



Time Series Analysis and Forecasting

# Chapter 2: ARMA Models

Seminar



# Seminar Outline

- 1 Multiple Choice Quiz
- 2 True/False Questions
- 3 Calculation Exercises
- 4 Python Exercises
- 5 Real Data Analysis
- 6 Discussion Questions
- 7 Summary

## Quiz 1: Lag Operator

### Question

What is the result of applying  $(1 - L)^2$  to  $X_t$ ?

- A.  $X_t - X_{t-1}$
- B.  $X_t - 2X_{t-1} + X_{t-2}$
- C.  $X_t + X_{t-1} + X_{t-2}$
- D.  $X_t - X_{t-2}$

*Answer on next slide...*

## Quiz 1: Solution

Answer:  $B - X_t - 2X_{t-1} + X_{t-2}$

Explanation:

$$\begin{aligned}(1 - L)^2 X_t &= (1 - 2L + L^2) X_t \\&= X_t - 2LX_t + L^2 X_t \\&= X_t - 2X_{t-1} + X_{t-2}\end{aligned}$$

This is the second difference of  $X_t$ .

Note:  $(1 - L)$  is the first difference operator,  $(1 - L)^2$  is the second difference.

## Quiz 2: AR(1) Stationarity

### Question

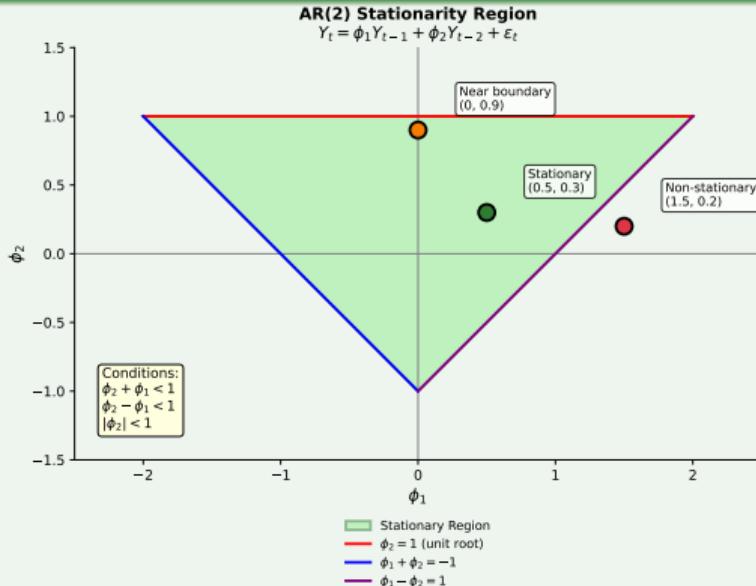
For which value of  $\phi$  is the AR(1) process  $X_t = 0.5 + \phi X_{t-1} + \varepsilon_t$  stationary?

- A.  $\phi = 1.2$
- B.  $\phi = 1.0$
- C.  $\phi = -0.8$
- D.  $\phi = -1.5$

*Answer on next slide...*

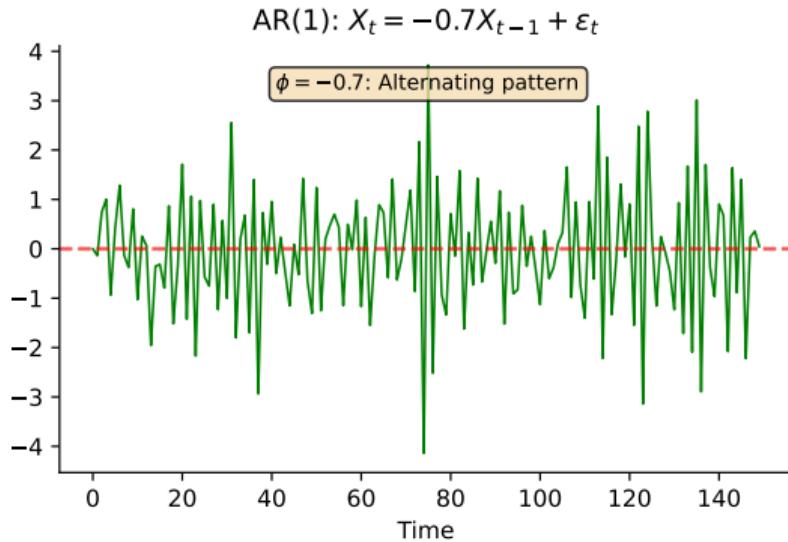
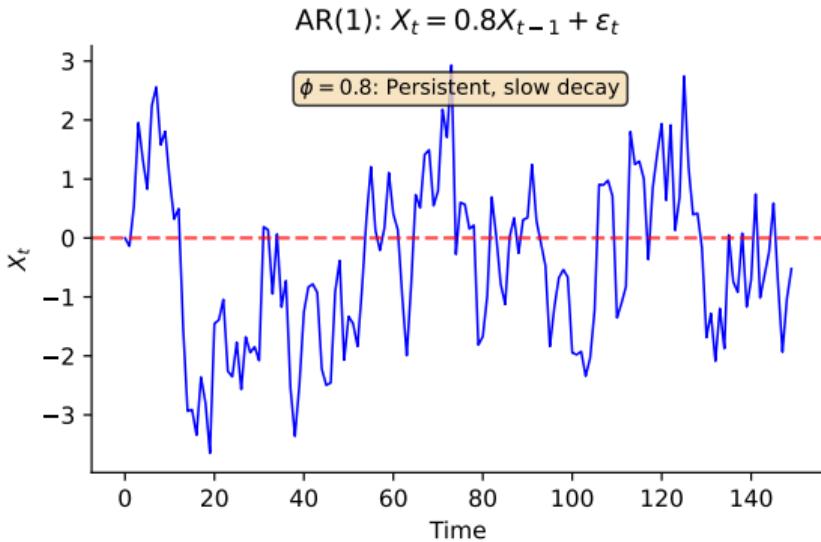
## Quiz 2: Solution

Answer:  $C - \phi = -0.8$  (Stationary)



AR(1) stationarity:  $|\phi| < 1$  (root outside unit circle). Only C satisfies:  $| - 0.8 | = 0.8 < 1$

## Visual: AR(1) Process Behavior



Positive  $\phi$ : persistent, smooth patterns. Negative  $\phi$ : oscillating behavior around mean.

