



# Time Series Analysis and Forecasting

Seminar 4: SARIMA 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

## Seminar Outline

### Today's Activities:

- 1. Review Quiz** — Checking understanding of SARIMA concepts
- 2. True/False Questions** — Conceptual checks
- 3. Practice Problems** — Calculations with SARIMA
- 4. Worked Examples** — Real-world applications
- 5. Real Data Analysis** — Case study with airline data
- 6. Discussion Topics** — Practical applications
- 7. Summary** — Key takeaways
- 8. AI-Assisted Exercises** — Human vs. AI modeling
- 9. Key Formulas** — Essential SARIMA formulas



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

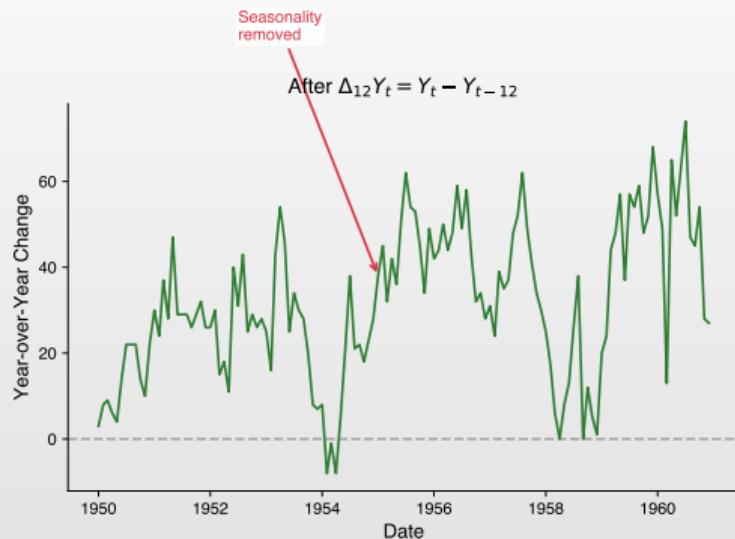
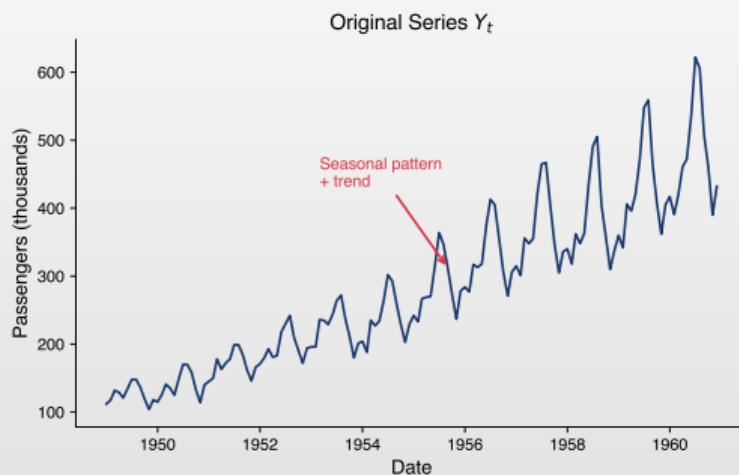
**Example (January sales):**  $Y_{Jan2025} - Y_{Jan2024}$

**Effect:** Removes stable annual seasonal pattern

**Note:**  $(1 - L^s)$  for any seasonal period  $s$  (quarterly:  $s = 4$ , weekly:  $s = 52$ )



## Visual: Seasonal Difference



Seasonal differencing removes annual patterns by comparing same periods across years.

Q [TSA\\_ch4\\_def\\_seasonal\\_diff](#)



## Quiz 2: SARIMA Notation

### Question

What does  $\text{SARIMA}(1, 1, 1) \times (1, 1, 1)_{12}$  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: Answer

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

### SARIMA( $p, d, q \times (P, D, Q)_s$ ) Notation

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

Regular (Non-Seasonal)			Seasonal		
$p$	= AR order	(Number of AR lags)	$P$	= Seasonal AR	(SAR lags at s, 2s, ...)
$d$	= Differencing	(Regular differences)	$D$	= Seasonal Diff	((1 - $L^s$ ) $D$ )
$q$	= MA order	(Number of MA lags)	$Q$	= Seasonal MA	(SMA lags at s, 2s, ...)
			$s$	= Period	(Seasonal period)

