Chapter 4: SARIMA Models

Seasonal Time Series

Outline

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)

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!

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 is constant over time
- Can be 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: Overview

Visual Methods

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

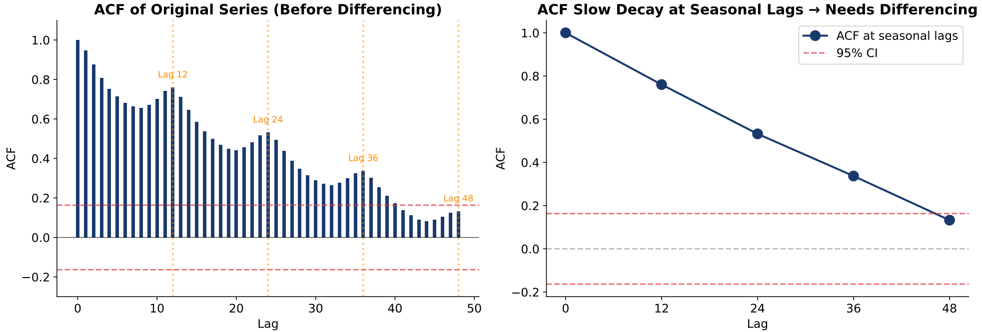
Statistical Tests

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

Key Principle

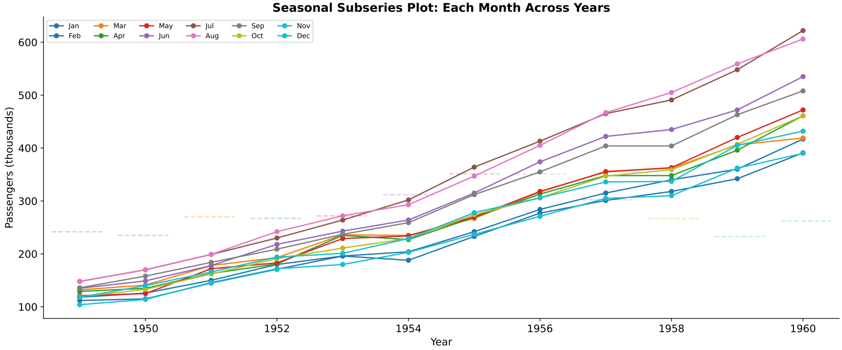
Always use **multiple methods** to confirm seasonality before modeling!

Visual Method 1: ACF for Seasonality Detection



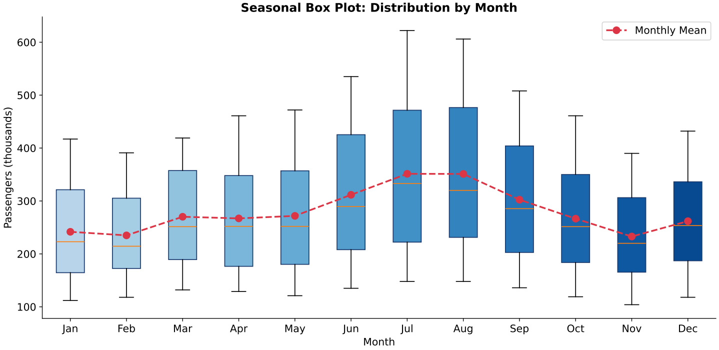
- **Left**: ACF of original series shows spikes at lags 12, 24, 36 (seasonal lags)
- **Right**: Slow decay at seasonal lags $\Rightarrow$ indicates **seasonal unit root**
- When ACF decays slowly at seasonal lags, apply seasonal differencing $(1 - L^s)$

Visual Method 2: Seasonal Subseries Plot



- Each line shows one month's values across all years
- Reveals: (1) Seasonal pattern (summer months higher), (2) Trend within each month
- If lines are roughly parallel $\Rightarrow$ additive seasonality
- If lines diverge (spread increases) $\Rightarrow$ multiplicative seasonality

