Time Series Analysis and Forecasting

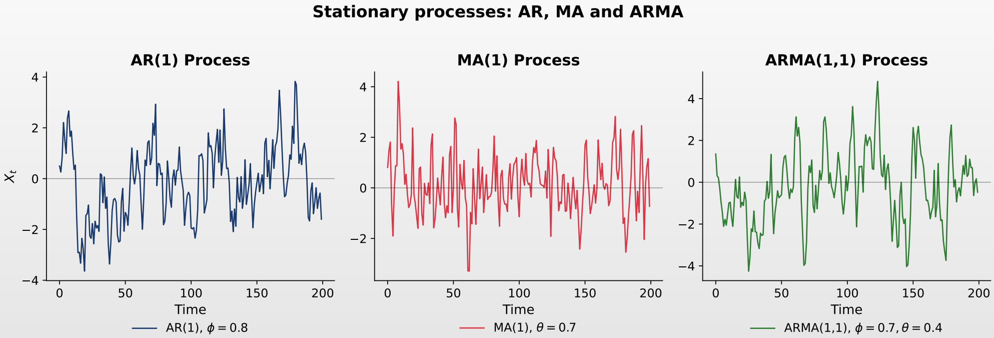
Chapter 2: ARMA Models

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

Outline

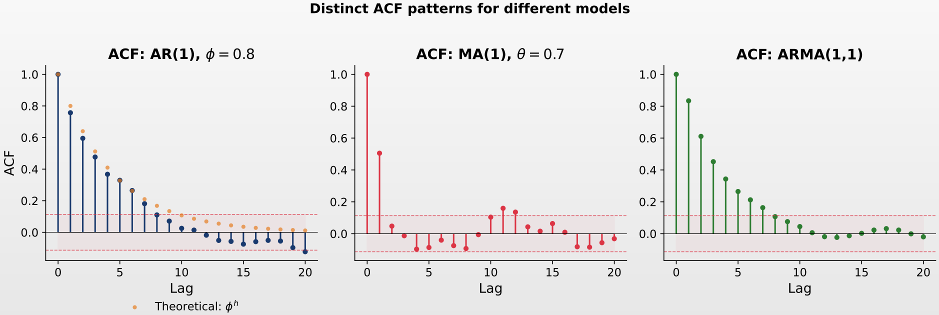
Motivating Example: Stationary Processes



- **AR processes**: Current value depends on past values — mean-reverting behavior
- **MA processes**: Current value depends on past shocks — short memory
- **ARMA**: Combines both mechanisms for flexible modeling

TSA_ch2_motivation

Model Identification via ACF Patterns



The ACF Reveals Model Structure

- Different ARMA models produce distinct ACF patterns
- We can identify the model by examining the data!

TSA_ch2_motivation

Recap: Stationarity

From Chapter 1: A process $\{X_t\}$ is **weakly stationary** if:

1. $\mathbb{E}[X_t] = \mu$ (constant mean)
2. $\text{Var}(X_t) = \sigma^2 < \infty$ (constant, finite variance)
3. $\text{Cov}(X_t, X_{t+h}) = \gamma(h)$ (covariance depends only on lag h)

Why stationarity matters for ARMA:

- ARMA models assume the underlying process is stationary
- Non-stationary data must be differenced first (ARIMA)
- Stationarity ensures stable model parameters

Today: We build models for stationary time series using past values and past errors.

The Lag Operator (Backshift Operator)

Definition 1 (Lag Operator)

The **lag operator** L (or backshift operator B) shifts a time series back by one period:

$$LX_t = X_{t-1}$$

Properties:

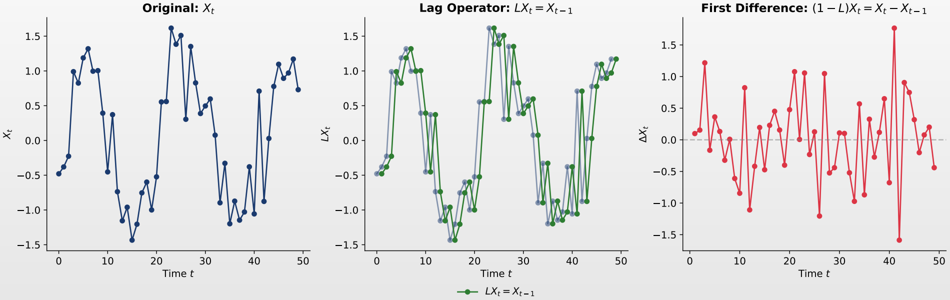
- $L^k X_t = X_{t-k}$ (shift back k periods)
- $L^0 X_t = X_t$ (identity)
- $(1 - L)X_t = X_t - X_{t-1} = \Delta X_t$ (first difference)
- $(1 - L)^d X_t = \Delta^d X_t$ (d-th difference)

Lag Polynomials:

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$$
$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$$

TSA_ch2_lag_operator

Lag Operator: Visual Illustration



Key Insight

- L shifts observations back: $LX_t = X_{t-1}$; simplifies ARMA expressions

White Noise Process

Definition 2 (White Noise)

A process $\{\varepsilon_t\}$ is **white noise**, denoted $\varepsilon_t \sim WN(0, \sigma^2)$, if:

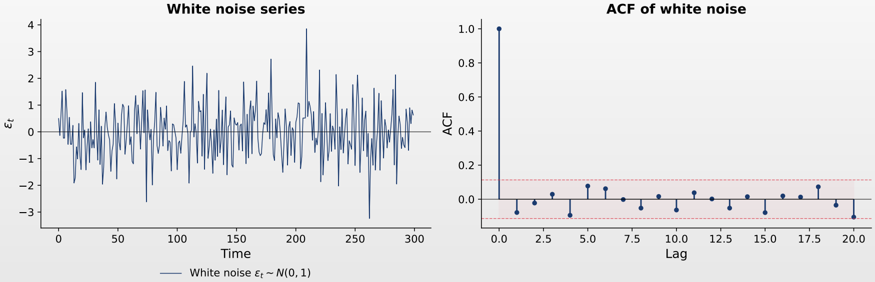
1. $\mathbb{E}[\varepsilon_t] = 0$ for all t
2. $\mathrm{Var}(\varepsilon_t) = \sigma^2$ for all t
3. $\mathrm{Cov}(\varepsilon_t, \varepsilon_s) = 0$ for all $t \neq s$

Properties:

- White noise is the "building block" of ARMA models
- ACF: $\rho(0) = 1$, $\rho(h) = 0$ for $h \neq 0$
- PACF: same pattern
- **Gaussian white noise:** additionally $\varepsilon_t \sim N(0, \sigma^2)$

Note: White noise is *not* predictable — it's pure randomness.

White Noise: Visual Illustration



Key Characteristics

- **Left**: Series fluctuates randomly around mean zero with no patterns
- **Right**: ACF shows only spike at lag 0; all others within confidence bounds

AR(1) Model: Definition

Definition 3 (AR(1) Process)

An **autoregressive process of order 1** is:

$$X_t = c + \phi X_{t-1} + \varepsilon_t$$

where $\varepsilon_t \sim WN(0, \sigma^2)$ and $|\phi| < 1$ for stationarity.

