Chapter 4: SARIMA Models

Seminar

Seminar Outline

Quiz 1: Seasonal Differencing

Question

For monthly data with annual seasonality, what does the operator $(1 - L^{12})$ do?

- A) Takes 12 consecutive differences
- B) Computes $Y_t - Y_{t-12}$
- C) Averages over 12 months
- D) Removes the first 12 observations

Quiz 1: Seasonal Differencing

Question

For monthly data with annual seasonality, what does the operator $(1 - L^{12})$ do?

- A) Takes 12 consecutive differences
- B) Computes $Y_t - Y_{t-12}$
- C) Averages over 12 months
- D) Removes the first 12 observations

Answer: B – Computes $Y_t - Y_{t-12}$

Seasonal difference operator:

$$(1 - L^{12})Y_t = Y_t - L^{12}Y_t = Y_t - Y_{t-12}$$

Example (January sales): $Y_{Jan2025} - Y_{Jan2024}$
Effect: Removes stable annual seasonal pattern
Note: $(1 - L^s)$ for any seasonal period s (quarterly: $s = 4$, weekly: $s = 52$)

Quiz 2: SARIMA Notation

Question

What does SARIMA$(1, 1, 1) \times (1, 1, 1)_{12}$ represent?

- A) 12 different ARIMA models
- B) ARIMA with 12 AR and 12 MA terms
- C) ARIMA(1,1,1) with seasonal ARIMA(1,1,1) at period 12
- D) A model requiring 12 years of data

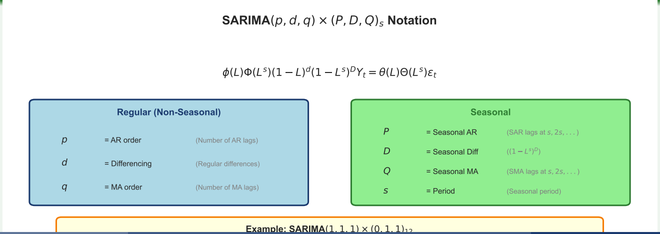
Quiz 2: SARIMA Notation

Question

What does SARIMA$(1,1,1) \times (1,1,1)_{12}$ represent?

- A) 12 different ARIMA models
- B) ARIMA with 12 AR and 12 MA terms
- C) ARIMA(1,1,1) with seasonal ARIMA(1,1,1) at period 12
- D) A model requiring 12 years of data

Answer: C – ARIMA(1,1,1) with seasonal ARIMA(1,1,1) at period 12



Quiz 3: The Airline Model

Question

The "airline model" refers to SARIMA$(0,1,1) \times (0,1,1)_{12}$. How many parameters does it have (excluding variance)?

- A) 2 parameters
- B) 4 parameters
- C) 6 parameters
- D) 12 parameters

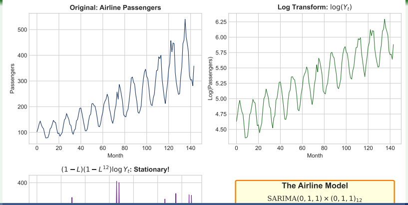
Quiz 3: The Airline Model

Question

The "airline model" refers to SARIMA$(0, 1, 1) \times (0, 1, 1)_{12}$. How many parameters does it have (excluding variance)?

- A) 2 parameters
- B) 4 parameters
- C) 6 parameters
- D) 12 parameters

Answer: A – 2 parameters (θ_1 and Θ_1)



Quiz 4: ACF of Seasonal Data

Question

For monthly data with strong seasonality, where would you expect to see significant ACF spikes?

- A) Only at lag 1
- B) Only at lag 12
- C) At lags 12, 24, 36, ...
- D) Randomly distributed

Quiz 4: ACF of Seasonal Data

Question

For monthly data with strong seasonality, where would you expect to see significant ACF spikes?

- A) Only at lag 1
- B) Only at lag 12
- C) At lags 12, 24, 36, ...
- D) Randomly distributed

Answer: C – At lags 12, 24, 36, ...

Intuition: January 2024 is similar to January 2023, 2022, etc.
ACF pattern for seasonal data:

- Spikes at lag s: $\rho_{12} \neq 0$
- Spikes at lag $2s$: $\rho_{24} \neq 0$
- Spikes at lag $3s$: $\rho_{36} \neq 0$

Diagnostic: If ACF decays slowly at seasonal lags $\Rightarrow$ seasonal unit root ($D = 1$)
If ACF cuts off after lag s $\Rightarrow$ seasonal MA ($Q = 1$)

Quiz 5: Multiplicative Structure

Question

In SARIMA, what does "multiplicative structure" mean?

- A) The seasonal amplitude grows proportionally
- B) Regular and seasonal polynomials are multiplied
- C) We multiply the data by seasonal factors
- D) The model is estimated using multiplication

Quiz 5: Multiplicative Structure

Question

In SARIMA, what does "multiplicative structure" mean?