Visual Method 3: Seasonal Box Plot



- Shows distribution of values for each month (season)
- Clear pattern: July-August peaks (summer travel), lower in winter
- Increasing variance by month $\Rightarrow$ suggests log transformation
- Red line shows monthly means – reveals the seasonal shape

Additive vs Multiplicative Seasonality

Additive Model

$$Y_t = T_t + S_t + \varepsilon_t$$

- Seasonal amplitude is **constant**
- Use when variance is stable
- Difference: $Y_t - Y_{t-s}$

Multiplicative Model

$$Y_t = T_t \times S_t \times \varepsilon_t$$

- Seasonal amplitude **grows with level**
- Use when variance increases
- Log transform: $\log(Y_t)$

Airline Data

Seasonal amplitude grows over time $\Rightarrow$ **multiplicative**
Solution: Model $\log(Y_t)$ instead of Y_t

F-test for Seasonal Dummies

Model with Seasonal Dummies

$$Y_t = \alpha + \sum_{j=1}^{s-1} \gamma_j D_{jt} + \varepsilon_t$$

where $D_{jt} = 1$ if observation t is in season j, 0 otherwise.

F-statistic

Test $H_0 : \gamma_1 = \gamma_2 = \cdots = \gamma_{s-1} = 0$ (no seasonality)

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

where SSR_R = restricted sum of squares, SSR_U = unrestricted.

Decision Rule

Reject H_0 if $F > F_{\alpha,s-1,n-s}$ (critical value) or if p-value $< \alpha$.
Rejection $\Rightarrow$ significant deterministic seasonality present.

Kruskal-Wallis Test for Seasonality

Setup

Nonparametric test comparing s seasonal groups. Let n_j = observations in season j, N = total observations, R_j = sum of ranks in season j.

Test Statistic

$$H = \frac{12}{N(N+1)} \sum_{j=1}^{s} \frac{R_j^2}{n_j} - 3(N+1)$$

Under H_0 (no seasonal differences): $H \sim \chi^2_{s-1}$ (approximately)

Decision Rule

Reject H_0 if $H > \chi^2_{\alpha,s-1}$ or if p-value $< \alpha$.
Advantage: No normality assumption; robust to outliers.

HEGY Test: Setup (Quarterly Data)

Auxiliary Regression

For quarterly data ($s = 4$), define transformed variables:

$$\begin{aligned} y_{1t} &= (1+L)(1+L^2)Y_t = Y_t + Y_{t-1} + Y_{t-2} + Y_{t-3} \\ y_{2t} &= -(1-L)(1+L^2)Y_t = -Y_t + Y_{t-1} - Y_{t-2} + Y_{t-3} \\ y_{3t} &= -(1-L^2)Y_t = -Y_t + Y_{t-2} \end{aligned}$$

Test Regression

$$(1-L^4)Y_t = \pi_1 y_{1,t-1} + \pi_2 y_{2,t-1} + \pi_3 y_{3,t-2} + \pi_4 y_{3,t-1} + \varepsilon_t$$

Include deterministic terms (constant, trend, seasonal dummies) as needed.

HEGY Test: Hypotheses and Decision Rules

Unit Root Hypotheses

- $H_0^{(1)}: \pi_1 = 0 \Rightarrow$ unit root at zero frequency (regular unit root)
- $H_0^{(2)}: \pi_2 = 0 \Rightarrow$ unit root at π frequency (semi-annual)
- $H_0^{(3)}: \pi_3 = \pi_4 = 0 \Rightarrow$ unit roots at $\pm\pi/2$ frequencies (annual)

Test Statistics

- $t_1 = \hat{\pi}_1 / SE(\hat{\pi}_1)$ for $H_0^{(1)}$ (use HEGY critical values)
- $t_2 = \hat{\pi}_2 / SE(\hat{\pi}_2)$ for $H_0^{(2)}$
- F_{34} for joint test of $H_0^{(3)}$

Decision Rule

Use HEGY critical values (not standard t or F). If $H_0^{(1)}$ not rejected $\Rightarrow d = 1$. If $H_0^{(2)}$ or $H_0^{(3)}$ not rejected $\Rightarrow D = 1$.