Interpretation:

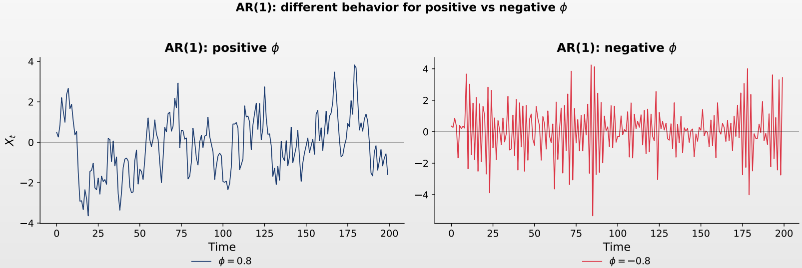
- c: constant (intercept)
- ϕ: autoregressive coefficient — measures persistence
- ε_t: innovation (unpredictable shock)

Using lag operator:

$$(1 - \phi L)X_t = c + \varepsilon_t$$

$$\phi(L)X_t = c + \varepsilon_t \quad \text{where } \phi(L) = 1 - \phi L$$

AR(1): Visual Illustration



- **Positive** ϕ: Persistent, smooth fluctuations; gradual mean reversion
- **Negative** ϕ: Oscillating behavior; rapid sign changes
- Higher $|\phi| \Rightarrow$ slower mean reversion, more persistence

TSA_ch2_ar1_simulation

AR(1) Stationarity Condition

For AR(1) to be stationary: $|\phi| < 1$

Intuition:

- If $|\phi| < 1$: shocks decay over time $\rightarrow$ stationary
- If $|\phi| = 1$: random walk $\rightarrow$ non-stationary (unit root)
- If $|\phi| > 1$: explosive process $\rightarrow$ non-stationary

Characteristic equation:

$$\phi(z) = 1 - \phi z = 0 \implies z = \frac{1}{\phi}$$

Stationarity requires the root $z = 1/\phi$ to lie **outside the unit circle**, i.e., $|z| > 1$, which means $|\phi| < 1$.

AR(1) Properties

For a stationary AR(1) with $|\phi| < 1$:

Mean:

$$\mu = \mathbb{E}[X_t] = \frac{c}{1-\phi}$$

Variance:

$$\gamma(0) = \mathrm{Var}(X_t) = \frac{\sigma^2}{1-\phi^2}$$

Autocovariance:

$$\gamma(h) = \phi^h \gamma(0) = \frac{\phi^h \sigma^2}{1-\phi^2}$$

Autocorrelation (ACF):

$$\rho(h) = \phi^h$$

Key insight: ACF decays exponentially at rate ϕ

TSA_ch2_ar1

Proof: AR(1) Mean

Claim: For AR(1): $X_t = c + \phi X_{t-1} + \varepsilon_t$, the mean is $\mu = \frac{c}{1-\phi}$

Proof: Take expectations of both sides:

$$\mathbb{E}[X_t] = \mathbb{E}[c + \phi X_{t-1} + \varepsilon_t] = c + \phi \mathbb{E}[X_{t-1}] + \mathbb{E}[\varepsilon_t]$$

By stationarity, $\mathbb{E}[X_t] = \mathbb{E}[X_{t-1}] = \mu$, and $\mathbb{E}[\varepsilon_t] = 0$:

$$\mu = c + \phi\mu$$

Solving for μ:

$$\mu - \phi\mu = c \implies \mu(1-\phi) = c \implies \boxed{\mu = \frac{c}{1-\phi}}$$

Requirement

This requires $\phi \neq 1$. If $\phi = 1$ (unit root), the mean is undefined.

Proof: AR(1) Variance

Claim: $\text{Var}(X_t) = \frac{\sigma^2}{1-\phi^2}$

Proof: Without loss of generality, assume $c = 0$ (centered process):

$$\text{Var}(X_t) = \text{Var}(\phi X_{t-1} + \varepsilon_t) = \phi^2 \text{Var}(X_{t-1}) + \text{Var}(\varepsilon_t) + 2\phi \text{Cov}(X_{t-1}, \varepsilon_t)$$

Since ε_t is independent of X_{t-1}, $\text{Cov}(X_{t-1}, \varepsilon_t) = 0$:

$$\gamma(0) = \phi^2 \gamma(0) + \sigma^2$$

By stationarity, $\text{Var}(X_t) = \text{Var}(X_{t-1}) = \gamma(0)$:

$$\gamma(0) - \phi^2 \gamma(0) = \sigma^2 \implies \gamma(0)(1 - \phi^2) = \sigma^2 \implies \boxed{\gamma(0) = \frac{\sigma^2}{1 - \phi^2}}$$

Note

- Requires $|\phi| < 1$ for positive variance
- As $|\phi| \to 1$, variance $\to \infty$

Proof: AR(1) Autocorrelation Function

Claim: $\rho(h) = \phi^h$ for $h \geq 0$

Proof: First, find autocovariance $\gamma(h) = \text{Cov}(X_t, X_{t-h})$.
Multiply $X_t = \phi X_{t-1} + \varepsilon_t$ by X_{t-h} and take expectations:

$$\mathbb{E}[X_t X_{t-h}] = \phi \mathbb{E}[X_{t-1} X_{t-h}] + \mathbb{E}[\varepsilon_t X_{t-h}]$$

For $h \geq 1$: $\mathbb{E}[\varepsilon_t X_{t-h}] = 0$ (future shock uncorrelated with past values)

$$\gamma(h) = \phi \gamma(h-1)$$

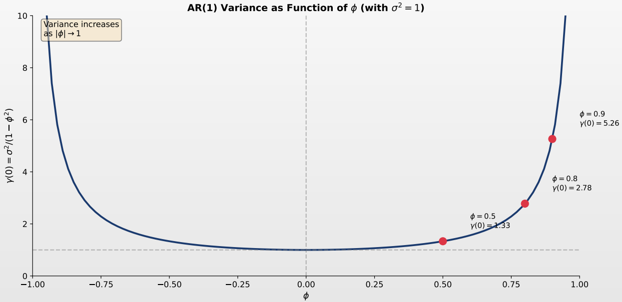
This is a recursive relation! Starting from $\gamma(0)$:

$$\gamma(1) = \phi\gamma(0), \quad \gamma(2) = \phi\gamma(1) = \phi^2\gamma(0), \quad \ldots \quad \boxed{\gamma(h) = \phi^h \gamma(0)}$$

The ACF is:

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{\phi^h \gamma(0)}{\gamma(0)} = \boxed{\phi^h}$$

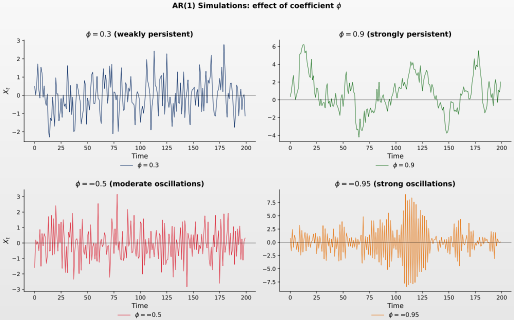
AR(1) Variance as Function of ϕ



Critical Insight

- As $|\phi| \to 1$, variance $\sigma^2/(1-\phi^2) \to \infty$
- Unit root processes ($\phi = 1$) are non-stationary: unbounded variance

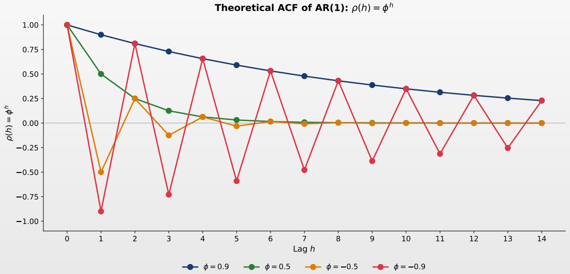
AR(1) Simulations: Effect of ϕ



- Different ϕ values produce distinct behavior: higher $|\phi|$ means more persistence
- Positive ϕ creates smooth, trending patterns; negative ϕ creates oscillations
- As $|\phi| \to 1$, the process becomes more persistent and approaches non-stationarity