- A) The seasonal amplitude grows proportionally
- B) Regular and seasonal polynomials are multiplied
- C) We multiply the data by seasonal factors
- D) The model is estimated using multiplication

Answer: B – Regular and seasonal polynomials are multiplied

Multiplicative SARIMA:

$$\underbrace{\phi(L)\Phi(L^s)}_{\text{AR}} \cdot \underbrace{(1-L)^d(1-L^s)^D}_{\text{Diff}} Y_t = \underbrace{\theta(L)\Theta(L^s)}_{\text{MA}} \varepsilon_t$$

Example: $\phi(L)\Phi(L^{12}) = (1-\phi_1 L)(1-\Phi_1 L^{12})$

$$= 1 - \phi_1 L - \Phi_1 L^{12} + \phi_1\Phi_1 L^{13}$$

Cross-term $\phi_1\Phi_1 L^{13}$: Captures interaction between short and long dynamics

Quiz 6: Seasonal vs Regular Differencing

Question

When would you apply both regular ($d = 1$) and seasonal ($D = 1$) differencing?

- A) When data has only a trend
- B) When data has only seasonality
- C) When data has both trend and seasonal non-stationarity
- D) Never – they cancel each other

Quiz 6: Seasonal vs Regular Differencing

Question

When would you apply both regular ($d = 1$) and seasonal ($D = 1$) differencing?

- A) When data has only a trend
- B) When data has only seasonality
- C) When data has both trend and seasonal non-stationarity
- D) Never – they cancel each other

Answer: C – Both trend and seasonal non-stationarity

Combined differencing:

$$W_t = (1 - L)(1 - L^{12})Y_t = Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}$$

When needed:

- ACF decays slowly at lags 1, 2, 3, ... $\Rightarrow$ need $d = 1$
- ACF decays slowly at lags 12, 24, 36, ... $\Rightarrow$ need $D = 1$

Classic example: Airline passengers, retail sales, energy demand

Quiz 7: Detecting Seasonality from ACF

Question

The ACF of a monthly time series shows slow decay at lags 12, 24, and 36. What does this suggest?

- A) The series is stationary
- B) The series needs regular differencing only
- C) The series has a seasonal unit root requiring $D = 1$
- D) The series is white noise

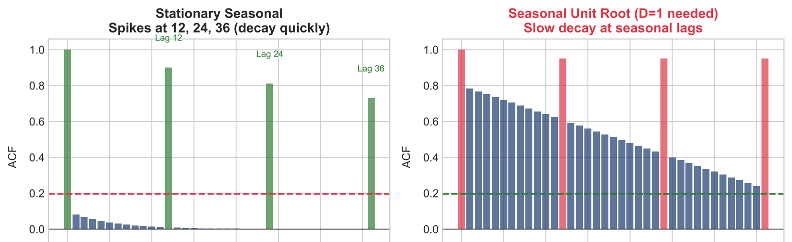
Quiz 7: Detecting Seasonality from ACF

Question

The ACF of a monthly time series shows slow decay at lags 12, 24, and 36. What does this suggest?

- A) The series is stationary
- B) The series needs regular differencing only
- C) The series has a seasonal unit root requiring $D = 1$
- D) The series is white noise

Answer: C – Seasonal unit root requiring $D = 1$



Quiz 8: Multiplicative vs Additive Seasonality

Question

If the seasonal amplitude of a time series grows proportionally with the level, this indicates:

A) Additive seasonality – use $(1 - L^s)$
B) Multiplicative seasonality – use log transformation
C) No seasonality present
D) Need for regular differencing only

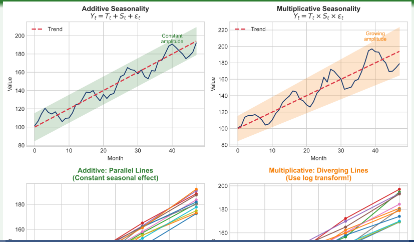
Quiz 8: Multiplicative vs Additive Seasonality

Question

If the seasonal amplitude of a time series grows proportionally with the level, this indicates:

- A) Additive seasonality – use $(1 - L^s)$
- B) Multiplicative seasonality – use log transformation
- C) No seasonality present
- D) Need for regular differencing only

Answer: B – Multiplicative seasonality, use log transformation



Quiz 9: Seasonal Subseries Plot

Question

In a seasonal subseries plot, what indicates multiplicative seasonality?

- A) Lines for each month are parallel
- B) Lines for each month diverge (spread increases over time)
- C) All months have the same mean
- D) Lines are horizontal

Quiz 9: Seasonal Subseries Plot

Question

In a seasonal subseries plot, what indicates multiplicative seasonality?

- A) Lines for each month are parallel
- B) Lines for each month diverge (spread increases over time)
- C) All months have the same mean
- D) Lines are horizontal

Answer: B – Lines diverge (spread increases over time)

Seasonal subseries plot: Groups data by month, plots each month's values across years
Interpretation:

- **Parallel lines**: Constant seasonal amplitude $\Rightarrow$ Additive
- **Diverging lines**: Growing amplitude $\Rightarrow$ Multiplicative
- **Horizontal lines**: No trend within months

Action: If multiplicative, apply log before fitting SARIMA

Quiz 10: Invertibility in SARIMA

Question

For SARIMA$(0,1,1)\times(0,1,1)_{12}$ to be invertible, which condition must hold?