Canova-Hansen Test

Key Difference from HEGY

- HEGY: H_0 = unit root (stochastic seasonality)
- CH: H_0 = **no** unit root (deterministic seasonality)

Test Statistic

Based on cumulative sum of residuals from regression with seasonal dummies:

$$L_s = \frac{1}{T^2} \sum_{t=1}^{T} \left(\sum_{i=1}^{t} \hat{f}_i \right)' \hat{\Omega}^{-1} \left(\sum_{i=1}^{t} \hat{f}_i \right)$$

where $\hat{f}_t$ are residuals projected onto seasonal frequencies.

Decision Rule

Reject H_0 if $L_s >$ critical value $\Rightarrow$ seasonal unit root present.
Use: When you suspect stable (deterministic) seasonality but want to confirm.

Summary: Which Test to Use?

Test	H_0	When to Use
F-test	No seasonality	Detect if any seasonality exists
Kruskal-Wallis	No seasonal diff.	Non-normal data, outliers
HEGY	Seasonal unit root	Determine d and D
Canova-Hansen	Deterministic season.	Confirm stable seasonality

Practical Workflow

1. F-test or Kruskal-Wallis: Is seasonality present?
2. If yes $\Rightarrow$ HEGY: Is it stochastic (unit root)?
3. HEGY rejects $\Rightarrow$ use seasonal dummies
4. HEGY fails to reject $\Rightarrow$ use $(1 - L^s)$ differencing

Kruskal-Wallis test for seasonality groups = [y[y.index.month == m] for m in range(1, 13)] stat, pvalue = stats.kruskal(*groups) print(f"Kruskal-Wallis: stat=stat:.2f, p=pvalue:.4f")
Check ACF at seasonal lags acf$_v als = acf(y, nlags = 36) seasonal_a cf =$
$[acf_v als[12], acf_v als[24], acf_v als[36]] print(f" ACFatlags 12, 24, 36 : seasonal_a cf")$

The Seasonal Difference Operator

Definition 2 (Seasonal Difference)

The **seasonal difference operator** Δ_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

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)

Seasonal Integration

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:

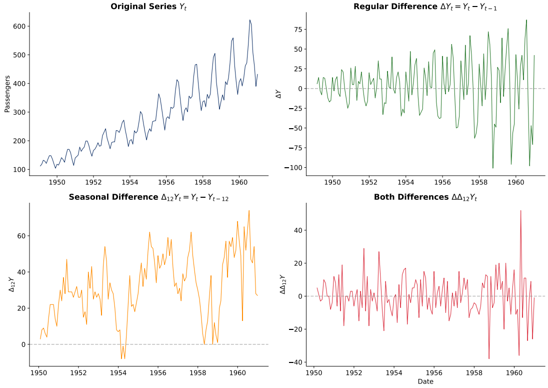
$$(1 - L)^d (1 - L^s)^D Y_t$$

is stationary.

Common Cases

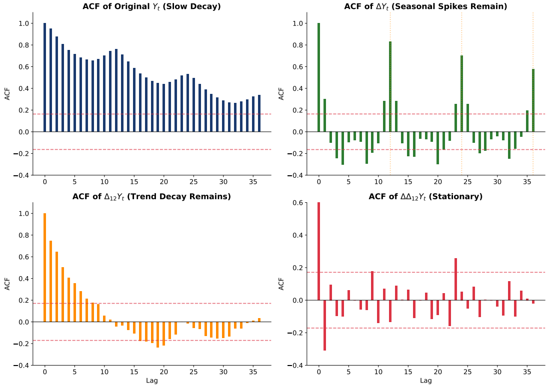
- $I(1, 0)_{12}$: Regular unit root only (monthly)
- $I(0, 1)_{12}$: Seasonal unit root only
- $I(1, 1)_{12}$: Both regular and seasonal unit roots