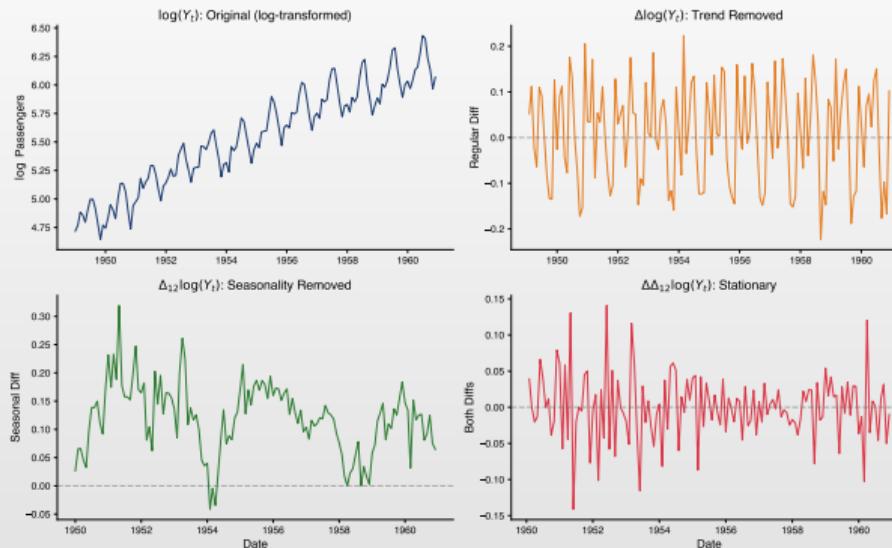
#### Example: SARIMA(1, 1, 1) $\times$ (0, 1, 1)<sub>12</sub>

Monthly data with: AR(1), MA(1), one regular diff,  
one seasonal diff at lag 12, seasonal MA(1)

$$(1 - \phi_1 L)(1 - \Phi_1 L^{12})(1 - L)(1 - L^{12}) Y_t = (1 + \theta_1 L)(1 + \Theta_1 L^{12}) \varepsilon_t$$



## Visual: SARIMA Model Structure



SARIMA combines regular ARIMA components with seasonal components at lag  $s$ .

Q TSA\_ch4\_def\_sarima

## Quiz 3: The Airline Model

### Question

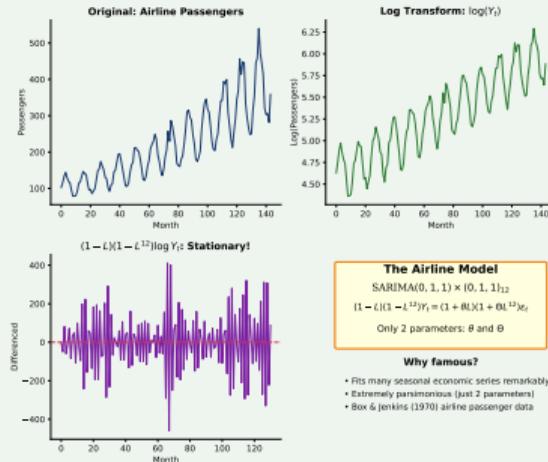
The “airline model” refers to  $\text{SARIMA}(0, 1, 1) \times (0, 1, 1)_{12}$ . How many parameters does it have (excluding variance)?

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



## Quiz 3: Answer

Answer: A – 2 parameters ( $\theta_1$  and  $\Theta_1$ )



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

Remarkably fits many seasonal economic series (Box & Jenkins, 1970)