- A) $|\theta_1| < 1$ only
- B) $|\Theta_1| < 1$ only
- C) Both $|\theta_1| < 1$ and $|\Theta_1| < 1$
- D) No invertibility condition exists for MA models

Quiz 10: Invertibility in SARIMA

Question

For SARIMA$(0,1,1) \times (0,1,1)_{12}$ to be invertible, which condition must hold?

- A) $|\theta_1| < 1$ only
- B) $|\Theta_1| < 1$ only
- C) Both $|\theta_1| < 1$ and $|\Theta_1| < 1$
- D) No invertibility condition exists for MA models

Answer: C – Both $|\theta_1| < 1$ and $|\Theta_1| < 1$

Invertibility condition: All MA roots must be outside unit circle
Multiplicative MA: $(1 + \theta_1 L)(1 + \Theta_1 L^{12})$
Roots:

- Regular: $z = -1/\theta_1$ must satisfy $|z| > 1 \Leftrightarrow |\theta_1| < 1$
- Seasonal: $z^{12} = -1/\Theta_1$ must satisfy $|z| > 1 \Leftrightarrow |\Theta_1| < 1$

Both conditions required for overall invertibility!

Quiz 11: HEGY Test

Question

The HEGY test is used to:

- A) Estimate SARIMA parameters
- B) Test for unit roots at different frequencies (trend and seasonal)
- C) Check residual normality
- D) Compare SARIMA models using information criteria

Quiz 11: HEGY Test

Question

The HEGY test is used to:

- A) Estimate SARIMA parameters
- B) Test for unit roots at different frequencies (trend and seasonal)
- C) Check residual normality
- D) Compare SARIMA models using information criteria

Answer: B – Test for unit roots at different frequencies

HEGY test (Hylleberg-Engle-Granger-Yoo, 1990):
Tests for unit roots at:

- Zero frequency ($\omega = 0$): trend unit root $\Rightarrow d = 1$
- Nyquist frequency ($\omega = \pi$): biannual cycle
- Seasonal frequencies: annual cycles $\Rightarrow D = 1$

Decision:

- Reject at all frequencies $\Rightarrow$ use seasonal dummies

Quiz 12: Seasonal MA Identification

Question

After applying $(1-L)(1-L^{12})$, the ACF shows a single significant spike at lag 12 only (no spike at lag 1). The PACF decays at seasonal lags. This suggests:

- A) SARIMA$(0,1,0) \times (0,1,1)_{12}$
- B) SARIMA$(0,1,1) \times (0,1,0)_{12}$
- C) SARIMA$(1,1,0) \times (1,1,0)_{12}$
- D) SARIMA$(0,1,1) \times (0,1,1)_{12}$

Quiz 12: Seasonal MA Identification

Question

After applying $(1-L)(1-L^{12})$, the ACF shows a single significant spike at lag 12 only (no spike at lag 1). The PACF decays at seasonal lags. This suggests:

- A) $\text{SARIMA}(0,1,0) \times (0,1,1)_{12}$
- B) $\text{SARIMA}(0,1,1) \times (0,1,0)_{12}$
- C) $\text{SARIMA}(1,1,0) \times (1,1,0)_{12}$
- D) $\text{SARIMA}(0,1,1) \times (0,1,1)_{12}$

Answer: A – $\text{SARIMA}(0,1,0) \times (0,1,1)_{12}$

Pattern analysis:

	ACF	**PACF**
Regular lags	No spikes	No spikes
Seasonal lags	Cuts off at s	Decays

Interpretation:

Quiz 13: Over-differencing

Question

After differencing, the ACF shows a large negative spike at lag 1 or lag s. This typically indicates:

A) The model needs more AR terms
B) The series has been over-differenced
C) The series is perfectly stationary
D) Heteroskedasticity is present

Quiz 13: Over-differencing

Question

After differencing, the ACF shows a large negative spike at lag 1 or lag s. This typically indicates:

- A) The model needs more AR terms
- B) The series has been over-differenced
- C) The series is perfectly stationary
- D) Heteroskedasticity is present

Answer: B – The series has been over-differenced

Over-differencing signature:

- ACF at lag 1 $\approx -0.5 \Rightarrow$ over-differenced at d
- ACF at lag $s \approx -0.5 \Rightarrow$ over-differenced at D

Why? If $\Delta Y_t = \varepsilon_t$, then $\Delta^2 Y_t = \varepsilon_t - \varepsilon_{t-1}$
This is MA(1) with $\theta = -1$, giving $\rho_1 = \frac{-1}{1+1} = -0.5$
Fix: Reduce d or D by one and re-examine ACF/PACF

Quiz 14: Forecasting Horizon

Question

For a SARIMA model with $D = 1$, what happens to forecast confidence intervals as the horizon $h \to \infty$?

- A) They converge to a fixed width
- B) They grow without bound
- C) They shrink to zero
- D) They oscillate seasonally

Quiz 14: Forecasting Horizon

Question

For a SARIMA model with $D = 1$, what happens to forecast confidence intervals as the horizon $h \to \infty$?