## Quiz 3: ACF Pattern

### Question

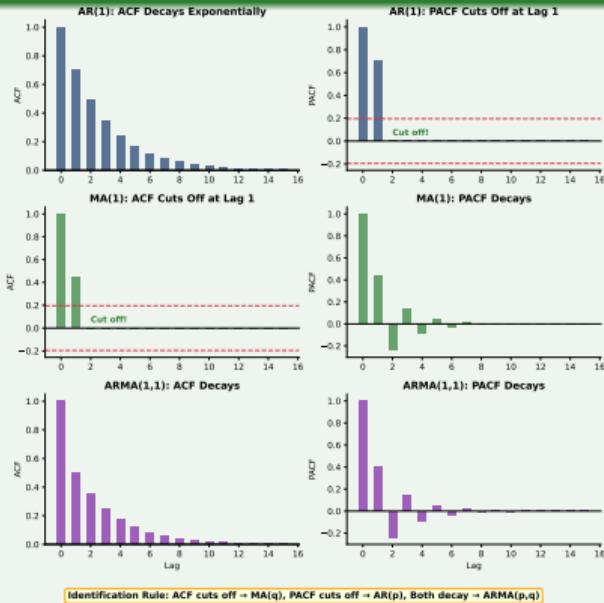
You observe the following ACF pattern: significant spike at lag 1, then all other lags are within confidence bands. The PACF shows gradual decay. What model is suggested?

- A AR(1)
- B MA(1)
- C ARMA(1,1)
- D White noise

*Answer on next slide...*

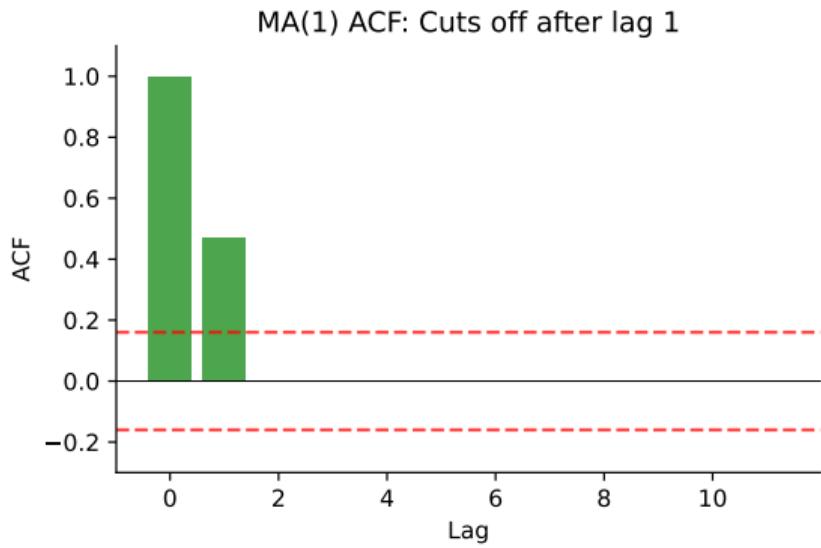
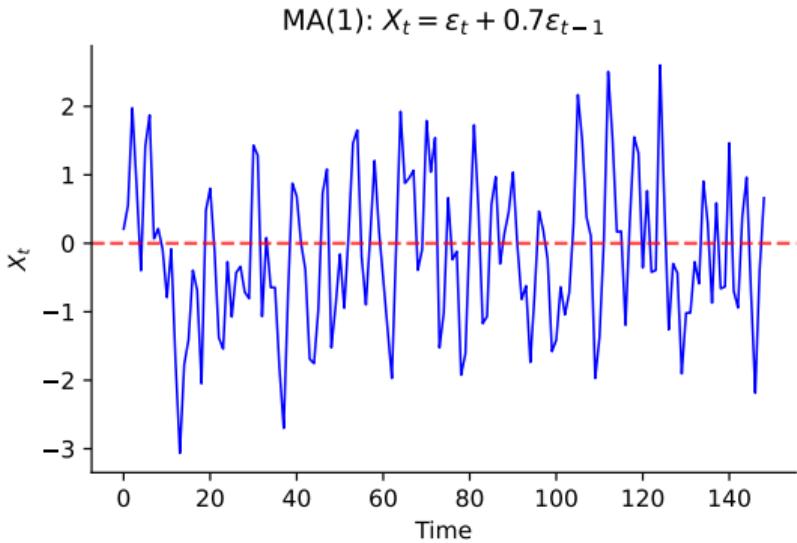
## Quiz 3: Solution

Answer: B – MA(1)



Pattern: ACF cuts off after lag 1  $\Rightarrow$  MA(1); PACF decays  $\Rightarrow$  confirms MA structure (not AR)

## Visual: MA(1) Process and ACF



MA(1) process (left). Key signature: ACF cuts off sharply after lag 1 (right).