## 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, ...

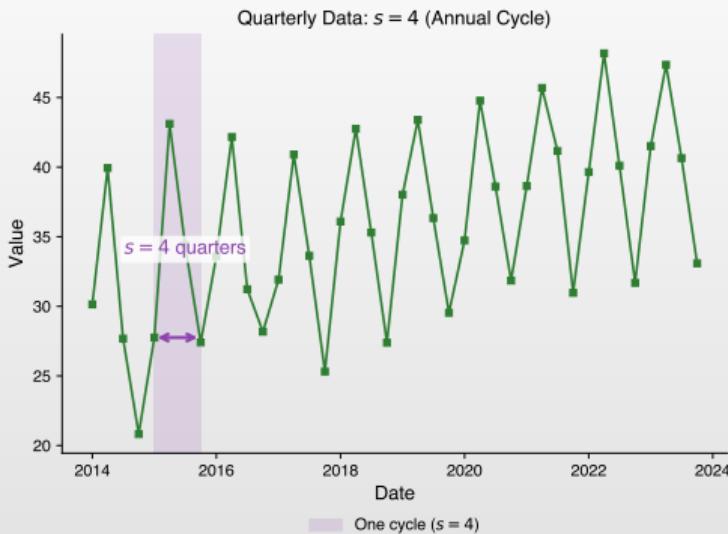
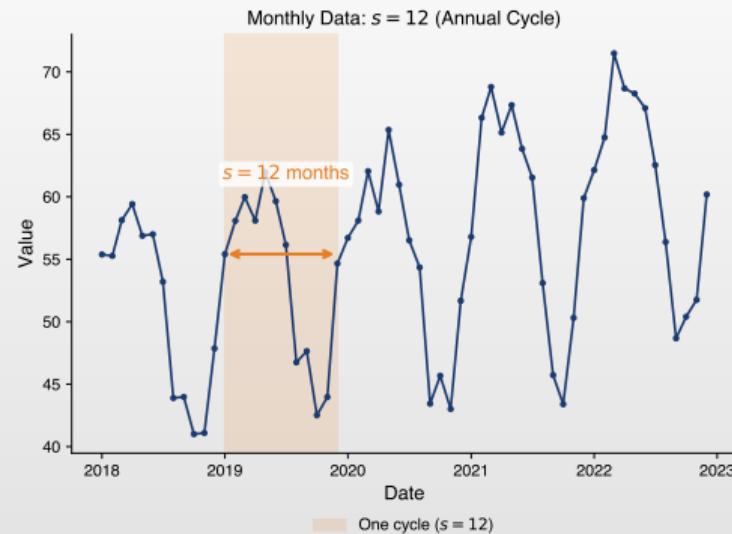
**Intuition:** January 2024 is similar to January 2023, 2022, etc.

**ACF pattern:** Spikes at lags  $s, 2s, 3s, \dots$  ( $\rho_{12}, \rho_{24}, \rho_{36} \neq 0$ )

**Diagnostic:** Slow decay at seasonal lags  $\Rightarrow D = 1$ ; Cutoff after lag  $s \Rightarrow Q = 1$



## Visual: Seasonality Patterns



Seasonal patterns repeat at regular intervals (monthly, quarterly, etc.) and may be additive or multiplicative.

[TSA\\_ch4\\_def\\_seasonality](#)



## 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:**  $\phi(L)\Phi(L^s)(1 - L)^d(1 - L^s)^D Y_t = \theta(L)\Theta(L^s)\varepsilon_t$

**Example:**  $(1 - \phi_1 L)(1 - \Phi_1 L^{12}) = 1 - \phi_1 L - \Phi_1 L^{12} + \phi_1 \Phi_1 L^{13}$

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

When needed: ACF slow decay at lags 1,2,3...  $\Rightarrow d = 1$ ; at lags 12,24,36...  $\Rightarrow D = 1$

Examples: Airline passengers, retail sales, energy demand



## Quiz 7: Detecting Seasonality from ACF

### Question

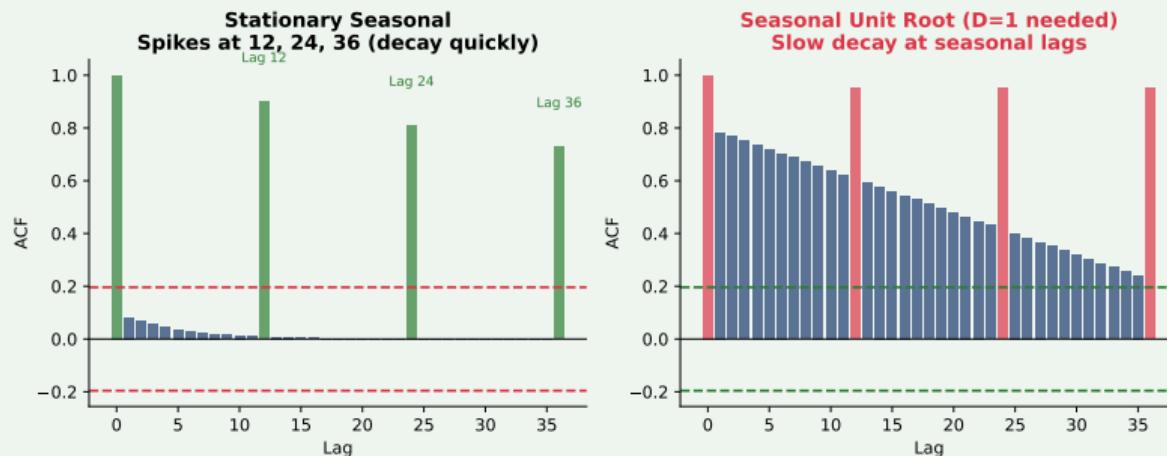
The ACF of a monthly time series shows slow decay at lags 12, 24, and 36. What does this suggest?