- A) They converge to a fixed width
- B) They grow without bound
- C) They shrink to zero
- D) They oscillate seasonally

Answer: B – They grow without bound

Unit root property: Any unit root (regular or seasonal) causes unbounded forecast variance
For SARIMA with $D = 1$:

$$\text{Var}(\hat{Y}_{T+h} - Y_{T+h}) \to \infty \text{ as } h \to \infty$$

Intuition: Seasonal shocks accumulate over time
Practical implication: Long-range forecasts for seasonal integrated series have very wide confidence intervals

Quiz 15: Seasonal Period Selection

Question

You have daily data showing clear weekly patterns. What seasonal period s should you use in a SARIMA model?

- A) $s = 12$ (monthly)
- B) $s = 7$ (weekly)
- C) $s = 365$ (yearly)
- D) $s = 24$ (hourly)

Quiz 15: Seasonal Period Selection

Question

You have daily data showing clear weekly patterns. What seasonal period s should you use in a SARIMA model?

- A) $s = 12$ (monthly)
- B) $s = 7$ (weekly)
- C) $s = 365$ (yearly)
- D) $s = 24$ (hourly)

Answer: B – $s = 7$ (weekly)

Seasonal period selection:

Data	**Pattern**	**Period** s
Daily	Weekly	7
Daily	Annual	365 (or 252 for trading)
Monthly	Annual	12
Quarterly	Annual	4
Hourly	Daily	24

Quiz 16: Seasonal AR Component

Question

In the seasonal component $\Phi(L^s) = 1 - \Phi_1 L^s$, what does the coefficient $\Phi_1 = 0.8$ tell us?

- A) 80% of this period's value comes from the previous period
- B) There is 80% correlation between consecutive observations
- C) 80% of this period's value is explained by the same period last year
- D) The seasonal pattern explains 80% of variance

Quiz 16: Seasonal AR Component

Question

In the seasonal component $\Phi(L^s) = 1 - \Phi_1 L^s$, what does the coefficient $\Phi_1 = 0.8$ tell us?

- A) 80% of this period's value comes from the previous period
- B) There is 80% correlation between consecutive observations
- C) 80% of this period's value is explained by the same period last year
- D) The seasonal pattern explains 80% of variance

Answer: C – 80% explained by same period last year

SAR(1) equation: $(1 - \Phi_1 L^{12}) Y_t = \varepsilon_t$
Rearranged: $Y_t = \Phi_1 Y_{t-12} + \varepsilon_t$
With $\Phi_1 = 0.8$: $Y_{Jan2024} = 0.8 \cdot Y_{Jan2023} + \varepsilon_t$
Interpretation: Strong seasonal persistence — each month's value is 80% determined by the same month in the previous year
Stationarity: Requires $|\Phi_1| < 1$ (here satisfied: $|0.8| < 1$)

Quiz 17: Seasonal Stationarity

Question

A seasonal process with $\Phi_1 = 1$ in SARIMA$(0,0,0) \times (1,0,0)_{12}$ is:

- A) Stationary
- B) Has a seasonal unit root (seasonally integrated)
- C) Explosive
- D) Undefined

Quiz 17: Seasonal Stationarity

Question

A seasonal process with $\Phi_1 = 1$ in SARIMA$(0,0,0) \times (1,0,0)_{12}$ is:

- A) Stationary
- B) Has a seasonal unit root (seasonally integrated)
- C) Explosive
- D) Undefined

Answer: B – Has a seasonal unit root

Model: $(1 - L^{12})Y_t = \varepsilon_t$ when $\Phi_1 = 1$
This implies: $Y_t = Y_{t-12} + \varepsilon_t$ (seasonal random walk)
Properties:

- Variance grows with time (non-stationary)
- Each month follows its own random walk
- Need $D = 1$ to make stationary

Analogy: Like regular RW but at seasonal frequency

Quiz 18: Model Comparison

Question

Model A: SARIMA$(1,1,1) \times (1,1,1)_{12}$ has AIC $= 520$. Model B: SARIMA$(0,1,1) \times (0,1,1)_{12}$ has AIC $= 525$. Which statement is most accurate?

- A) Model A is always better since it has lower AIC
- B) Model B should be preferred due to parsimony despite higher AIC
- C) The AIC difference of 5 suggests Model A is substantially better
- D) We cannot compare models with different orders

Quiz 18: Model Comparison

Question

Model A: $\text{SARIMA}(1,1,1) \times (1,1,1)_{12}$ has $\text{AIC} = 520$. Model B: $\text{SARIMA}(0,1,1) \times (0,1,1)_{12}$ has $\text{AIC} = 525$. Which statement is most accurate?