## Quiz 4: MA Invertibility

### Question

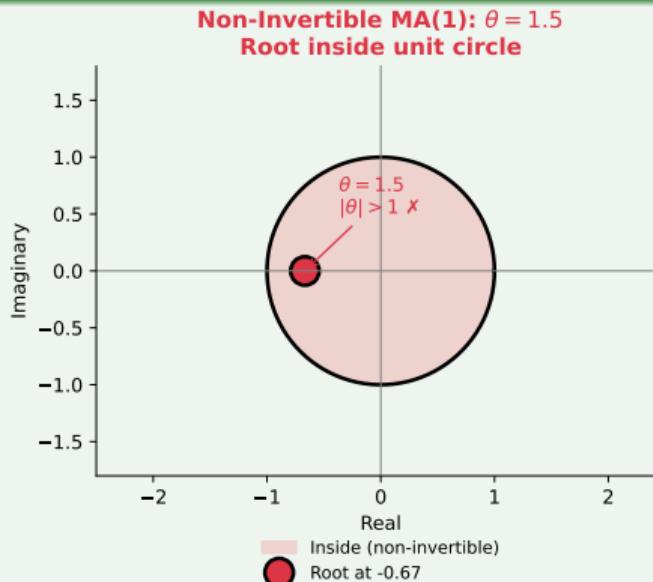
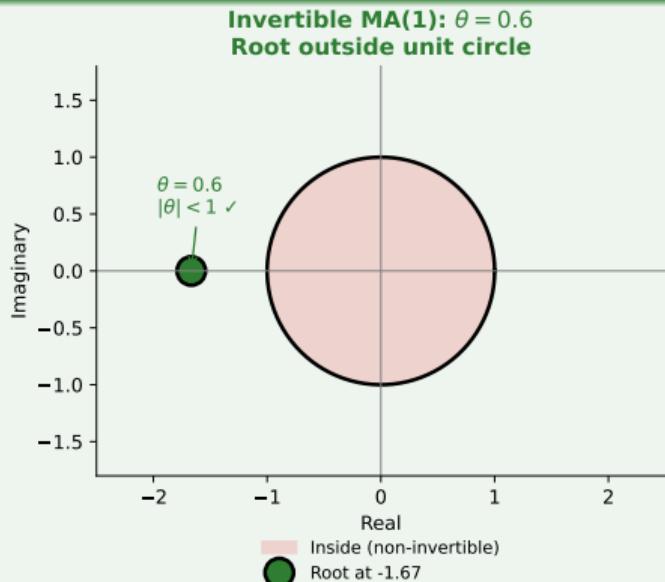
For the MA(1) process  $X_t = \varepsilon_t + 1.5\varepsilon_{t-1}$ , is the process invertible?

- A Yes, because MA processes are always invertible
- B Yes, because  $1.5 > 0$
- C No, because  $|\theta| = 1.5 > 1$
- D No, because MA processes are never invertible

*Answer on next slide...*

## Quiz 4: Solution

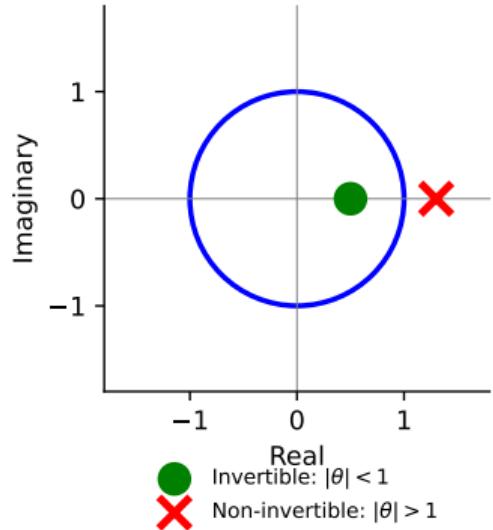
Answer: C – Not invertible ( $|\theta| = 1.5 > 1$ )



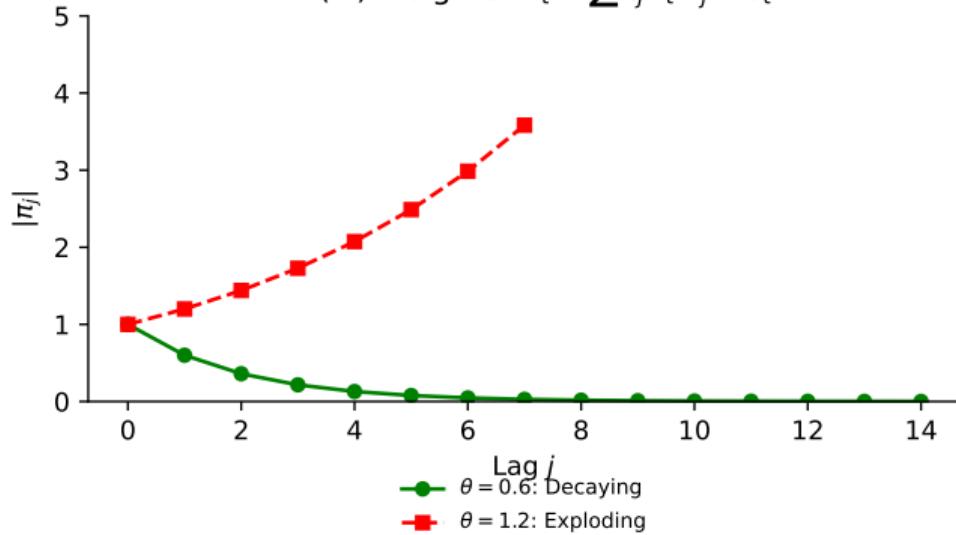
MA invertibility: Root  $z = -1/\theta$  must be outside unit circle  $\Leftrightarrow |\theta| < 1$ . Here  $z = -0.67$  is inside!

## Visual: Invertibility Concept

Invertibility: Root outside unit circle



AR( $\infty$ ) weights:  $X_t = \sum \pi_j X_{t-j} + \varepsilon_t$



Left: invertibility requires roots outside unit circle. Right: AR( $\infty$ ) weights decay only when  $|\theta| < 1$ .

## Quiz 5: ARMA Representation

### Question

The compact form  $\phi(L)X_t = \theta(L)\varepsilon_t$  represents which model?

- A Pure AR model
- B Pure MA model
- C ARMA model
- D None of the above

*Answer on next slide...*

## Quiz 5: Solution

Answer: C – ARMA model

- $\phi(L) = 1 - \phi_1L - \cdots - \phi_pL^p$  is the AR polynomial
- $\theta(L) = 1 + \theta_1L + \cdots + \theta_qL^q$  is the MA polynomial

