SARIMA(1, 1, 0)  $\times$  (1, 1, 0)<sub>12</sub>
- B SARIMA(0, 1, 1)  $\times$  (0, 1, 1)<sub>12</sub>
- C SARIMA(1, 1, 1)  $\times$  (1, 1, 1)<sub>12</sub>
- D SARIMA(0, 1, 0)  $\times$  (0, 1, 0)<sub>12</sub>

## Quiz Question 6: Answer

Correct Answer: (B)

SARIMA(0, 1, 1)  $\times$  (0, 1, 1)<sub>12</sub> (The Airline Model)

### ACF/PACF Identification Rules

For MA processes, ACF **cuts off** after lag  $q$ :

Pattern	Suggests
ACF spike at lag 1 only	MA(1) for non-seasonal part
ACF spike at lag 12 only	SMA(1) for seasonal part

Combined: MA(1)  $\times$  SMA(1) = (0,  $d$ , 1)  $\times$  (0,  $D$ , 1)<sub>12</sub>

With  $d = 1$  and  $D = 1$  (already differenced): (0, 1, 1)  $\times$  (0, 1, 1)<sub>12</sub>

## References

-  Box, G.E.P., Jenkins, G.M., Reinsel, G.C., & Ljung, G.M. (2015). *Time Series Analysis: Forecasting and Control*. 5th ed. Wiley.
-  Hyndman, R.J. & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*. 3rd ed. OTexts.
-  Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press.
-  Brockwell, P.J. & Davis, R.A. (2016). *Introduction to Time Series and Forecasting*. 3rd ed. Springer.