- (A) Model A is always better since it has lower AIC
- (B) Model B should be preferred due to parsimony despite higher AIC
- (C) The AIC difference of 5 suggests Model A is substantially better
- (D) We cannot compare models with different orders

Answer: C – AIC difference of 5 suggests Model A is substantially better

AIC interpretation: Lower AIC = better balance of fit and complexity
Rule of thumb (Burnham & Anderson):

- $\Delta\text{AIC} < 2$: Models are essentially equivalent
- ΔAIC $2 - 10$: Some evidence for better model
- $\Delta\text{AIC} > 10$: Strong evidence for better model

Here: $\Delta\text{AIC} = 5$ suggests Model A meaningfully better

Quiz 19: Seasonal Patterns in Residuals

Question

After fitting a SARIMA model, you notice significant ACF spikes at lags 12 and 24 in the residuals. What does this indicate?

- A) The model is correctly specified
- B) The seasonal component is inadequate
- C) The data is not seasonal
- D) Overfitting has occurred

Quiz 19: Seasonal Patterns in Residuals

Question

After fitting a SARIMA model, you notice significant ACF spikes at lags 12 and 24 in the residuals. What does this indicate?

- A) The model is correctly specified
- B) The seasonal component is inadequate
- C) The data is not seasonal
- D) Overfitting has occurred

Answer: B – The seasonal component is inadequate

Residual diagnostics: Good residuals should be white noise (no significant ACF)
Seasonal ACF in residuals means:

- Model hasn't fully captured seasonal structure
- Consider increasing P or Q
- Verify D is correct (not over/under-differenced)

Action: Try SARIMA with higher seasonal order, check Ljung-Box at seasonal lags

Quiz 20: Practical Forecasting

Question

You're forecasting monthly retail sales with SARIMA$(0,1,1) \times (0,1,1)_{12}$. For the 13-month-ahead forecast, which historical observations are most influential?

- A) Only the most recent observation
- B) The observation from the same month last year
- C) All observations equally
- D) Only observations from the same month in all previous years

Quiz 20: Practical Forecasting

Question

You're forecasting monthly retail sales with SARIMA$(0,1,1) \times (0,1,1)_{12}$. For the 13-month-ahead forecast, which historical observations are most influential?

- A) Only the most recent observation
- B) The observation from the same month last year
- C) All observations equally
- D) Only observations from the same month in all previous years

Answer: B – The observation from the same month last year

Airline model forecast structure: $(1-L)(1-L^{12})Y_t = (1+\theta L)(1+\Theta L^{12})\varepsilon_t$
For 13-month ahead forecast:

- Most influential: Y_{T-11} (same month last year)
- Also influenced by Y_T (most recent) and Y_{T-12}
- MA shocks decay quickly, seasonal structure persists

Intuition: "Next January looks like last January, adjusted for recent trend"

True/False Questions (1-6)

Determine whether each statement is True or False:

1. The seasonal period s for quarterly data with annual patterns is $s = 4$.
2. SARIMA models can only handle one seasonal frequency.
3. If AIC selects SARIMA$(1, 1, 1) \times (1, 1, 1)_{12}$ and BIC selects the airline model, BIC is always wrong.
4. The Kruskal-Wallis test can detect seasonality without assuming normality.
5. After fitting a SARIMA model, residuals should show no significant ACF at seasonal lags.
6. Log transformation converts multiplicative seasonality to additive.

True/False Solutions (1-6)

1. **TRUE**: Quarterly data with annual cycle has $s = 4$ quarters per year.
2. **TRUE**: Standard SARIMA handles one s; multiple seasonalities need TBATS or Fourier terms.
3. **FALSE**: BIC penalizes complexity more; simpler model may be better for interpretation/forecasting.
4. **TRUE**: Kruskal-Wallis is nonparametric, comparing distributions across seasons.
5. **TRUE**: Residual ACF should be within confidence bands at ALL lags including seasonal.
6. **TRUE**: $\log(T \times S \times \varepsilon) = \log T + \log S + \log \varepsilon$ (additive form).

Problem 1: Expanding the Seasonal Difference

Exercise

Expand $(1 - L)(1 - L^{12})Y_t$ fully. What observations are involved?

Problem 1: Expanding the Seasonal Difference

Exercise

Expand $(1 - L)(1 - L^{12})Y_t$ fully. What observations are involved?

Solution

$(1 - L)(1 - L^{12}) = 1 - L - L^{12} + L^{13}$

Therefore: $(1 - L)(1 - L^{12})Y_t = Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}$

Interpretation: This is the difference of differences:

- First seasonal difference: $Y_t - Y_{t-12}$ (this year vs last year)
- Then regular difference: $(Y_t - Y_{t-12}) - (Y_{t-1} - Y_{t-13})$

Problem 2: Airline Model Expansion

Exercise

Write out the full equation for the airline model SARIMA$(0, 1, 1) \times (0, 1, 1)_{12}$:
$(1 - L)(1 - L^{12})Y_t = (1 + \theta_1 L)(1 + \Theta_1 L^{12})\varepsilon_t$

Problem 2: Airline Model Expansion

Exercise

Write out the full equation for the airline model SARIMA$(0,1,1) \times (0,1,1)_{12}$:
$(1-L)(1-L^{12})Y_t = (1+\theta_1 L)(1+\Theta_1 L^{12})\varepsilon_t$

Solution

Expand the MA side: $(1+\theta_1 L)(1+\Theta_1 L^{12}) = 1 + \theta_1 L + \Theta_1 L^{12} + \theta_1\Theta_1 L^{13}$

Full model: $Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13} = \varepsilon_t + \theta_1\varepsilon_{t-1} + \Theta_1\varepsilon_{t-12} + \theta_1\Theta_1\varepsilon_{t-13}$

Note: The cross-term $\theta_1\Theta_1 L^{13}$ is the multiplicative interaction between regular and seasonal MA components.