- 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: Answer

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



**Left:** Stationary seasonal (fast decay at seasonal lags)  
**Right:** Seasonal unit root (slow decay  $\Rightarrow$  need  $D = 1$ )



## Quiz 8: Multiplicative vs Additive Seasonality

### Question

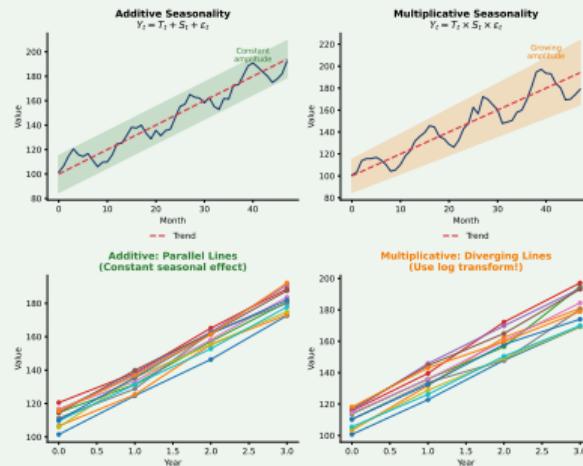
If the seasonal amplitude of a time series grows proportionally with the level, this indicates:

- 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: Answer

Answer: B – Multiplicative seasonality, use log transformation



**Multiplicative:** Seasonal amplitude grows with level (diverging lines)

**Solution:** Apply log transformation before fitting SARIMA



## 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)

**Subseries plot:** Groups data by month, plots each month's values across years

**Parallel**  $\Rightarrow$  Additive; **Diverging**  $\Rightarrow$  Multiplicative; **Horizontal**  $\Rightarrow$  No trend

**Action:** If multiplicative, apply log before fitting SARIMA



## Quiz 10: Invertibility in SARIMA

### Question

For SARIMA(0, 1, 1)  $\times$  (0, 1, 1)<sub>12</sub> 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)<sub>12</sub> 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:** All MA roots outside unit circle

**Multiplicative MA:**  $(1 + \theta_1 L)(1 + \Theta_1 L^{12})$

**Roots:** Regular  $|z| = |-1/\theta_1| > 1 \Leftrightarrow |\theta_1| < 1$ ; Seasonal  $|\Theta_1| < 1$

**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 at: Zero freq ( $\omega = 0$ )  $\Rightarrow d = 1$ ; Nyquist ( $\omega = \pi$ ); Seasonal  $\Rightarrow D = 1$

**Decision:** Reject all  $\Rightarrow$  seasonal dummies; Don't reject seasonal  $\Rightarrow$  seasonal differencing



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



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

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

**Pattern:** Regular lags – no spikes in ACF/PACF; Seasonal lags – ACF cuts off at  $s$ , PACF decays

**Interpretation:** No regular MA ( $q = 0$ ); Seasonal MA(1) indicated ( $Q = 1$ )

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



## 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

Signature: ACF at lag 1  $\approx -0.5 \Rightarrow$  over-diff at  $d$ ; ACF at lag  $s \approx -0.5 \Rightarrow$  over-diff at  $D$

Why?  $\Delta^2 Y_t = \varepsilon_t - \varepsilon_{t-1}$  is MA(1) with  $\theta = -1$ , giving  $\rho_1 = -0.5$

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  $h \rightarrow \infty$ ?

- 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  $h \rightarrow \infty$ ?

- 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 causes unbounded forecast variance

**For SARIMA with  $D = 1$ :**  $\text{Var}(\hat{Y}_{T+h} - Y_{T+h}) \rightarrow \infty$  as  $h \rightarrow \infty$

**Intuition:** Seasonal shocks accumulate; long-range forecasts have wide CIs



## 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)

Data	Pattern	Period $s$
Daily	Weekly	7
Monthly	Annual	12
Quarterly	Annual	4

Rule:  $s = \text{observations per cycle of dominant pattern}$



## 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):  $Y_t = \Phi_1 Y_{t-12} + \varepsilon_t$

With  $\Phi_1 = 0.8$ :  $Y_{Jan2024} = 0.8 \cdot Y_{Jan2023} + \varepsilon_t$

Interpretation: Strong seasonal persistence – 80% explained by same month last year