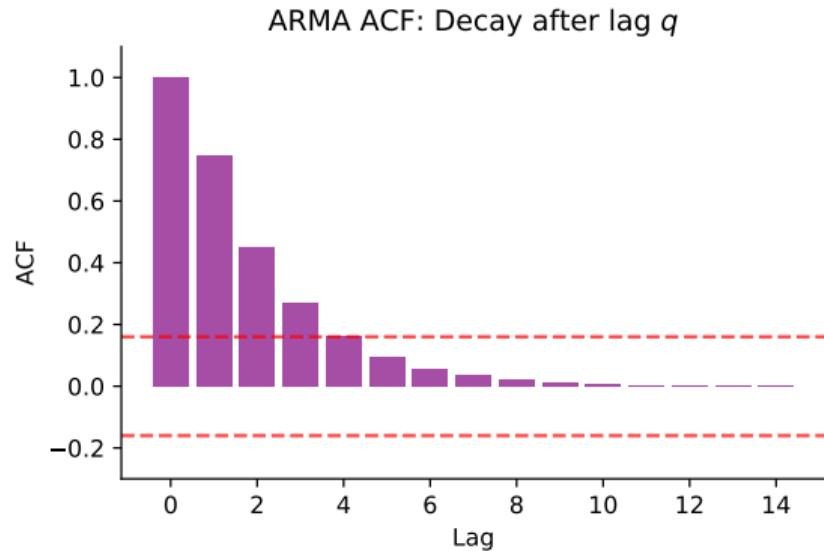
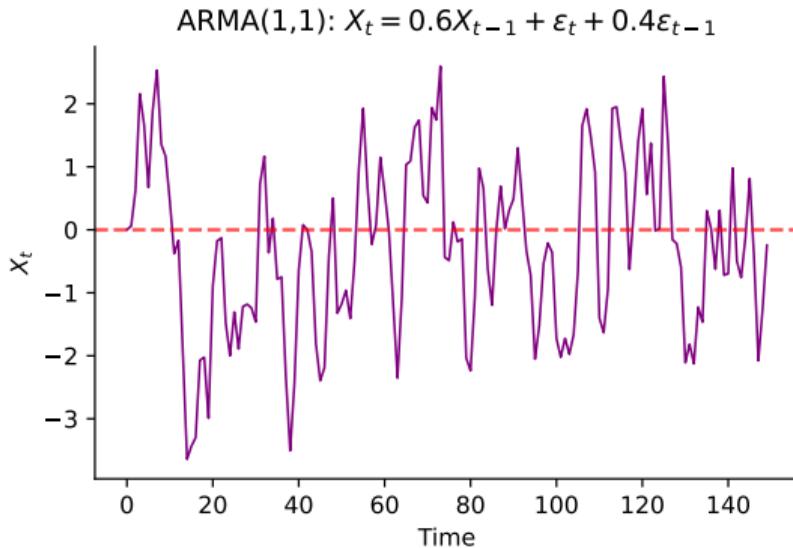
The equation  $\phi(L)X_t = \theta(L)\varepsilon_t$  expands to:

$$X_t - \phi_1X_{t-1} - \cdots - \phi_pX_{t-p} = \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q}$$

This is the general ARMA(p,q) model.

**Special cases:**  $\theta(L) = 1$  (no MA): Pure AR;  $\phi(L) = 1$  (no AR): Pure MA

## Visual: ARMA Process



ARMA(1,1) combines AR and MA components. ACF shows decay after initial lag.

## Quiz 6: Information Criteria

### Question

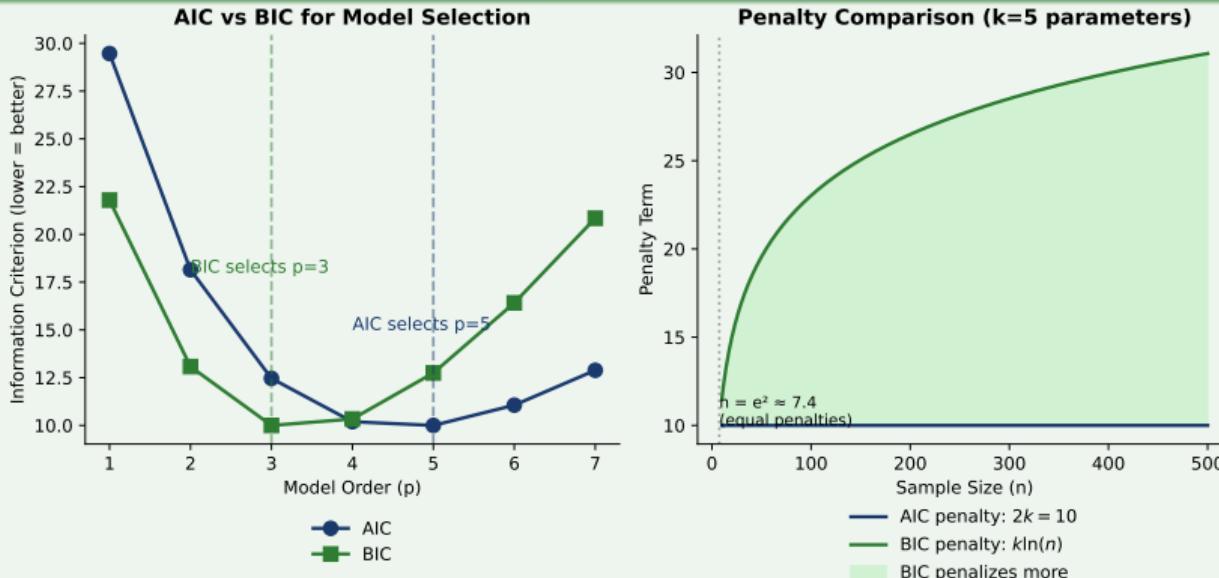
When comparing ARMA(1,1) and ARMA(2,1) using BIC, which statement is correct?

- A Lower BIC always means better forecasts
- B BIC penalizes complexity less than AIC
- C The model with lower BIC is preferred
- D BIC can only compare models with same number of parameters

*Answer on next slide...*

## Quiz 6: Solution

Answer: C – Lower BIC is preferred



$$\text{AIC: } -2 \ln(\hat{L}) + 2k \quad \text{BIC: } -2 \ln(\hat{L}) + k \ln(n) \quad \text{BIC penalizes complexity more} \Rightarrow \text{more parsimonious models}$$

## Quiz 7: Ljung-Box Test

### Question

After fitting an ARMA(2,1) model, you run the Ljung-Box test on residuals and get p-value = 0.02. What do you conclude?

- A The model is adequate
- B Residuals are white noise
- C There is significant autocorrelation in residuals
- D The model has too many parameters

*Answer on next slide...*

## Quiz 7: Solution

Answer: C – There is significant autocorrelation in residuals

The Ljung-Box test has:

- $H_0$ : Residuals are white noise (no autocorrelation)
- $H_1$ : Residuals have significant autocorrelation

With p-value = 0.02 < 0.05:

- We **reject**  $H_0$
- Conclusion: residuals are **not** white noise
- The model is **inadequate** — significant structure remains

**Next step:** Try a different model (e.g., increase  $p$  or  $q$ )

## Quiz 8: Forecasting

### Question

For an AR(1) model with  $\phi = 0.6$  and mean  $\mu = 10$ , what happens to forecasts as horizon  $h \rightarrow \infty$ ?