Problem 3: Parameter Count

Exercise

How many parameters (excluding σ^2) are in SARIMA$(2,1,1) \times (1,0,1)_4$?

Problem 3: Parameter Count

Exercise

How many parameters (excluding σ^2) are in SARIMA$(2, 1, 1) \times (1, 0, 1)_4$?

Solution

- Regular AR($p = 2$): $\phi_1, \phi_2 \Rightarrow$ 2 parameters
- Regular MA($q = 1$): $\theta_1 \Rightarrow$ 1 parameter
- Seasonal AR($P = 1$): $\Phi_1 \Rightarrow$ 1 parameter
- Seasonal MA($Q = 1$): $\Theta_1 \Rightarrow$ 1 parameter

Total: 5 parameters

Note: The differencing orders ($d = 1$, $D = 0$) don't add parameters – they're transformations applied to the data.

Problem 4: SARIMA Forecasting

Exercise

Given the airline model with $\theta_1 = -0.4$ and $\Theta_1 = -0.6$, and:

- $Y_T = 500, \ Y_{T-1} = 495, \ Y_{T-11} = 480, \ Y_{T-12} = 470$
- $\varepsilon_T = 5$, $\varepsilon_{T-11} = -3$, $\varepsilon_{T-12} = 2$

Forecast Y_{T+1}.

Problem 4: SARIMA Forecasting

Exercise

Given the airline model with $\theta_1 = -0.4$ and $\Theta_1 = -0.6$, and:

- $Y_T = 500,\ Y_{T-1} = 495,\ Y_{T-11} = 480,\ Y_{T-12} = 470$
- $\varepsilon_T = 5$, $\varepsilon_{T-11} = -3$, $\varepsilon_{T-12} = 2$

Forecast Y_{T+1}.

Solution

From the model: $Y_{T+1} = Y_T + Y_{T-11} - Y_{T-12} + \varepsilon_{T+1} + \theta_1\varepsilon_T + \Theta_1\varepsilon_{T-11} + \theta_1\Theta_1\varepsilon_{T-12}$
Setting $\mathbb{E}[\varepsilon_{T+1}] = 0$:
$\hat{Y}_{T+1} = 500 + 480 - 470 + 0 + (-0.4)(5) + (-0.6)(-3) + (-0.4)(-0.6)(2)$
$= 510 - 2 + 1.8 + 0.48 = \mathbf{510.28}$

Problem 5: Identifying Seasonal Period

Exercise

Match each data type with its typical seasonal period s:

1. Quarterly GDP data
2. Monthly retail sales
3. Weekly restaurant reservations
4. Daily electricity demand

Problem 5: Identifying Seasonal Period

Exercise

Match each data type with its typical seasonal period s:

1. Quarterly GDP data
2. Monthly retail sales
3. Weekly restaurant reservations
4. Daily electricity demand

Solution

1. Quarterly GDP: $s = 4$ (annual cycle over 4 quarters)
2. Monthly retail sales: $s = 12$ (annual cycle over 12 months)
3. Weekly restaurant reservations: $s = 7$ (weekly cycle) or $s = 52$ (annual)
4. Daily electricity demand: $s = 7$ (weekly pattern) or $s = 365$ (annual)

Note: Some series have multiple seasonal patterns (e.g., daily data may have weekly AND annual cycles).

Example: Monthly Retail Sales Analysis

Scenario

You have 5 years of monthly retail sales data showing clear December peaks and January troughs. Build an appropriate SARIMA model.

Step-by-step Approach

1. **Visual inspection**: Plot shows upward trend + strong December spikes
2. **Seasonal period**: Monthly data with annual pattern $\Rightarrow s = 12$
3. **Transformation**: Consider $\log(Y_t)$ if seasonal amplitude grows with level
4. **Differencing**: Try $(1-L)(1-L^{12})Y_t$ – check ACF/PACF
5. **Model selection**: Start with airline model, compare via AIC

Example: ACF/PACF Interpretation for Seasonal Data

Observed Patterns (after differencing)

- ACF: Significant at lags 1, 12; cuts off after lag 1 and lag 12
- PACF: Significant at lags 1, 12, 13; decays at multiples of 12

Interpretation

Regular component: ACF cuts off at 1 $\Rightarrow$ MA(1)
Seasonal component: ACF significant only at lag 12 $\Rightarrow$ seasonal MA(1)
Suggested model: SARIMA$(0, d, 1) \times (0, D, 1)_{12}$ – the airline model!

Alternative check: If PACF showed cutoff at seasonal lags instead of ACF, consider seasonal AR terms.