Stationarity: Requires  $|\Phi_1| < 1$  (satisfied here)



## 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  $\text{SARIMA}(0, 0, 0) \times (1, 0, 0)_{12}$  is:

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

Answer: B – Has a seasonal unit root

Model:  $Y_t = Y_{t-12} + \varepsilon_t$  (seasonal random walk)

Properties: Variance grows with time; each month follows its own RW; need  $D = 1$

Analogy: Like regular random walk but at seasonal frequency



## Quiz 18: Model Comparison

### Question

Model A: SARIMA(1, 1, 1)  $\times$  (1, 1, 1)<sub>12</sub> has AIC = 520. Model B: SARIMA(0, 1, 1)  $\times$  (0, 1, 1)<sub>12</sub> 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: SARIMA(1, 1, 1)  $\times$  (1, 1, 1)<sub>12</sub> has AIC = 520. Model B: SARIMA(0, 1, 1)  $\times$  (0, 1, 1)<sub>12</sub> 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

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

**Rule of thumb:**  $\Delta\text{AIC} < 2$ : equivalent; 2–10: some evidence;  $> 10$ : strong evidence

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

**Always:** Also check residual diagnostics and forecast performance!



## 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

**Diagnostics:** Good residuals should be white noise (no significant ACF)

**Seasonal ACF in residuals:** Model hasn't captured seasonal structure; try increasing  $P$  or  $Q$ ; verify  $D$  is correct

**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  $\text{SARIMA}(0, 1, 1) \times (0, 1, 1)_{12}$ . 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)<sub>12</sub>. 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

**For 13-month ahead:** Most influential is  $Y_{T-11}$  (same month last year), also  $Y_T$  and  $Y_{T-12}$

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



## True/False Questions (1-6)

### Question

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)<sub>12</sub> 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.

*Answer on next slide...*



## True/False Solutions (1-6)

### Answers

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

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

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

**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)<sub>12</sub>:

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



## Problem 2: Airline Model Expansion

### Exercise

Write out the full equation for the airline model SARIMA(0, 1, 1)  $\times$  (0, 1, 1)<sub>12</sub>:

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

### 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  $\sigma^2$ ) are in SARIMA(2, 1, 1)  $\times$  (1, 0, 1)<sub>4</sub>?



## Problem 3: Parameter Count

### Exercise

How many parameters (excluding  $\sigma^2$ ) are in SARIMA(2, 1, 1)  $\times$  (1, 0, 1)<sub>4</sub>?

### 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  $\mathbb{E}[\varepsilon_{T+1}] = 0$ :

$$\begin{aligned}\hat{Y}_{T+1} &= 500 + 480 - 470 + 0 + (-0.4)(5) + (-0.6)(-3) + (-0.4)(-0.6)(2) \\ &= 510 - 2 + 1.8 + 0.48 = 510.28\end{aligned}$$



## 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)<sub>12</sub> – 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