Effect of Differencing: Time Series View



- Regular differencing ΔY_t removes trend but seasonality remains
- Seasonal differencing $\Delta_{12} Y_t$ removes seasonality but trend remains
- Both needed: $\Delta\Delta_{12} Y_t$ appears stationary

Effect of Differencing: ACF View



- Original: slow decay (non-stationary)
- After Δ: seasonal spikes at 12, 24 remain
- After Δ_{12}: decay at low lags remains
- After both: ACF cuts off quickly $\Rightarrow$ ready for SARIMA modeling

Deciding the Order of Differencing

Step-by-Step Procedure

1. Plot the series – does it have trend? Seasonality?
2. Check ACF – slow decay at lags 1,2,3... $\Rightarrow$ need $d \geq 1$
3. Check ACF at seasonal lags – slow decay at $s, 2s, 3s$ $\Rightarrow$ need $D \geq 1$
4. Apply differencing and repeat until ACF cuts off

Common Pitfalls

- **Over-differencing**: Introduces artificial patterns; ACF shows negative spike at lag 1 or s
- **Under-differencing**: ACF remains slowly decaying
- Rule: Rarely need $d > 1$ or $D > 1$

Airline Data Decision

ACF shows: (1) slow decay $\Rightarrow d = 1$, (2) spikes at 12, 24 $\Rightarrow D = 1$
Result: Apply $(1 - L)(1 - L^{12})$ before modeling

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

- $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$: Non-seasonal AR
- $\Phi(L^s) = 1 - \Phi_1 L^s - \cdots - \Phi_P L^{Ps}$: Seasonal AR
- $\theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$: Non-seasonal MA
- $\Theta(L^s) = 1 + \Theta_1 L^s + \cdots + \Theta_Q L^{Qs}$: Seasonal MA
- $(1-L)^d$: Regular differencing
- $(1-L^s)^D$: Seasonal differencing

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$(1, 1, 1) \times (1, 1, 1)_{12}$: Monthly data with AR(1), MA(1), seasonal AR(1), seasonal MA(1), regular and seasonal differencing.

Common SARIMA Models

Airline Model: 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 for many economic series (Box & Jenkins, 1970).

SARIMA$(1,0,0) \times (1,0,0)_s$

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

Pure seasonal and non-seasonal autoregressive model.

SARIMA$(0,1,1) \times (0,1,0)_s$

$$(1-L)(1-L^s)Y_t = (1+\theta L)\varepsilon_t$$

Random walk with seasonal differencing and MA(1) errors.

The Multiplicative Structure

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$(1,0,0) \times (1,0,0)_{12}$

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

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.

ACF/PACF for Seasonal Models

Key Insight

Seasonal models show patterns at both:

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

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

Example: Airline Model ACF/PACF

SARIMA$(0,1,1) \times (0,1,1)_{12}$

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

Expected PACF Pattern

- Exponential decay at lags $1, 2, 3, \ldots$
- Exponential decay at lags $12, 24, 36, \ldots$

Model Identification Guidelines

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 $1, 2, \ldots, s-1$
4. Seasonal behavior at lags $s, 2s, 3s, \ldots$

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

Estimation Methods

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 $s, 2s, \ldots$
- Need sufficient seasonal cycles (at least 3-4 years of monthly data)

Stationarity and Invertibility

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:

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

Diagnostic Checking

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.

Model Selection Criteria

Information Criteria

Compare competing SARIMA models using:

- $\text{AIC} = -2\ln(L) + 2k$
- $\text{BIC} = -2\ln(L) + k\ln(n)$
- $\text{AICc} = \text{AIC} + \frac{2k(k+1)}{n-k-1}$ (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 $(p, d, q) \times (P, D, Q)_s$.