TSA_ch2_ar1_simulation

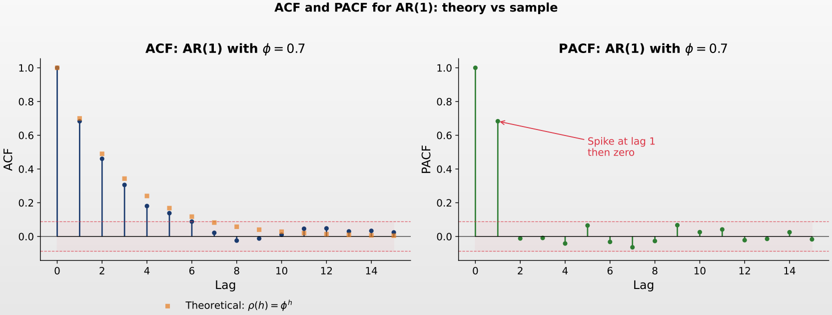
AR(1) Theoretical ACF



ACF Pattern

- For AR(1): $\rho(h) = \phi^h$
- Positive ϕ: smooth decay; negative ϕ: alternating decay

AR(1) ACF and PACF: Theory vs Sample



- **ACF**: Exponential decay at rate ϕ – theoretical formula: $\rho(h) = \phi^h$
- **PACF**: Single spike at lag 1, then cuts off – this identifies AR(1)
- Sample estimates (bars) fluctuate around theoretical values; use confidence bands

TSA_ch2_acf_pacf_patterns

AR(1) ACF and PACF Patterns

ACF of AR(1):

- Decays exponentially: $\rho(h) = \phi^h$
- If $\phi > 0$: all positive, gradual decay
- If $\phi < 0$: alternating signs, decay in magnitude

PACF of AR(1):

- **Cuts off after lag 1**
- $\pi_1 = \phi$, $\pi_k = 0$ for $k > 1$

	ACF	**PACF**
AR(1)	Exponential decay	Cuts off at lag 1

This is the key identification pattern for AR(1)!

AR(p) Model: General Form

Definition 4 (AR(p) Process)

An **autoregressive process of order p** is:

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t$$

Using lag operator:

$$\phi(L)X_t = c + \varepsilon_t$$

where $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$

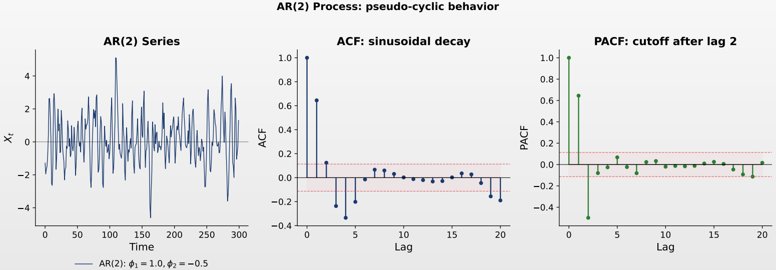
Stationarity condition:

- All roots of $\phi(z) = 0$ must lie **outside** the unit circle
- Equivalently: all roots have modulus > 1

PACF pattern:

- PACF cuts off after lag p
- ACF decays (exponentially or with damped oscillations)

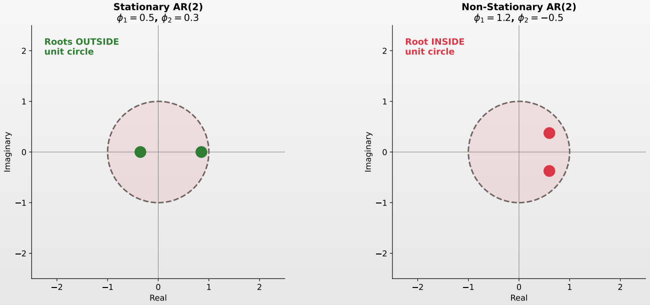
AR(p): Visual Illustration



- **AR(2) with real roots**: Mixture of two exponential decays in ACF
- **AR(2) with complex roots**: Pseudo-cyclic behavior
  - Damped sinusoidal ACF pattern
  - Period related to argument of complex roots
- **Key identification**: PACF cuts off after lag p (here, lag 2)
- Higher-order AR models can capture richer dynamics

TSA_ch2_ar2

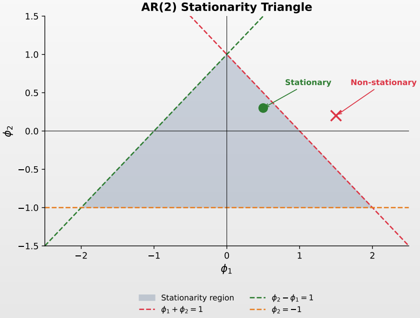
AR(2) Stationarity: Unit Circle Visualization



Stationarity Condition

- All roots of $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 = 0$ must lie **outside** the unit circle
- Equivalently, all roots must have modulus > 1

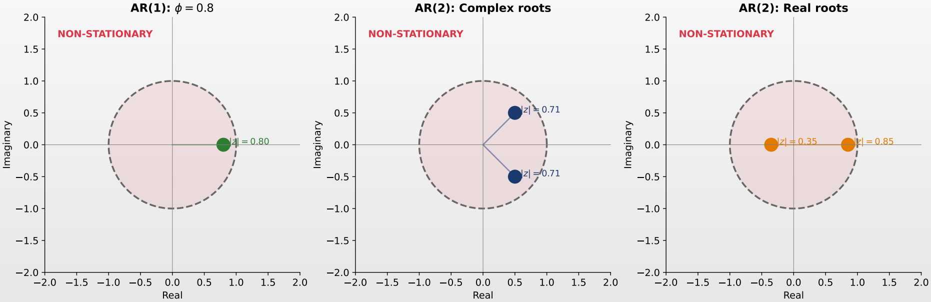
AR(2) Stationarity Triangle



- The triangular region defines all stationary AR(2) parameter combinations
- Boundaries: $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$, and $|\phi_2| < 1$
- Points outside this region lead to non-stationary or explosive processes

TSA_ch2_stationarity

Characteristic Polynomial Roots



Interpretation

- Complex conjugate roots produce oscillatory ACF behavior
- Closer to unit circle $\Rightarrow$ more persistent oscillations; real roots $\Rightarrow$ monotonic decay

AR(2) Model

Definition 5 (AR(2) Process)

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$$

Stationarity conditions for AR(2):

1. $\phi_1 + \phi_2 < 1$
2. $\phi_2 - \phi_1 < 1$
3. $|\phi_2| < 1$

ACF behavior depends on roots:

- **Real roots:** mixture of two exponential decays
- **Complex roots:** damped sinusoidal pattern (pseudo-cycles)

PACF: Cuts off after lag 2 ($\pi_k = 0$ for $k > 2$)

TSA_ch2_ar2

Quiz: AR Stationarity

Question

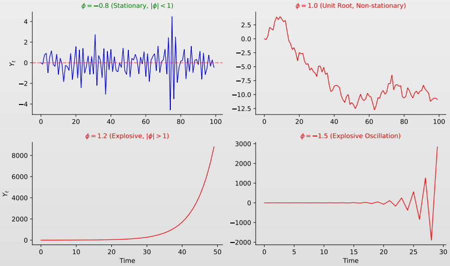
For which value of ϕ is the AR(1) process $X_t = c + \phi X_{t-1} + \varepsilon_t$ stationary?

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

Quiz: AR Stationarity – Answer

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

- AR(1) stationary iff $|\phi| < 1$; only $|-0.8| = 0.8 < 1$



TSA_ch2_ar1_simulation

MA(1) Model: Definition

Definition 6 (MA(1) Process)

A **moving average process of order 1** is:

$$X_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$$

where $\varepsilon_t \sim WN(0, \sigma^2)$.