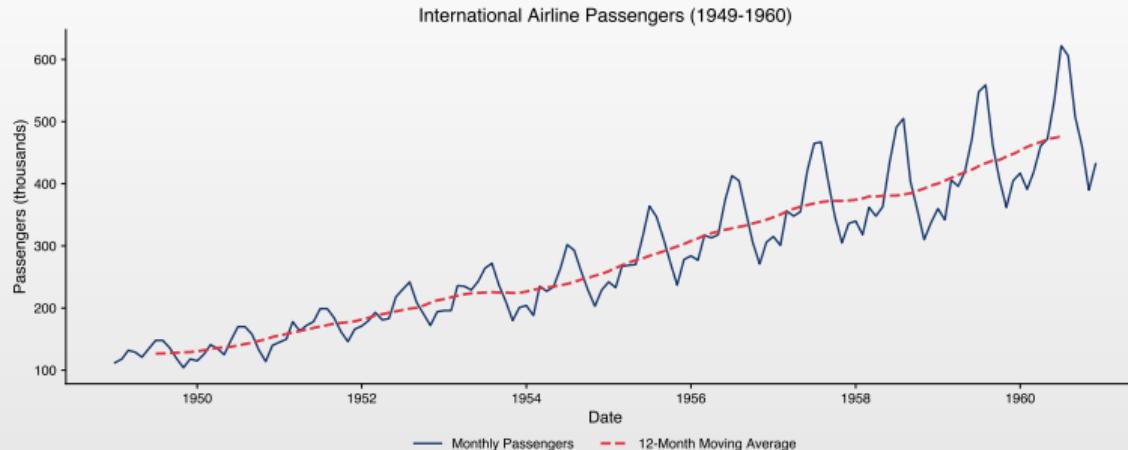
```
SARIMAX Results
=====
Model:      SARIMAX(0,1,1)x(0,1,1,12)   AIC:    1348.52
                           BIC:    1358.21
=====
            coef    std err      z     P>|z|
-----
ma.L1      -0.4018    0.072   -5.58    0.000
ma.S.L12    -0.5521    0.081   -6.82    0.000
sigma2     1254.3201  142.856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



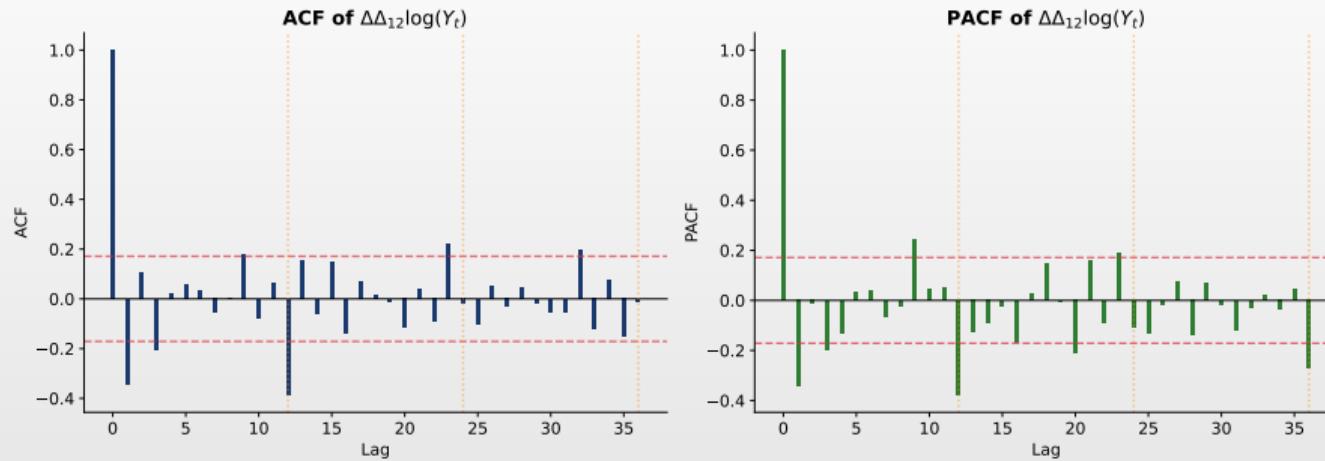
## Case Study: Airline Passengers (1949–1960)



- Classic Box-Jenkins dataset: 144 monthly observations
- Clear **upward trend** and **seasonal pattern** (summer peaks)
- Seasonal amplitude **grows with level** ⇒ 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)<sub>12</sub>** (airline model)



## SARIMA Estimation Results: Airline Data

Model: SARIMA(0, 1, 1)(0, 1, 1)<sub>12</sub> on log(Passengers)

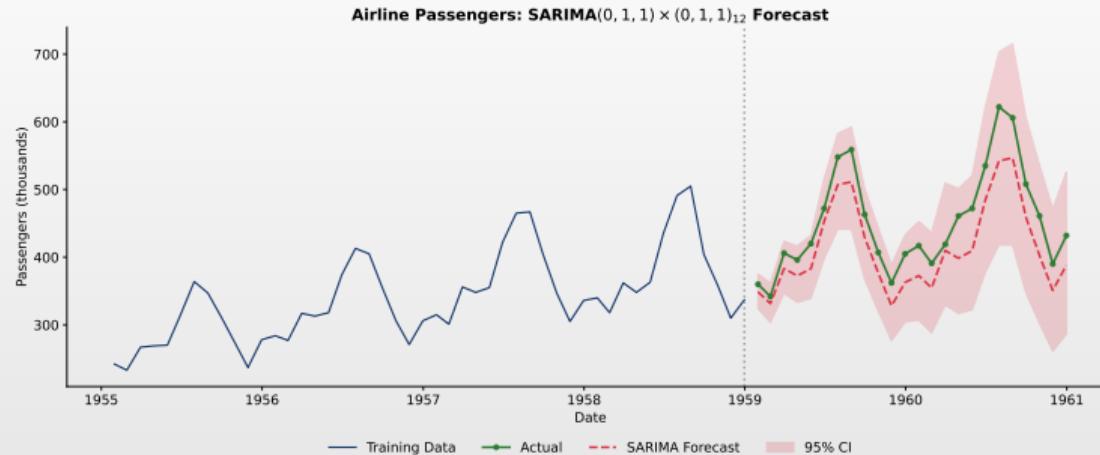
Parameter	Estimate	Std. Error	z-stat	p-value
$\theta_1$ (MA.L1)	-0.4018	0.0896	-4.48	< 0.001
$\Theta_1$ (MA.S.L12)	-0.5569	0.0731	-7.62	< 0.001
$\sigma^2$	0.00135	-	-	-

## Model Fit Statistics

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



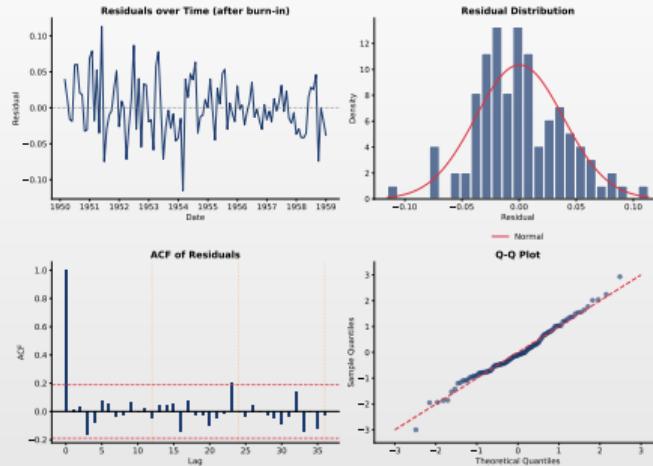
## Forecast: 24 Months Ahead



- 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)<sub>4</sub> 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)<sub>12</sub> and SARIMA(1, 1, 0)  $\times$  (1, 1, 0)<sub>12</sub> 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 Points from Today's Seminar

### What We Covered

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:** SARIMA(0, 1, 1)(0, 1, 1)<sub>12</sub> is surprisingly effective
4. **Multiplicative structure:** Creates interaction terms between components
5. **ACF/PACF diagnostics:** Patterns at both regular and seasonal lags
6. **Log transformation:** Often needed for multiplicative seasonality

### Next Seminar

GARCH and Volatility Modeling: volatility clustering, ARCH/GARCH estimation, financial risk applications (VaR).



## AI in SARIMA Modeling

### Context

AI tools can fit SARIMA models and generate seasonal forecasts automatically. The critical skill is **evaluating whether the seasonal methodology is correct**.

**Key questions to ask about any AI-generated SARIMA analysis:**

1. Did it check for multiplicative seasonality and apply log if needed?
2. Is the seasonal period  $s$  correctly identified from the data frequency?
3. Did it test for both regular and seasonal unit roots before choosing  $d$  and  $D$ ?