Example: Python Implementation

Fitting SARIMA in Python

```
from statsmodels.tsa.statespace.sarimax import SARIMAX
import pmdarima as pm

# Manual fit
model = SARIMAX(y, order=(0,1,1), seasonal_order=(0,1,1,12))
results = model.fit()
print(results.summary())

# Automatic selection
auto_model = pm.auto_arima(y, seasonal=True, m=12,
                           start_p=0, max_p=2,
                           start_q=0, max_q=2,
                           d=1, D=1,
                           trace=True)
```

Example: Interpreting SARIMA Output

Sample statsmodels Output

```
                           SARIMAX Results
==============================================================
Model:              SARIMAX(0,1,1)x(0,1,1,12)   AIC:    1348.52
                                                 BIC:    1358.21
==============================================================
                  coef    std err       z     P>|z|
--------------------------------------------------------------
ma.L1          -0.4018      0.072   -5.58     0.000
ma.S.L12       -0.5521      0.081   -6.82     0.000
sigma2       1254.3201    142.856    8.78     0.000
```

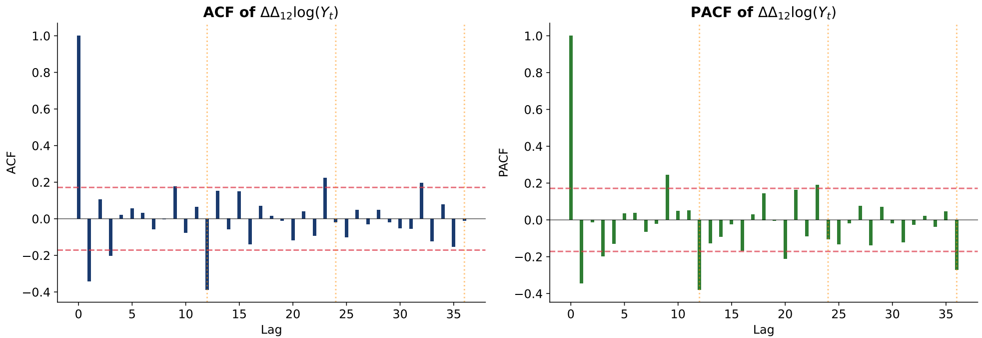
Interpretation

- $\hat{\theta}_1 = -0.40$: Negative MA means positive shocks reduce next period's value
- $\hat{\Theta}_1 = -0.55$: Same-season correlation is captured
- Both coefficients significant ($p < 0.001$); $|\theta|, |\Theta| < 1$ – invertible

Case Study: Airline Passengers (1949–1960)

- Classic Box-Jenkins dataset: 144 monthly observations
- Clear **upward trend** and **seasonal pattern** (summer peaks)
- Seasonal amplitude **grows with level** $\Rightarrow$ multiplicative seasonality
- Suggests: log transformation + SARIMA modeling

ACF/PACF Analysis After Differencing



- After $(1-L)(1-L^{12})\log(Y_t)$: series appears stationary
- Significant spike at lag 1 in ACF $\Rightarrow$ MA(1) component
- Significant spike at lag 12 in ACF $\Rightarrow$ Seasonal MA(1) component
- Pattern suggests: **SARIMA**$(0,1,1)(0,1,1)_{12}$ (airline model)

SARIMA Estimation Results: Airline Data

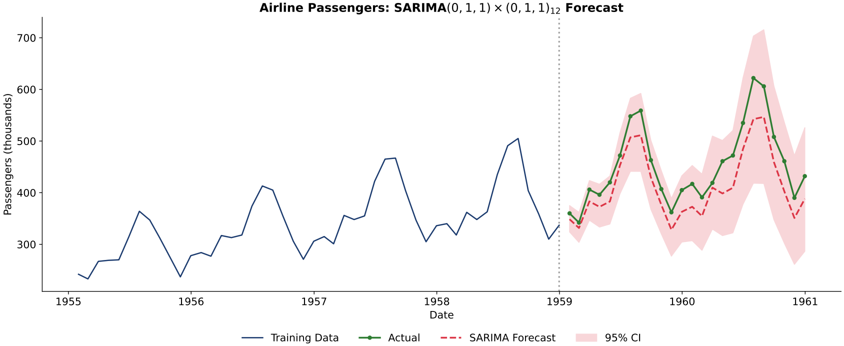
Model: SARIMA$(0,1,1)(0,1,1)_{12}$ on log(Passengers)

Parameter	Estimate	Std. Error	z-stat	p-value
θ_1 (MA.L1)	-0.4018	0.0896	-4.48	< 0.001
Θ_1 (MA.S.L12)	-0.5569	0.0731	-7.62	< 0.001
σ^2	0.00135	–	–	–

Model Fit Statistics

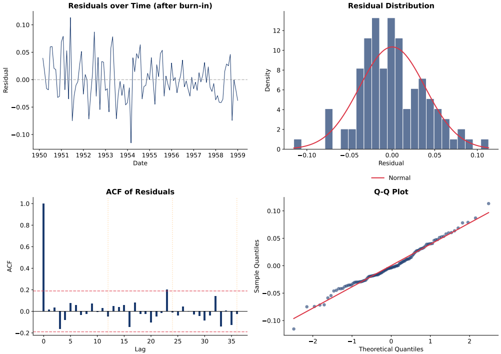
- Log-Likelihood: 244.70
- AIC: -483.40, BIC: -474.53
- Both MA coefficients significant and within invertibility bounds