- A. Forecasts grow without bound
- B. Forecasts converge to 0
- C. Forecasts converge to  $\mu = 10$
- D. Forecasts oscillate forever

*Answer on next slide...*

## Quiz 8: Solution

Answer: C – Forecasts converge to  $\mu = 10$

For AR(1), the  $h$ -step ahead forecast is:

$$\hat{X}_{n+h|n} = \mu + \phi^h(X_n - \mu)$$

Since  $|\phi| = 0.6 < 1$ :

$$\lim_{h \rightarrow \infty} \phi^h = 0$$

Therefore:

$$\lim_{h \rightarrow \infty} \hat{X}_{n+h|n} = \mu + 0 \cdot (X_n - \mu) = \mu = 10$$

**Key insight:** Long-run forecasts from stationary ARMA models always converge to the unconditional mean.

## Quiz 9: AR(2) Roots

### Question

An AR(2) process has characteristic roots  $z_1 = 0.8$  and  $z_2 = -0.5$ . Is it stationary?

- A Yes, because both roots are inside the unit circle
- B No, because one root is negative
- C No, because roots must be outside the unit circle
- D Cannot determine without more information

*Answer on next slide...*

## Quiz 9: Solution

Answer: C – Roots must be outside the unit circle

For AR stationarity, roots of  $\phi(z) = 0$  must lie **outside** the unit circle, i.e.,  $|z| > 1$ .

Here:  $|z_1| = 0.8 < 1$  and  $|z_2| = 0.5 < 1$  – both **inside** unit circle.

→ **Non-stationary** (actually explosive)

Note: Equivalent condition: coefficients  $\phi_1, \phi_2$  must satisfy stationarity triangle.

## Quiz 10: MA(q) Properties

### Question

For an MA(2) process, the ACF:

- A Decays exponentially
- B Cuts off after lag 2
- C Cuts off after lag 1
- D Never cuts off

*Answer on next slide...*

## Quiz 10: Solution

Answer: B – Cuts off after lag 2

For  $\text{MA}(q)$ , the ACF is exactly zero for lags  $> q$ .

- $\text{MA}(1)$ : ACF cuts off after lag 1
- $\text{MA}(2)$ : ACF cuts off after lag 2
- $\text{MA}(q)$ : ACF cuts off after lag  $q$

This is the key identification feature: ACF cutoff  $\Rightarrow$  MA order.

Meanwhile, PACF of MA processes decays (doesn't cut off).

## Quiz 11: ARMA Parsimony

### Question

Why might ARMA(1,1) be preferred over AR(5) even if both fit equally well?

- A. ARMA models are always better
- B. Fewer parameters reduce overfitting risk
- C. AR models cannot capture trends
- D. MA components are more stable

*Answer on next slide...*

## Quiz 11: Solution

Answer: B – Fewer parameters reduce overfitting risk

**Parsimony principle:** prefer simpler models.

- ARMA(1,1): 2 parameters ( $\phi_1, \theta_1$ )
- AR(5): 5 parameters ( $\phi_1, \dots, \phi_5$ )

Fewer parameters means:

- Less risk of overfitting
- Better out-of-sample forecasts
- More interpretable model

BIC penalizes complexity more than AIC, often selecting more parsimonious models.

## Quiz 12: Residual Diagnostics

### Question

After fitting an ARMA model, the residual ACF shows a significant spike at lag 5. This suggests:

- A. The model is adequate
- B. The model may need higher order terms
- C. Residuals are white noise
- D. The data is non-stationary

*Answer on next slide...*

## Quiz 12: Solution

Answer: B – The model may need higher order terms

Good residuals should be white noise with no significant ACF.

A significant spike at lag 5 indicates remaining autocorrelation structure not captured by the model.

**Actions:**

- Consider adding AR or MA terms
- Check if AR(5) or MA(5) component helps
- Re-run Ljung-Box test after modification

## Quiz 13: Wold Decomposition

### Question

The Wold decomposition theorem states that any stationary process can be written as:

- A. A finite AR process
- B. A finite MA process
- C. An infinite MA process plus a deterministic component
- D. An ARIMA process

*Answer on next slide...*

## Quiz 13: Solution

Answer: C – An infinite MA process plus a deterministic component

Wold's theorem: Any stationary process can be written as:

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + \eta_t$$

where  $\eta_t$  is deterministic and  $\sum \psi_j^2 < \infty$ .

**Implication:** MA( $\infty$ ) is the most general representation. ARMA models are parsimonious approximations to this infinite MA.

## Quiz 14: Unit Root vs Trend Stationary

### Question

How do you make a unit root process stationary?

- A. Subtract a linear trend
- B. Take first differences
- C. Apply moving average
- D. Use seasonal adjustment

*Answer on next slide...*

## Quiz 14: Solution

Answer: B – Take first differences

- **Unit root** (stochastic trend): Use **differencing**
- **Trend stationary** (deterministic trend): Use **detrending** (regression)

For random walk  $X_t = X_{t-1} + \varepsilon_t$ :

$$\Delta X_t = X_t - X_{t-1} = \varepsilon_t$$

which is stationary white noise.

## True/False Questions

### Question

Determine if each statement is True or False:

- ① An AR(2) process can exhibit pseudo-cyclical behavior.
- ② MA processes require a stationarity condition.
- ③ The PACF of an AR( $p$ ) process cuts off after lag  $p$ .
- ④ If AIC selects ARMA(2,1) and BIC selects ARMA(1,1), they cannot both be correct.
- ⑤ Forecast confidence intervals narrow as the forecast horizon increases.
- ⑥ The Yule-Walker equations can be used to estimate MA parameters.