4. Are residuals free of seasonal patterns (check ACF at lags  $s, 2s, 3s$ )?
5. Did it compare against a seasonal naive benchmark?



## AI Exercise 1: Critique an AI SARIMA Analysis

### Scenario

You asked an AI: "Fit the best SARIMA model to this monthly retail sales data." It returned:

- Fitted SARIMA(2,1,2)(2,1,2)<sub>12</sub> with AIC = 1520
- No check for multiplicative vs additive seasonality
- Applied  $D = 1$  "because the data is monthly"
- Ljung-Box p-value = 0.03 on residuals
- 24-month forecast with constant-width confidence intervals

### Your critique:

1. Why is the model likely over-parameterized? How many parameters does it have?
2. Why should multiplicative seasonality be checked before fitting?
3. Why is Ljung-Box p = 0.03 **not** acceptable at 5% level?
4. Why must SARIMA confidence intervals widen (not stay constant)?



## AI Exercise 2: Prompt Refinement for SARIMA

### Task

Iteratively improve prompts for fitting a SARIMA model to monthly airline data.

**Round 1** (vague): “*Fit a seasonal model to monthly airline passengers*”

- What did the AI produce? Did it check for multiplicative seasonality?

**Round 2** (better): “*Check if seasonality is multiplicative (log transform if so), apply seasonal differencing, fit SARIMA using AIC, check residual ACF at seasonal lags*”

- Did the AI follow the seasonal Box-Jenkins methodology?

**Round 3** (expert): “*(1) Plot series, check if amplitude grows with level, (2) if multiplicative, apply log, (3) test ADF on levels & seasonal differences, (4) fit airline model + 2 alternatives, (5) compare AIC/BIC, (6) Ljung-Box on residuals, (7) 24-month rolling forecast with RMSE*”

- Compare results across all three rounds



## AI Exercise 3: Model Selection Competition

### Task

Download the classic airline passenger dataset (Box-Jenkins, 1970).

#### Your approach (manual):

- Check multiplicative vs additive seasonality → log if needed
- Determine  $d$  and  $D$  from ADF/KPSS and seasonal ACF
- ACF/PACF of transformed series → candidate models
- Compare airline model vs SARIMA(1, 1, 0)(1, 1, 0)<sub>12</sub> via AIC/BIC
- Rolling 1-step forecast on last 24 observations

#### AI approach:

- Ask AI to “find the best SARIMA model and forecast airline passengers”

#### Compare:

- Did the AI apply log transformation? Which model did each select?
- Compare out-of-sample RMSE; did the AI check seasonal residual patterns?
- Submit:** 1-page reflection on AI strengths and weaknesses



## Key Formulas Summary

Concept	Formula