Interpretation:

- μ: mean of the process
- θ: MA coefficient — measures impact of past shock
- Current value depends on current and one past shock

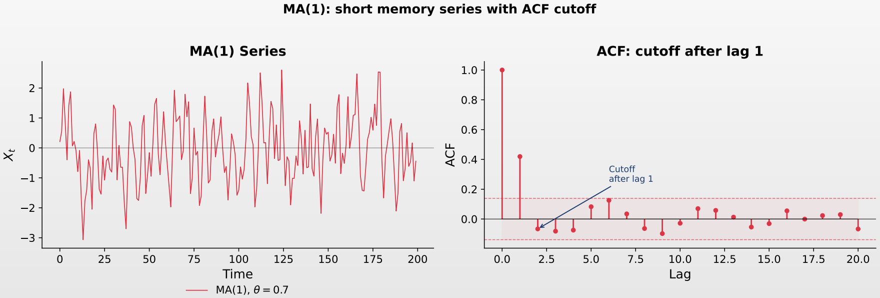
Using lag operator:

$$X_t = \mu + \theta(L)\varepsilon_t$$

where $\theta(L) = 1 + \theta L$

Key property: MA processes are **always stationary** for any finite θ

MA(1): Visual Illustration



- **Left panel**: MA(1) series — less persistent than AR(1), rapid mean reversion
- **Right panel**: ACF shows characteristic **cutoff after lag 1**
  - Only $\rho(1) \neq 0$; all higher lags are zero
  - This sharp cutoff is the key identifier for MA models
- PACF decays exponentially (opposite pattern to AR)

TSA_ch2_ma1

MA(1) Properties

For MA(1): $X_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$

Mean:

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

Variance:

$$\gamma(0) = \mathrm{Var}(X_t) = \sigma^2(1 + \theta^2)$$

Autocovariance:

$$\gamma(1) = \theta\sigma^2, \quad \gamma(h) = 0 \text{ for } h > 1$$

Autocorrelation (ACF):

$$\rho(1) = \frac{\theta}{1 + \theta^2}, \quad \rho(h) = 0 \text{ for } h > 1$$

Key insight: ACF **cuts off** after lag 1

TSA_ch2_ma1

Proof: MA(1) Variance and Autocovariance

Setup: $X_t = \varepsilon_t + \theta\varepsilon_{t-1}$ (assuming $\mu = 0$)

Variance:

$$\begin{aligned}\gamma(0) &= \text{Var}(X_t) = \text{Var}(\varepsilon_t + \theta\varepsilon_{t-1}) \\ &= \text{Var}(\varepsilon_t) + \theta^2\text{Var}(\varepsilon_{t-1}) + 2\theta\text{Cov}(\varepsilon_t, \varepsilon_{t-1}) \\ &= \sigma^2 + \theta^2\sigma^2 + 0 = \boxed{\sigma^2(1+\theta^2)}\end{aligned}$$

Autocovariance at lag 1:

$$\begin{aligned}\gamma(1) &= \text{Cov}(X_t, X_{t-1}) = \text{Cov}(\varepsilon_t + \theta\varepsilon_{t-1}, \varepsilon_{t-1} + \theta\varepsilon_{t-2}) \\ &= \text{Cov}(\varepsilon_t, \varepsilon_{t-1}) + \theta\text{Cov}(\varepsilon_t, \varepsilon_{t-2}) + \theta\text{Cov}(\varepsilon_{t-1}, \varepsilon_{t-1}) + \theta^2\text{Cov}(\varepsilon_{t-1}, \varepsilon_{t-2}) \\ &= 0 + 0 + \theta\sigma^2 + 0 = \boxed{\theta\sigma^2}\end{aligned}$$

Autocovariance at lag $h \geq 2$: No overlapping ε terms $\Rightarrow \gamma(h) = 0$

Proof: MA(1) ACF Maximum

Claim: $|\rho(1)| \leq 0.5$ for any value of θ

Proof: The ACF at lag 1 is:

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{\theta\sigma^2}{\sigma^2(1+\theta^2)} = \frac{\theta}{1+\theta^2}$$

To find the maximum, take derivative w.r.t. θ and set to zero:

$$\frac{d\rho(1)}{d\theta} = \frac{(1+\theta^2) - \theta(2\theta)}{(1+\theta^2)^2} = \frac{1-\theta^2}{(1+\theta^2)^2} = 0$$

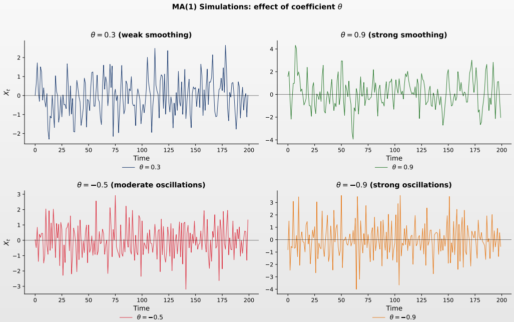
Solution: $\theta = \pm 1$. At these values:

$$\rho(1)\big|_{\theta=1} = \frac{1}{1+1} = \frac{1}{2}, \quad \rho(1)\big|_{\theta=-1} = \frac{-1}{1+1} = -\frac{1}{2}$$

Implication

- If you estimate $|\hat{\rho}(1)| > 0.5$ from data, the process is **not** MA(1)

MA(1) Simulations: Effect of θ



- MA(1) is always stationary regardless of θ – finite memory of only one lag
- Positive θ smooths the series; negative θ creates more rapid fluctuations
- Unlike AR(1), MA(1) shocks only affect the process for one period

TSA_ch2_ma1

MA(1) ACF and PACF Patterns

ACF of MA(1):

- **Cuts off after lag 1**
- $\rho(1) = \frac{\theta}{1+\theta^2}$, $\rho(h) = 0$ for $h > 1$
- Note: $|\rho(1)| \leq 0.5$ always (maximum at $\theta = \pm 1$)

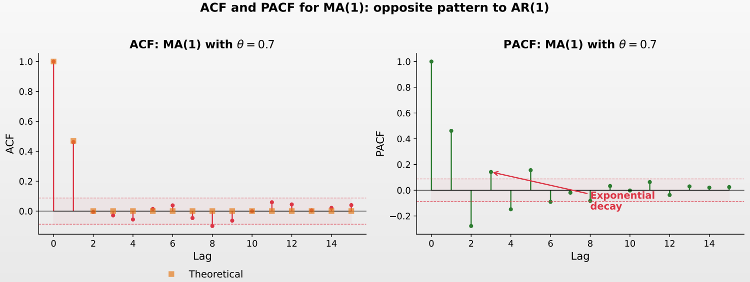
PACF of MA(1):

- Decays exponentially (or with alternating signs)
- Does *not* cut off

	ACF	**PACF**
MA(1)	Cuts off at lag 1	Exponential decay

This is the opposite pattern from AR(1)!