*Answer on next slide...*

## True/False: Solutions

### Answers

- ① **TRUE:** AR(2) with complex roots shows damped oscillations
- ② **FALSE:** MA processes are always stationary; they need *invertibility* condition
- ③ **TRUE:** This is the key identification feature of AR( $p$ )
- ④ **FALSE:** Both can be “correct” — they optimize different criteria (AIC favors fit, BIC favors parsimony)
- ⑤ **FALSE:** Confidence intervals *widen* as horizon increases (more uncertainty)
- ⑥ **FALSE:** Yule-Walker is for AR models only; MA uses MLE

## Exercise 1: AR(1) Properties

**Problem:** Consider the AR(1) process:

$$X_t = 2 + 0.7X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim WN(0, 9)$$

Calculate:

- ① The mean  $\mu$
- ② The variance  $\gamma(0)$
- ③ The autocovariance  $\gamma(1)$  and  $\gamma(2)$
- ④ The autocorrelation  $\rho(1)$  and  $\rho(2)$

## Exercise 1: Solution

Given:  $c = 2$ ,  $\phi = 0.7$ ,  $\sigma^2 = 9$

### 1. Mean:

$$\mu = \frac{c}{1 - \phi} = \frac{2}{1 - 0.7} = \frac{2}{0.3} = 6.67$$

### 2. Variance:

$$\gamma(0) = \frac{\sigma^2}{1 - \phi^2} = \frac{9}{1 - 0.49} = \frac{9}{0.51} = 17.65$$

### 3. Autocovariance:

$$\gamma(1) = \phi \cdot \gamma(0) = 0.7 \times 17.65 = 12.35$$

$$\gamma(2) = \phi^2 \cdot \gamma(0) = 0.49 \times 17.65 = 8.65$$

### 4. Autocorrelation:

$$\rho(1) = \phi = 0.7, \quad \rho(2) = \phi^2 = 0.49$$

## Exercise 2: MA(1) Properties

**Problem:** Consider the MA(1) process:

$$X_t = 5 + \varepsilon_t - 0.4\varepsilon_{t-1}, \quad \varepsilon_t \sim WN(0, 4)$$

Calculate:

- ① The mean  $\mu$
- ② The variance  $\gamma(0)$
- ③ The autocovariance  $\gamma(1)$
- ④ The autocorrelation  $\rho(1)$
- ⑤ Is this process invertible?

## Exercise 2: Solution

Given:  $\mu = 5$ ,  $\theta = -0.4$ ,  $\sigma^2 = 4$

### 1. Mean:

$$\mathbb{E}[X_t] = \mu = 5$$

### 2. Variance:

$$\gamma(0) = \sigma^2(1 + \theta^2) = 4(1 + 0.16) = 4 \times 1.16 = 4.64$$

### 3. Autocovariance at lag 1:

$$\gamma(1) = \theta\sigma^2 = -0.4 \times 4 = -1.6$$

### 4. Autocorrelation:

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{-1.6}{4.64} = -0.345$$

### 5. Invertibility: $|\theta| = 0.4 < 1 \rightarrow \text{Yes, invertible}$

## Exercise 3: Characteristic Roots

**Problem:** Consider the AR(2) process:

$$X_t = 0.5X_{t-1} + 0.3X_{t-2} + \varepsilon_t$$

- ① Write the characteristic equation
- ② Find the characteristic roots
- ③ Is this process stationary?

## Exercise 3: Solution

### 1. Characteristic equation:

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 = 1 - 0.5z - 0.3z^2 = 0$$

Or:  $0.3z^2 + 0.5z - 1 = 0$

### 2. Roots (using quadratic formula):

$$z = \frac{-0.5 \pm \sqrt{0.25 + 1.2}}{0.6} = \frac{-0.5 \pm 1.204}{0.6}$$

$$z_1 = \frac{0.704}{0.6} = 1.17, \quad z_2 = \frac{-1.704}{0.6} = -2.84$$

### 3. Stationarity check:

Both roots have  $|z| > 1$ :  $|z_1| = 1.17 > 1$  and  $|z_2| = 2.84 > 1$

→ **Stationary** (roots outside unit circle)

## Exercise 4: Forecasting

**Problem:** You have fit an AR(1) model:

$$X_t = 3 + 0.8X_{t-1} + \varepsilon_t, \quad \sigma^2 = 4$$

Given  $X_{100} = 20$ , calculate:

- ① The 1-step ahead forecast  $\hat{X}_{101|100}$
- ② The 2-step ahead forecast  $\hat{X}_{102|100}$
- ③ The long-run forecast  $\hat{X}_{100+h|100}$  as  $h \rightarrow \infty$
- ④ The 95% confidence interval for  $\hat{X}_{101|100}$

## Exercise 4: Solution

Given:  $c = 3$ ,  $\phi = 0.8$ ,  $\sigma^2 = 4$ ,  $X_{100} = 20$

Mean:  $\mu = \frac{3}{1-0.8} = 15$

### 1. One-step forecast:

$$\hat{X}_{101|100} = c + \phi X_{100} = 3 + 0.8 \times 20 = 19$$

### 2. Two-step forecast:

$$\hat{X}_{102|100} = c + \phi \hat{X}_{101|100} = 3 + 0.8 \times 19 = 18.2$$

### 3. Long-run forecast:

$$\lim_{h \rightarrow \infty} \hat{X}_{100+h|100} = \mu = 15$$

### 4. 95% CI for 1-step:

$$\text{MSFE}(1) = \sigma^2 = 4, \quad \sqrt{\text{MSFE}(1)} = 2$$

$$CI : 19 \pm 1.96 \times 2 = [15.08, 22.92]$$

# Python Exercise 1: Simulate and Fit AR(1)

## Task:

- ① Simulate 500 observations from AR(1) with  $\phi = 0.7$
- ② Plot the series and ACF/PACF
- ③ Fit an AR(1) model and check if  $\hat{\phi} \approx 0.7$
- ④ Examine residual diagnostics

## Hint code:

```
np.random.seed(42)
n = 500
phi = 0.7
x = np.zeros(n)
for t in range(1, n):
    x[t] = phi * x[t-1] + np.random.randn()
```

## Python Exercise 2: Model Selection

### Task:

- ① Load a real time series (e.g., stock returns)
- ② Check stationarity using ADF test
- ③ Compare AIC/BIC for ARMA(1,0), ARMA(0,1), ARMA(1,1), ARMA(2,1)
- ④ Select the best model
- ⑤ Generate forecasts with confidence intervals