Seasonal difference	$(1 - L^s)Y_t = Y_t - Y_{t-s}$
Combined differencing	$(1 - L)(1 - L^s)Y_t = Y_t - Y_{t-1} - Y_{t-s} + Y_{t-s-1}$
SARIMA( $p, d, q$ )( $P, D, Q$ ) $_s$	$\phi(L)\Phi(L^s)(1 - L)^d(1 - L^s)^D Y_t = \theta(L)\Theta(L^s)\varepsilon_t$
Airline model	$(1 - L)(1 - L^{12})Y_t = (1 + \theta_1 L)(1 + \Theta_1 L^{12})\varepsilon_t$
Multiplicative AR	$(1 - \phi_1 L)(1 - \Phi_1 L^s) = 1 - \phi_1 L - \Phi_1 L^s + \phi_1 \Phi_1 L^{s+1}$
Multiplicative MA	$(1 + \theta_1 L)(1 + \Theta_1 L^s) = 1 + \theta_1 L + \Theta_1 L^s + \theta_1 \Theta_1 L^{s+1}$
Invertibility	$ \theta_i  < 1$ and $ \Theta_j  < 1$ for all $i, j$
Stationarity	$ \Phi_j  < 1$ and regular AR roots outside unit circle
Log transform	$\log(T \times S \times \varepsilon) = \log T + \log S + \log \varepsilon$

Notation:  $s$  = seasonal period,  $\phi/\Phi$  = regular/seasonal AR,  $\theta/\Theta$  = regular/seasonal MA,  $d/D$  = regular/seasonal differencing order



## What's Next?

### Seminar 5: GARCH and Volatility Modeling

- Volatility clustering:** Why financial returns show time-varying variance
- ARCH/GARCH models:** Modeling conditional heteroskedasticity
- Extensions:** EGARCH, GJR-GARCH, leverage effects
- Applications:** Value-at-Risk and risk management

Questions?



## Bibliography I

### Fundamental textbooks

- Hyndman, R.J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*, 3rd ed., OTexts.
- Shumway, R.H., & Stoffer, D.S. (2017). *Time Series Analysis and Its Applications*, 4th ed., Springer.
- Brockwell, P.J., & Davis, R.A. (2016). *Introduction to Time Series and Forecasting*, 3rd ed., Springer.

### Financial time series

- Tsay, R.S. (2010). *Analysis of Financial Time Series*, 3rd ed., Wiley.
- Franke, J., Härdle, W.K., & Hafner, C.M. (2019). *Statistics of Financial Markets*, 4th ed., Springer.



## Bibliography II

### Modern approaches and Machine Learning

- Nielsen, A. (2019). *Practical Time Series Analysis*, O'Reilly Media.
- Petropoulos, F., et al. (2022). *Forecasting: Theory and Practice*, International Journal of Forecasting.
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition, International Journal of Forecasting.

### Online resources and code

- **Quantlet:** <https://quantlet.com> — Code repository for statistics
- **Quantinar:** <https://quantinar.com> — Learning platform for quantitative methods
- **GitHub TSA:** [https://github.com/QuantLet/TSA/tree/main/TSA\\_ch4](https://github.com/QuantLet/TSA/tree/main/TSA_ch4) — Python code for this seminar