MA(1) ACF and PACF: Visual Comparison



Key Identification Pattern

- **ACF**: Single spike at lag 1, then cuts off — the MA(1) signature
- **PACF**: Exponential decay — opposite of AR(1)
- This ACF/PACF reversal distinguishes MA from AR

Invertibility of MA Models

Definition 7 (Invertibility)

An MA process is **invertible** if it can be written as an infinite AR process:

$$X_t = \mu + \sum_{j=1}^{\infty} \pi_j (X_{t-j} - \mu) + \varepsilon_t$$

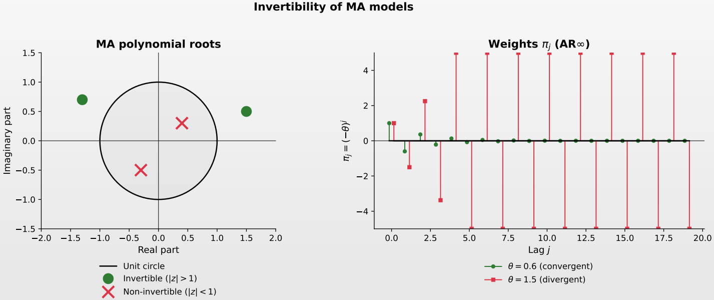
For MA(1): Invertible if $|\theta| < 1$
For MA(q): All roots of $\theta(z) = 0$ must lie outside the unit circle

Why invertibility matters:

- Ensures unique representation
- Required for forecasting and estimation
- Creates correspondence: AR(∞) $\leftrightarrow$ MA(q)

Note: Stationarity is for AR, Invertibility is for MA

Invertibility: Visual Illustration



- **Left**: Roots of $\theta(z) = 0$ must lie outside unit circle (for MA(1): $|\theta| < 1$)
- **Right**: Invertible weights $\pi_j = (-\theta)^j$ decay; non-invertible explode
- **Importance**: Ensures unique model representation and valid forecasts

MA(q) Model: General Form

Definition 8 (MA(q) Process)

A **moving average process of order q** is:

$$X_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}$$

Using lag operator:

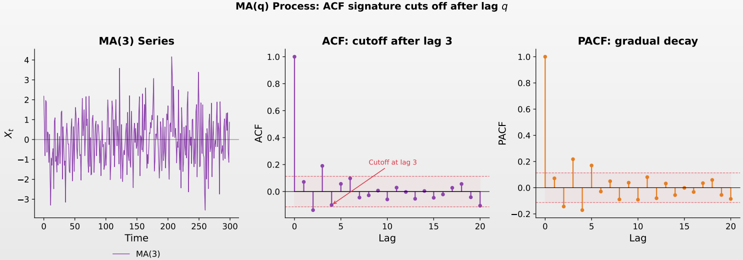
$$X_t = \mu + \theta(L)\varepsilon_t$$

where $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$

Properties:

- Always stationary (finite variance)
- ACF cuts off after lag q: $\rho(h) = 0$ for $h > q$
- PACF decays gradually
- Invertible if all roots of $\theta(z) = 0$ lie outside unit circle

MA(q): Visual Illustration



- **MA(3) example**: Process depends on current and 3 past shocks
- **ACF signature**: Cuts off sharply after lag q (here, lag 3)
  - Non-zero correlations only at lags 1, 2, 3
  - All higher lags are exactly zero (within sampling error)
- **PACF**: Gradual exponential/oscillating decay
- Sharp ACF cutoff is the key identifier for pure MA processes

Quiz: ACF/PACF Pattern Recognition

Question

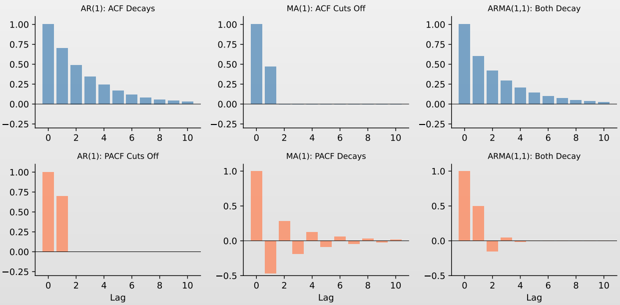
You observe: ACF has spike at lag 1, then cuts off. PACF decays gradually. What model?

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

Quiz: ACF/PACF Pattern Recognition – Answer

Correct Answer: (B) MA(1)

ACF cuts off → MA process; PACF decays → confirms MA(1)



Quiz: MA Invertibility

Question

Is MA(1) $X_t = \varepsilon_t + 1.5\varepsilon_{t-1}$ invertible?

(A) Yes, MA processes are always invertible

(B) Yes, because $1.5 > 0$

(C) No, because $|\theta| = 1.5 > 1$

(D) No, MA processes are never invertible

Quiz: MA Invertibility

Question

Is MA(1) $X_t = \varepsilon_t + 1.5\varepsilon_{t-1}$ invertible?

(A) Yes, MA processes are always invertible

(B) Yes, because $1.5 > 0$

(C) No, because $|\theta| = 1.5 > 1$

(D) No, MA processes are never invertible

Answer: (C)

Invertibility requires $|\theta| < 1$. Here $|\theta| = 1.5 > 1$, so not invertible.

ARMA(p,q) Model: Definition

Definition 9 (ARMA(p,q) Process)

An **autoregressive moving average process** of order (p,q) is:

$$X_t = c + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

Compact form using lag operators:

$$\phi(L)(X_t - \mu) = \theta(L)\varepsilon_t$$

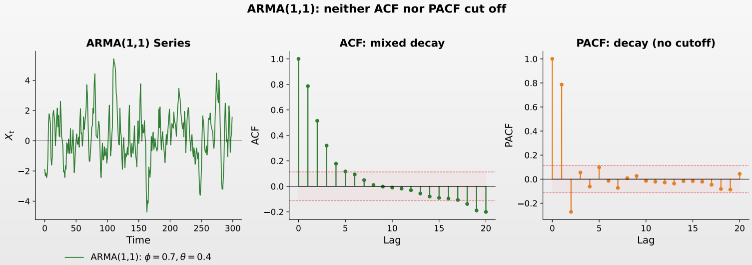
or equivalently:

$$\phi(L)X_t = c + \theta(L)\varepsilon_t$$

where $\mu = \frac{c}{1-\phi_1-\cdots-\phi_p}$

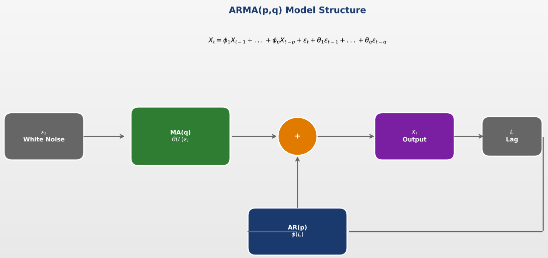
Key idea: Combines AR and MA components for more flexible modeling

ARMA: Visual Illustration



- **ARMA(1,1) combines** AR persistence with MA shock response
- **ACF pattern**: Decays after initial lag (not sharp cutoff like pure MA)
  - First lag influenced by both ϕ and θ
  - Subsequent lags decay geometrically like AR
- **PACF pattern**: Also decays (no sharp cutoff like pure AR)
- Neither ACF nor PACF cuts off — key identifier for mixed models

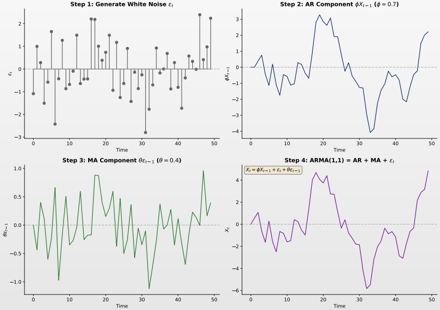
ARMA Model Structure



Model Components

- ARMA combines AR (past values) and MA (past shocks) components
- AR captures persistence; MA captures short-term shock effects

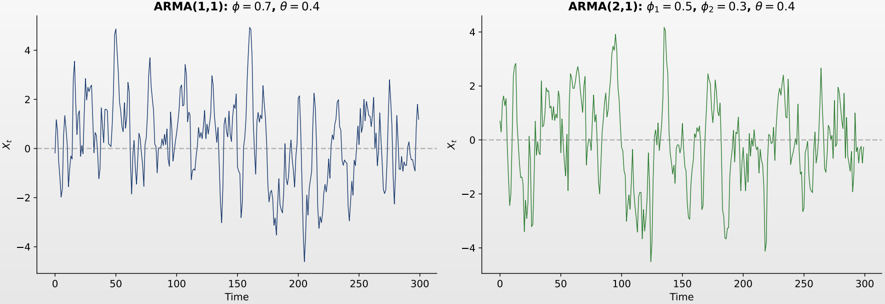
How ARMA Simulation Works



Simulation Algorithm

- Generate white noise ε_t, apply MA filter, then AR recursion

ARMA Examples



Key Observation

- Different ARMA specifications produce visually similar series but distinct ACF/PACF
- Model identification requires examining autocorrelation structure