### Key functions:

- `adfuller()` for stationarity test
- `ARIMA(data, order=(p,0,q)).fit()` for fitting
- `results.aic, results.bic` for criteria
- `results.get_forecast(h)` for predictions

# Python Exercise 3: Diagnostic Checking

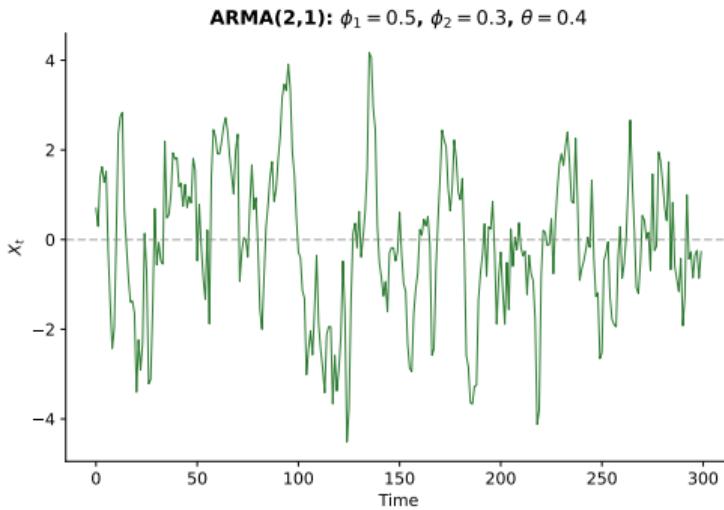
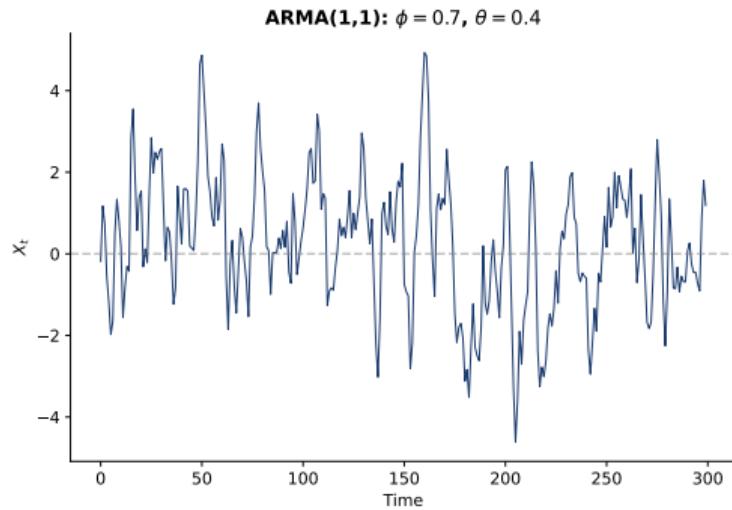
**Task:** After fitting a model, perform complete diagnostics:

- ① Plot residuals over time
- ② Plot ACF of residuals
- ③ Create Q-Q plot
- ④ Run Ljung-Box test
- ⑤ Check if AR/MA roots are outside unit circle

**Key functions:**

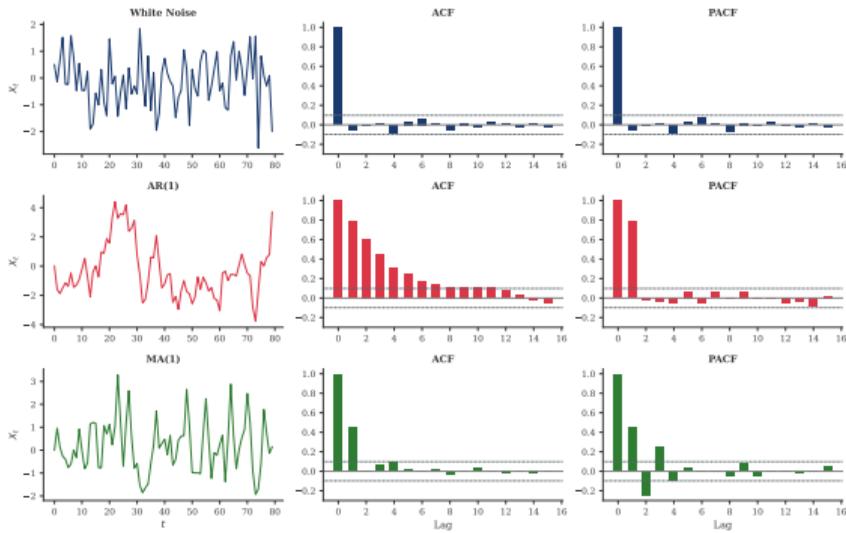
- `results.resid` for residuals
- `plot_acf(resid)` for ACF plot
- `stats.probplot(resid)` for Q-Q plot
- `acorr_ljungbox(resid)` for portmanteau test
- `results.arroots, results.maroots` for roots

## Case Study: Industrial Production Index



- US Industrial Production: monthly data, already stationary (growth rates)
- Shows typical ARMA patterns: mean-reverting with short-term dependence
- Volatility clustering visible – ARMA captures the conditional mean
- Suitable for ARMA modeling without differencing

# ACF/PACF Pattern Recognition



- ACF shows gradual decay – suggests AR component
- PACF cuts off after lag 2 – suggests AR(2) might be appropriate
- Some significant lags in ACF beyond lag 2 – MA terms may help
- Pattern consistent with ARMA(2,1) or similar low-order models