Point Forecasts

Forecast Computation

SARIMA forecasts are computed recursively:

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

Seasonal Pattern in Forecasts

SARIMA forecasts naturally capture seasonality:

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

Forecast Intervals

Uncertainty Quantification

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

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

Variance computed from MA(∞) 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

Long-Horizon Forecasts

Behavior as $h \to \infty$

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

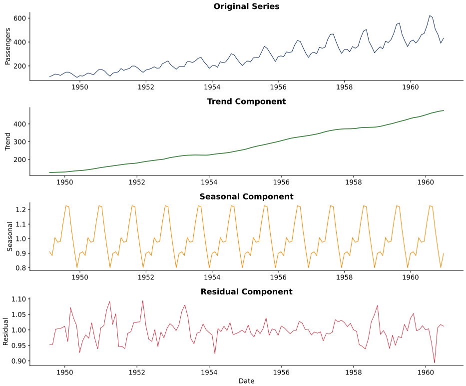
Practical Implication

- 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

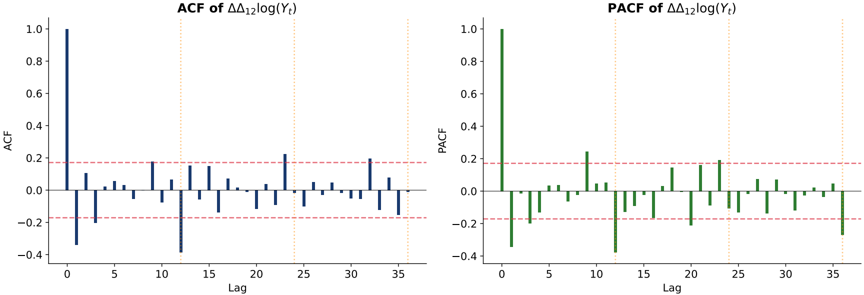
- Classic dataset: Monthly international airline passengers (1949-1960)
- Clear upward trend and growing seasonal amplitude

Seasonal Decomposition



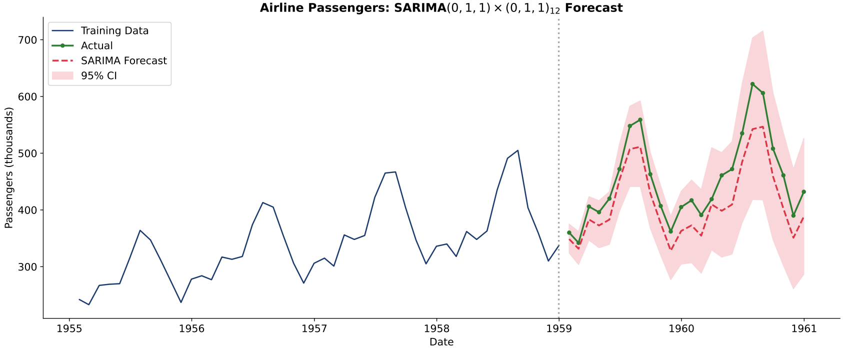
- Trend: Strong upward growth
- Seasonality: Summer peaks (vacation travel)
- Residual: Random variation after removing trend and season

ACF/PACF Analysis



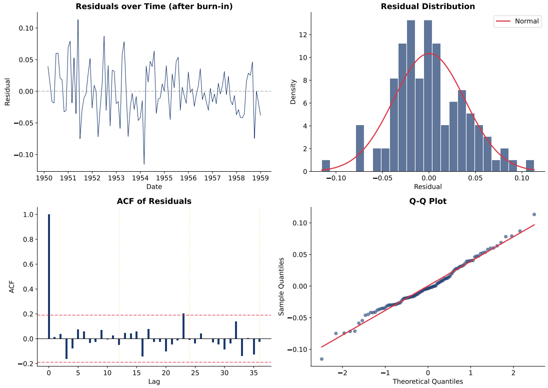
- After $\Delta\Delta_{12}$ differencing: spikes at lags 1 and 12
- Suggests SARIMA$(0,1,1) \times (0,1,1)_{12}$ (Airline model)

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

Fit SARIMA(0,1,1)(0,1,1)[12] model = SARIMAX(y, order=(0,1,1),
seasonal$_o$$rder = (0, 1, 1, 12))results = model.fit()print(results.summary())$
Forecast forecast = results.get$_f$$orecast(steps = 24)$

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.

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.