ARMA(1,1) Model

Definition 10 (ARMA(1,1) Process)

$$X_t = c + \phi X_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$$

Properties (assuming stationarity and invertibility):

- Mean: $\mu = \frac{c}{1-\phi}$
- Variance: $\gamma(0) = \frac{(1+2\phi\theta+\theta^2)\sigma^2}{1-\phi^2}$

ACF:

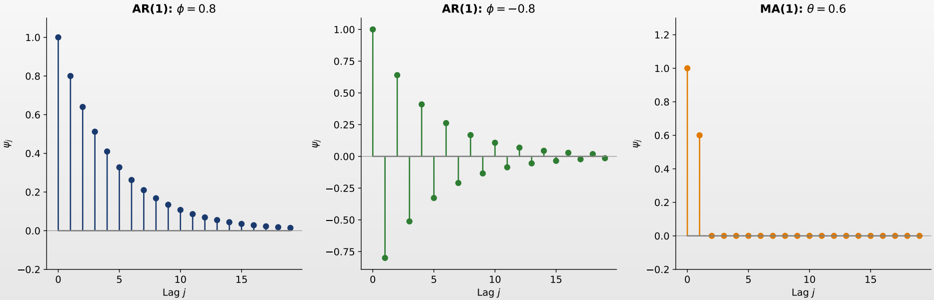
$$\rho(1) = \frac{(1+\phi\theta)(\phi+\theta)}{1+2\phi\theta+\theta^2}$$

$$\rho(h) = \phi \cdot \rho(h-1) \quad \text{for } h \geq 2$$

Pattern: ACF decays exponentially after lag 1 (like AR), but starting point depends on both ϕ and θ

TSA_ch2_arma

Impulse Response Functions



Interpretation

- IRF shows how a unit shock $\varepsilon_t = 1$ propagates over time
- For stationary processes, IRF decays to zero as $h \to \infty$

Stationarity and Invertibility Summary

For ARMA(p,q) to be well-behaved:

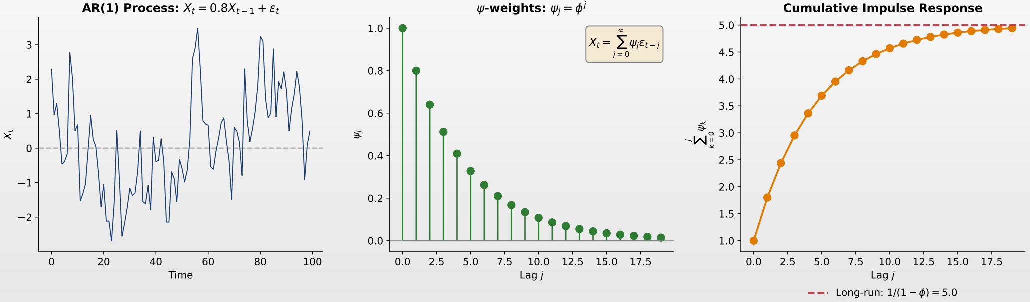
Condition	Requirement
Stationarity	Roots of $\phi(z) = 0$ outside unit circle
Invertibility	Roots of $\theta(z) = 0$ outside unit circle

Implications:

- **Stationarity:** Can write as MA(∞): $X_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$
- **Invertibility:** Can write as AR(∞): $X_t = \mu + \sum_{j=1}^{\infty} \pi_j (X_{t-j} - \mu) + \varepsilon_t$

Causal representation: X_t depends only on *past* shocks (not future)

Wold's Decomposition Theorem



Fundamental Result

- Any stationary process: $X_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ with $\psi_0 = 1$, $\sum \psi_j^2 < \infty$
- ARMA models are parsimonious approximations to this infinite representation

Quiz: 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

Quiz: 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: (C) ARMA model

- $\phi(L)$ is the AR polynomial, $\theta(L)$ is the MA polynomial $\rightarrow$ ARMA(p,q)

Quiz: Lag Operator

Question

What is $(1 - L)^2 X_t$?

(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) $X_t - X_{t-2}$

Quiz: Lag Operator

Question

What is $(1 - L)^2 X_t$?

(A) $X_t - X_{t-1}$

(B) $X_t - 2X_{t-1} + X_{t-2}$

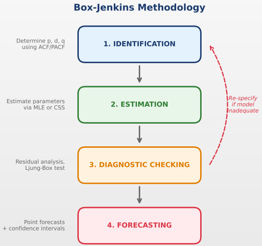
(C) $X_t + X_{t-1} + X_{t-2}$

(D) $X_t - X_{t-2}$

Answer: (B)

- $(1 - L)^2 = 1 - 2L + L^2$
- $(1 - L)^2 X_t = X_t - 2X_{t-1} + X_{t-2}$

The Box-Jenkins Methodology



Three-Stage Process

- **Identification**: ACF/PACF $\rightarrow$ orders (p, q); **Estimation**: MLE; **Diagnostics**: white noise residuals

Model Identification Summary Table

Model Identification: ACF/PACF Patterns

Model	ACF Pattern	PACF Pattern
AR(p)	Exponential decay or damped oscillation	Cuts off after lag p
MA(q)	Cuts off after lag q	Exponential decay or damped oscillation
ARMA(p,q)	Exponential decay after lag q-p	Exponential decay after lag p-q

Practical tip: Start simple (low p, q), increase if diagnostics fail

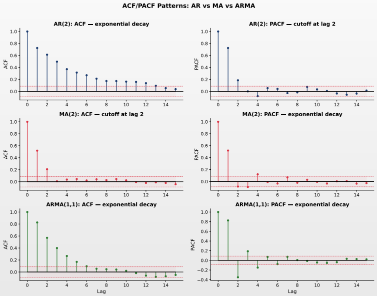
TSA_ch2_acf_pacf_patterns

ACF/PACF Identification Rules

Theoretical patterns for stationary processes:

Model	ACF Pattern	PACF Pattern
AR(1)	Exponential decay	Spike at lag 1, then 0
AR(2)	Damped exponential/sine	Spikes at lags 1-2, then 0
AR(p)	Decays gradually	Cuts off after lag p
MA(1)	Spike at lag 1, then 0	Exponential decay
MA(2)	Spikes at lags 1-2, then 0	Damped exponential/sine
MA(q)	Cuts off after lag q	Decays gradually
ARMA(p,q)	Decays	Decays

ACF/PACF Patterns: Visual Guide



Practical Identification

- **AR**: ACF decays, PACF cuts off — use PACF for order p
- **MA**: ACF cuts off, PACF decays — use ACF for order q
- **ARMA**: Both decay — use AIC/BIC

Information Criteria

Purpose: Balance goodness-of-fit against model complexity

Akaike Information Criterion (AIC):

$$\text{AIC} = -2\ln(\hat{L}) + 2k$$

Bayesian Information Criterion (BIC/SBC):

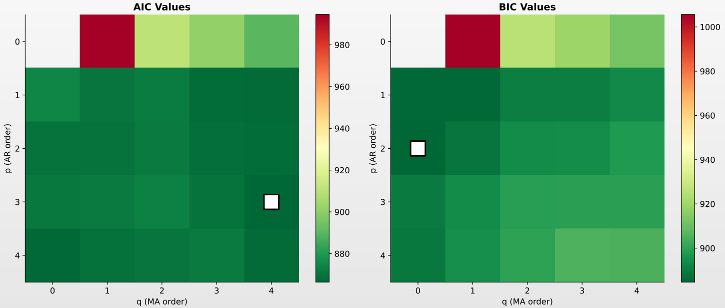
$$\text{BIC} = -2\ln(\hat{L}) + k\ln(n)$$

where $\hat{L}$ = maximized likelihood, k = number of parameters, n = sample size