Forecast: 24 Months Ahead



- Forecasts capture both trend and seasonal pattern
- 95% confidence intervals widen over forecast horizon
- Seasonal peaks (July-August) and troughs (February) clearly visible
- Model successfully extrapolates the multiplicative seasonal pattern

Model Diagnostics



- Residuals appear random with no systematic patterns
- Distribution approximately normal (Q-Q plot close to diagonal)
- ACF of residuals within confidence bounds – no significant autocorrelation
- Ljung-Box test: $p > 0.05$ at all tested lags $\Rightarrow$ adequate model

Discussion: Deterministic vs Stochastic Seasonality

Key Question

When should you use seasonal dummies vs SARIMA for seasonal data?

Considerations

Seasonal dummies (deterministic):

- Fixed, repeating pattern each year
- Same December effect every year
- Appropriate when seasonality is stable

SARIMA (stochastic):

- Evolving seasonal pattern
- This year's December depends on last year's December
- Better when seasonal amplitude varies

Discussion: Log Transformation

Key Question

When should you take logarithms before fitting SARIMA?

Guidelines

Use log transformation when:

- Seasonal fluctuations grow with the level (multiplicative seasonality)
- Variance increases over time
- Data is strictly positive (prices, sales, counts)

Avoid log when:

- Seasonal pattern is additive (constant amplitude)
- Data contains zeros or negatives
- Already on a rate/ratio scale

Tip: Compare AIC of models with and without log transformation.

Discussion: Multiple Seasonalities

Challenge

Daily sales data may have both weekly (7-day) and annual (365-day) seasonal patterns. How do you handle this?

Approaches

1. **Nested SARIMA**: Model at shorter frequency, include longer as exogenous
2. **TBATS/BATS models**: Explicitly handle multiple seasonalities
3. **Fourier terms**: Add sin/cos terms for each seasonal frequency
4. **Prophet/similar**: Modern tools designed for multiple seasonalities

Note: Standard SARIMA handles only one seasonal period. For complex seasonality, consider specialized methods.

Discussion: Forecasting Seasonal Data

Key Question

What are the unique challenges of forecasting seasonal time series?

Challenges and Solutions

- **Horizon matters**: 12-month forecast means predicting a full cycle
- **Uncertainty grows**: Seasonal forecasts compound regular uncertainty
- **Turning points**: Capturing when seasons peak/trough
- **Structural breaks**: COVID-19 disrupted many seasonal patterns

Best practices:

- Use rolling-origin cross-validation
- Compare against seasonal naive benchmark
- Report forecast intervals, especially at seasonal horizons

Take-Home Exercises

1. **Theoretical**: Show that $(1 - L)(1 - L^4)$ can be written as $(1 - L - L^4 + L^5)$ and explain what this transformation does to quarterly data with annual seasonality.

2. **Computation**: For SARIMA$(1, 0, 0) \times (1, 0, 0)_4$ with $\phi_1 = 0.5$ and $\Phi_1 = 0.8$, write out the full AR polynomial and identify all non-zero coefficients.

3. **Applied**: Download monthly airline passenger data and:
   - Plot the series and identify trend/seasonality
   - Apply appropriate transformations
   - Fit the airline model and interpret coefficients
   - Generate 24-month forecasts with confidence intervals

4. **Comparison**: Fit both SARIMA$(0, 1, 1) \times (0, 1, 1)_{12}$ and SARIMA$(1, 1, 0) \times (1, 1, 0)_{12}$ to the airline data. Compare using AIC, BIC, and residual diagnostics. Which is preferred?

Exercise Solutions Hints

Hints

1. Expand $(1-L)(1-L^4) = 1 \cdot 1 - 1 \cdot L^4 - L \cdot 1 + L \cdot L^4 = 1 - L - L^4 + L^5$
2. AR polynomial: $(1-\phi_1 L)(1-\Phi_1 L^4) = 1 - 0.5L - 0.8L^4 + 0.4L^5$
3. For airline data:
   - Use log transformation (multiplicative seasonality)
   - Both $d = 1$ and $D = 1$ needed
   - Typical estimates: $\theta_1 \approx -0.4$, $\Theta_1 \approx -0.6$
4. The MA-based airline model typically fits better than pure AR seasonal model for this data (lower AIC).

Key Takeaways from This Seminar

Main Points

1. Seasonal differencing $(1 - L^s)$ removes stochastic seasonality
2. SARIMA notation: $(p, d, q) \times (P, D, Q)_s$ separates regular and seasonal
3. The airline model is surprisingly effective for many datasets
4. Multiplicative structure creates interaction terms
5. ACF/PACF show patterns at both regular and seasonal lags
6. Log transformation often needed for multiplicative seasonality

Next Steps

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