## ARMA Estimation Results

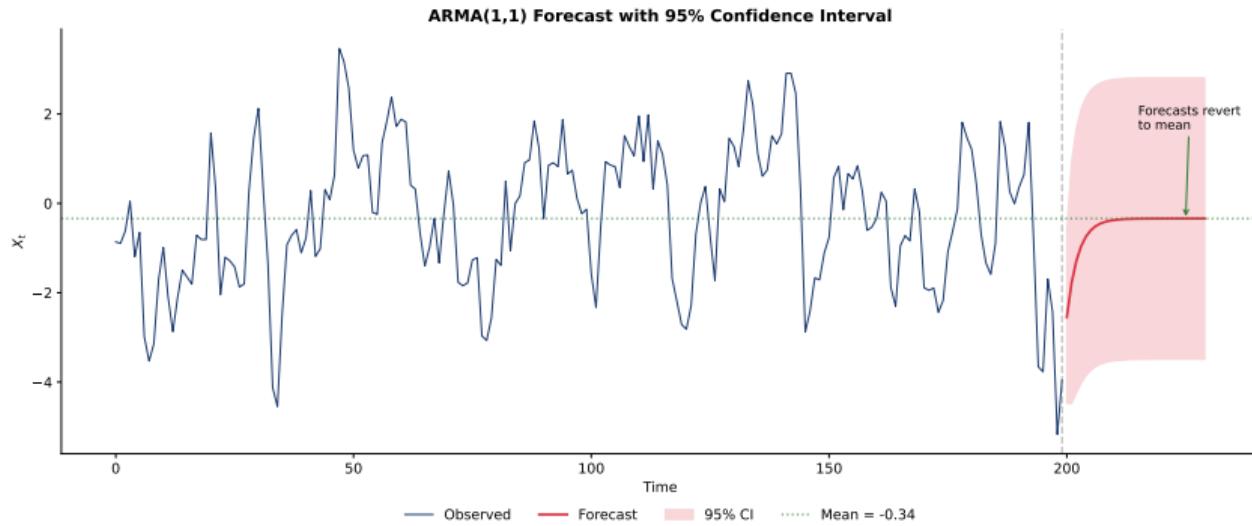
Model: ARMA(2,1) for Industrial Production Growth

Parameter	Estimate	Std. Error	z-stat	p-value
$c$ (const)	0.156	0.048	3.25	0.001
$\phi_1$ (AR.L1)	0.423	0.089	4.75	< 0.001
$\phi_2$ (AR.L2)	0.187	0.072	2.60	0.009
$\theta_1$ (MA.L1)	-0.156	0.091	-1.71	0.087

## Model Selection

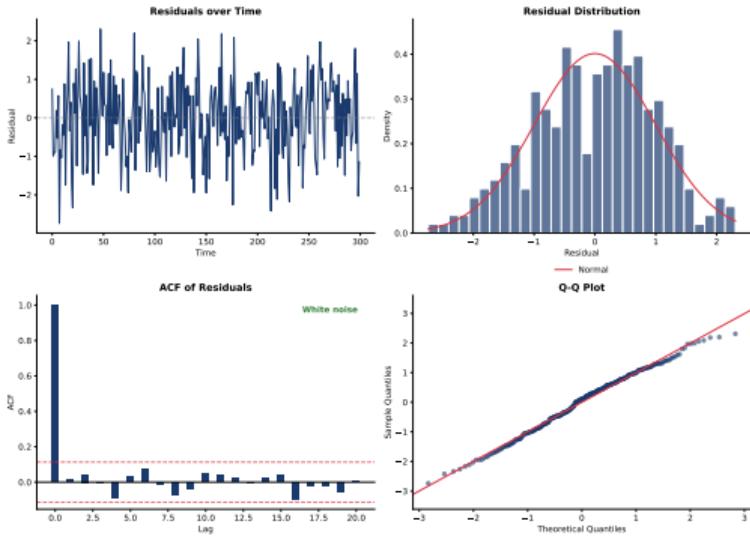
AIC: -412.5, BIC: -398.2. Model passes stationarity and invertibility checks.

## Forecast Performance



- ARMA forecasts mean-revert toward unconditional mean
- Short-term forecasts capture recent dynamics
- Confidence intervals expand with forecast horizon
- Comparison with naive forecast shows ARMA improvement

# Model Diagnostics



- Residuals appear random with no systematic patterns
- ACF of residuals within confidence bounds
- Q-Q plot shows approximate normality
- Ljung-Box test:  $p > 0.05$  – no significant autocorrelation in residuals

## Discussion 1: Model Selection

**Scenario:** You're modeling monthly inflation rates. After checking stationarity (passed), you find:

- ACF: significant at lags 1, 2, 3, then decays
- PACF: significant at lags 1, 2, then cuts off
- AIC selects ARMA(2,3)
- BIC selects ARMA(2,0) = AR(2)

**Questions:**

- ① What does the ACF/PACF pattern suggest?
- ② Why do AIC and BIC disagree?
- ③ Which model would you choose and why?
- ④ What additional checks would you perform?

## Discussion 2: Forecast Evaluation

**Scenario:** You fit an ARMA(1,1) model to daily stock returns. The in-sample fit looks good (Ljung-Box p-value = 0.45), but out-of-sample RMSE is worse than a simple random walk forecast.

### Questions:

- ① Is this surprising? Why or why not?
- ② What does this tell us about stock return predictability?
- ③ Should you conclude the ARMA model is useless?
- ④ What alternatives might you consider?

**Hint:** Think about the Efficient Market Hypothesis and what ARMA captures vs. what it doesn't (e.g., volatility clustering).

## Discussion 3: Real-World Application

**Scenario:** A central bank economist asks you to forecast quarterly GDP growth for policy planning.

**Questions:**

- ① What preliminary analysis would you do before fitting ARMA?
- ② GDP is often non-stationary — how would you handle this?
- ③ Would you use AIC or BIC for model selection? Why?
- ④ How would you communicate forecast uncertainty to policymakers?
- ⑤ What limitations of ARMA models should you mention?

# Key Takeaways from Today's Seminar

- ① **AR models:** Current value depends on past values
  - Stationarity:  $|\phi| < 1$  for AR(1)
  - PACF cuts off at lag  $p$
- ② **MA models:** Current value depends on past shocks
  - Always stationary; invertibility:  $|\theta| < 1$  for MA(1)
  - ACF cuts off at lag  $q$
- ③ **Model selection:** Use ACF/PACF patterns + information criteria
- ④ **Diagnostics:** Residuals must be white noise (Ljung-Box test)
- ⑤ **Forecasting:** Point forecasts converge to mean; uncertainty grows

Next Seminar: ARIMA and Seasonal Models

## References

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

# Data Sources and Software

## Software Tools:

- statsmodels – ARIMA models for Python
- pmdarima – Automatic ARIMA selection
- pandas – Time series data manipulation
- matplotlib – Visualization

## Data and Examples:

- Simulated AR, MA, and ARMA processes
- Examples based on Hyndman & Athanasopoulos (2021)

# Thank You!

Questions?