Usage:

- Lower values are better
- BIC penalizes complexity more strongly than AIC
- AIC tends to choose larger models; BIC more parsimonious
- Compare models fit to the *same data*

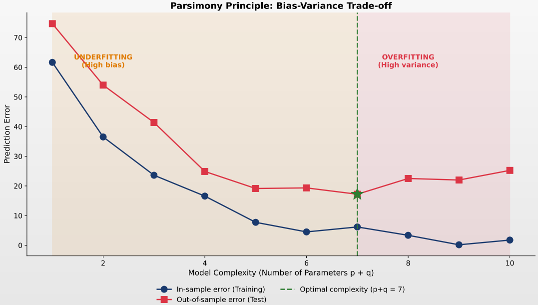
AIC vs BIC: Model Selection



Interpretation

- White square = best model; lower values (green) better; BIC prefers simpler models

Parsimony Principle: Bias-Variance Trade-off



Key Principle

- Simple models underfit (high bias); complex models overfit (high variance)
- Optimal model balances fit and complexity — "as simple as possible, but no simpler"

Automatic Model Selection

Grid search approach:

1. Fit ARMA(p,q) for $p = 0, 1, \ldots, p_{max}$ and $q = 0, 1, \ldots, q_{max}$
2. Select model with lowest AIC or BIC
3. Verify with diagnostic checks

In Python (statsmodels):

- `pm.auto_arima()` from `pmdarima` package
- Automatically tests stationarity, searches over orders
- Returns best model by AIC/BIC

Caution:

- Automatic selection is a starting point, not final answer
- Always check diagnostics
- Consider domain knowledge

Quiz: Information Criteria

Question

Comparing ARMA(1,1) vs ARMA(2,1) using BIC, which is correct?

(A) Lower BIC always means better forecasts

(B) BIC penalizes complexity less than AIC

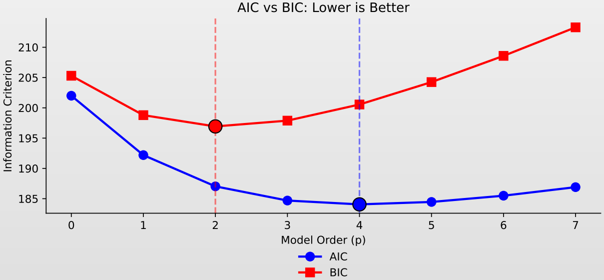
(C) The model with lower BIC is preferred

(D) BIC can only compare models with same # of parameters

Quiz: Information Criteria – Answer

Correct Answer: (C) Lower BIC is preferred

Lower BIC indicates better fit-complexity trade-off. BIC penalizes complexity more than AIC.



Estimation Methods Overview

Three main approaches:

1. Method of Moments / Yule-Walker (AR only)

- Match sample autocorrelations to theoretical values
- Simple, closed-form for AR models
- Not efficient for MA components

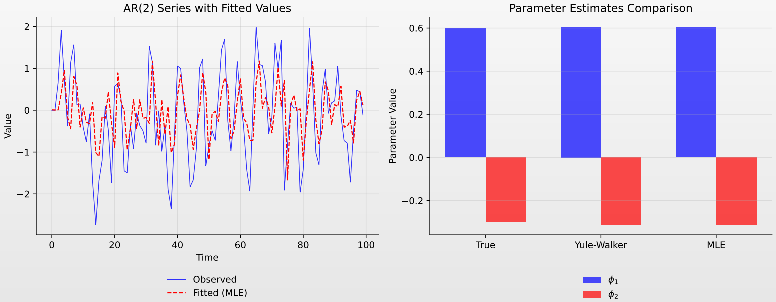
2. Maximum Likelihood Estimation (MLE)

- Most common approach
- Requires distributional assumption (usually Gaussian)
- Efficient and consistent

3. Conditional Least Squares

- Minimize sum of squared residuals
- Conditioning on initial observations
- Computationally simpler than exact MLE

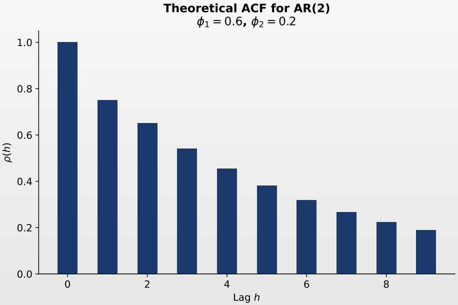
Estimation Methods Comparison



Method Selection

- **Yule-Walker**: Fast, closed-form for AR; **MLE**: Most efficient; **CSS**: Good balance

Yule-Walker Equations for AR(p)



Yule-Walker Equations

$$\rho(1) = \phi_1 + \phi_2 \rho(1)$$

$$\rho(2) = \phi_1 \rho(1) + \phi_2$$

Matrix form: $R \cdot \phi = \rho$

R = autocorrelation matrix

Solution: $\hat{\phi} = R^{-1}\rho$

Key Property

- Exploits relationship between AR parameters and autocorrelations
- Replace $\rho(k)$ with sample $\hat{\rho}(k)$ for estimation

Yule-Walker Equations: Matrix Form

For AR(p): $X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t$

Yule-Walker equations:

$$\rho(k) = \phi_1 \rho(k-1) + \phi_2 \rho(k-2) + \cdots + \phi_p \rho(k-p)$$

for $k = 1, 2, \ldots, p$

Matrix form:

$$\begin{pmatrix} \rho(0) & \rho(1) & \cdots & \rho(p-1) \\ \rho(1) & \rho(0) & \cdots & \rho(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(p-1) & \rho(p-2) & \cdots & \rho(0) \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(p) \end{pmatrix}$$

Estimation: Replace $\rho(k)$ with sample autocorrelations $\hat{\rho}(k)$

Proof: Yule-Walker Equations

Goal: Derive the relationship $\rho(k) = \phi_1\rho(k-1) + \cdots + \phi_p\rho(k-p)$

Proof: Start with AR(p): $X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t$
Multiply both sides by X_{t-k} and take expectations:

$$\mathbb{E}[X_t X_{t-k}] = \phi_1\mathbb{E}[X_{t-1}X_{t-k}] + \cdots + \phi_p\mathbb{E}[X_{t-p}X_{t-k}] + \mathbb{E}[\varepsilon_t X_{t-k}]$$

For $k \geq 1$: $\mathbb{E}[\varepsilon_t X_{t-k}] = 0$ (future shock uncorrelated with past)
Using $\gamma(k) = \mathbb{E}[X_t X_{t-k}]$ (assuming zero mean):

$$\gamma(k) = \phi_1\gamma(k-1) + \phi_2\gamma(k-2) + \cdots + \phi_p\gamma(k-p)$$

Dividing by $\gamma(0)$:

$$\boxed{\rho(k) = \phi_1\rho(k-1) + \phi_2\rho(k-2) + \cdots + \phi_p\rho(k-p)}$$

AR(1) Special Case

- $\rho(k) = \phi_1\rho(k-1) = \phi_1^k$ (using $\rho(0) = 1$)

Maximum Likelihood Estimation

Assuming Gaussian errors: $\varepsilon_t \sim N(0, \sigma^2)$

Log-likelihood for ARMA(p,q):

$$\ell(\phi, \boldsymbol{\theta}, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{n}\varepsilon_t^2$$

where ε_t are the innovations computed recursively.

Estimation procedure:

1. Initialize: use method of moments or OLS for starting values
2. Optimize: numerical methods (e.g., BFGS, Newton-Raphson)
3. Iterate until convergence

In practice: Use `statsmodels.tsa.arima.model.ARIMA`

Standard Errors and Inference

Asymptotic distribution of MLE:

$$\hat{\boldsymbol{\theta}} \xrightarrow{d} N\left(\boldsymbol{\theta}_0, \frac{1}{n}\mathrm{I}(\boldsymbol{\theta}_0)^{-1}\right)$$

where $\mathrm{I}(\boldsymbol{\theta})$ is the Fisher information matrix.

Standard errors: Square root of diagonal of $\frac{1}{n}\hat{\mathrm{I}}^{-1}$

Hypothesis testing:

- $H_0 : \phi_j = 0$ (or $\theta_j = 0$)
- Test statistic: $z = \frac{\hat{\phi}_j}{SE(\hat{\phi}_j)} \sim N(0, 1)$ asymptotically
- Reject if $|z| > 1.96$ at 5% level

Confidence interval: $\hat{\phi}_j \pm 1.96 \cdot SE(\hat{\phi}_j)$

Residual Analysis

If model is correctly specified, residuals should be white noise:

1. Plot residuals over time

- Should fluctuate around zero
- No obvious patterns or trends
- Constant variance (no heteroskedasticity)

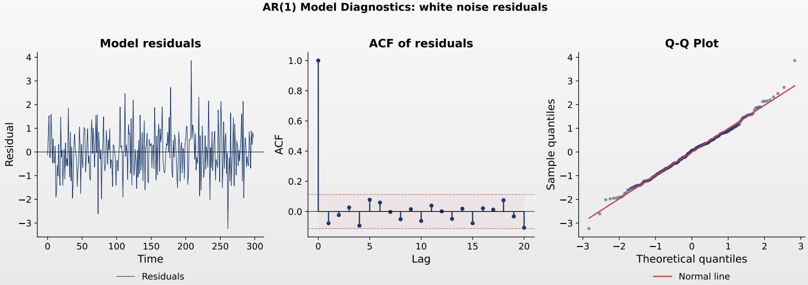
2. Check ACF of residuals

- All correlations should be within confidence bands
- No significant spikes $\rightarrow$ white noise

3. Check histogram / Q-Q plot

- Should be approximately normal (if assuming Gaussian)
- Heavy tails suggest non-normal errors

Residual Diagnostics: Example



What to Look For

- **Residual plot**: Random scatter around zero, constant variance
- **ACF**: No significant spikes (white noise)
- **Q-Q plot**: Points on diagonal (normality)

Ljung-Box Test

Definition 11 (Ljung-Box Test)

Tests whether residuals are independently distributed (no autocorrelation).

Test statistic:

$$Q(m) = n(n+2) \sum_{k=1}^{m} \frac{\hat{\rho}_k^2}{n-k}$$

Hypotheses:

- H_0: Residuals are white noise (no autocorrelation up to lag m)
- H_1: Residuals are autocorrelated

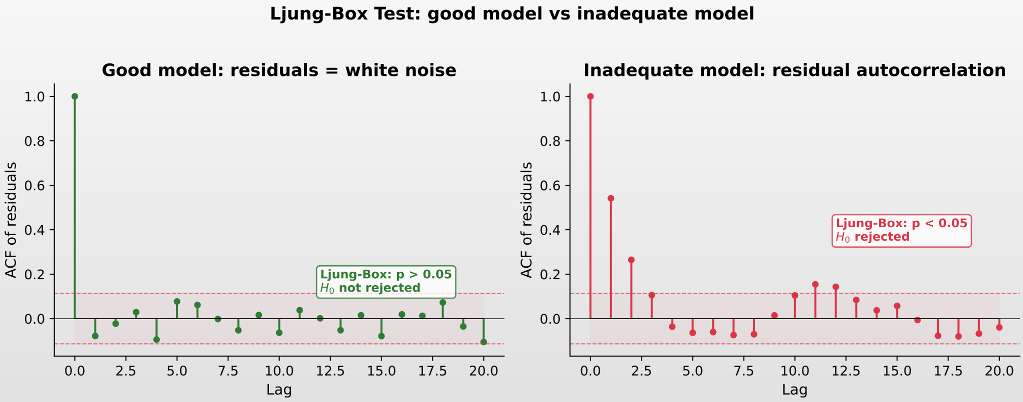
Distribution: Under H_0, $Q(m) \sim \chi^2(m-p-q)$ approximately

Decision:

- p-value $> 0.05 \rightarrow$ fail to reject $H_0 \rightarrow$ residuals look like white noise (good!)
- p-value $< 0.05 \rightarrow$ significant autocorrelation remains $\rightarrow$ model inadequate

TSA_ch2_diagnostics

Ljung-Box Test: Visual Illustration



Left: Good model – residuals are white noise (no significant ACF). Right: Poor model – residuals show autocorrelation.

Diagnostic Checklist

A good ARMA model should satisfy:

1. **Stationarity:** AR roots outside unit circle
   ✓ Check with `arroots`
2. **Invertibility:** MA roots outside unit circle
   ✓ Check with `maroots`
3. **White noise residuals:** No significant ACF
   ✓ ACF plot, Ljung-Box test
4. **Normal residuals:** (if assumed)
   ✓ Q-Q plot, Jarque-Bera test
5. **No heteroskedasticity:** Constant variance
   ✓ Plot residuals, ARCH test
6. **Parsimonious:** Lowest AIC/BIC among adequate models

If diagnostics fail: Return to identification, try different orders

Quiz: Ljung-Box Test

Question

After fitting an ARMA model, you run the Ljung-Box test on residuals and get p-value $= 0.03$. What does this mean?

(A) Model is adequate, residuals are white noise

(B) Model is inadequate, residuals have autocorrelation

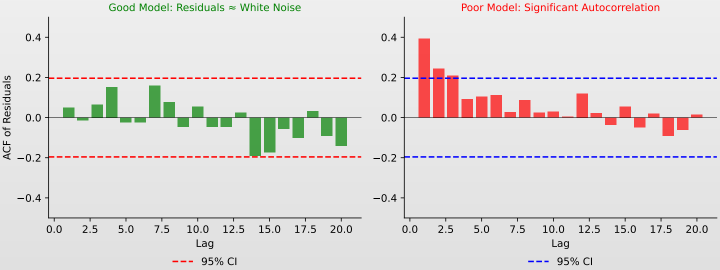
(C) Need to increase sample size

(D) Test is inconclusive

Quiz: Ljung-Box Test – Answer

Correct Answer: (B) Model is inadequate

- p-value < 0.05 rejects H_0 (white noise)
- Indicates remaining autocorrelation in residuals



Point Forecasts

Optimal forecast: Conditional expectation minimizes MSE

$$\hat{X}_{n+h|n} = \mathbb{E}[X_{n+h}|X_n, X_{n-1}, \ldots]$$

For AR(1): $X_t = c + \phi X_{t-1} + \varepsilon_t$

$$\hat{X}_{n+1|n} = c + \phi X_n$$
$$\hat{X}_{n+2|n} = c + \phi \hat{X}_{n+1|n} = c(1+\phi) + \phi^2 X_n$$
$$\hat{X}_{n+h|n} = \mu + \phi^h (X_n - \mu)$$

Key property: Forecasts converge to mean μ as $h \to \infty$

For MA(1): $X_t = \mu + \varepsilon_t + \theta \varepsilon_{t-1}$

$$\hat{X}_{n+1|n} = \mu + \theta \varepsilon_n$$
$$\hat{X}_{n+h|n} = \mu \quad \text{for } h > 1$$

Forecast Uncertainty

Forecast error:

$$e_{n+h|n} = X_{n+h} - \hat{X}_{n+h|n}$$

Mean squared forecast error (MSFE):

$$\mathrm{MSFE}(h) = \mathbb{E}[e^2_{n+h|n}] = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2$$

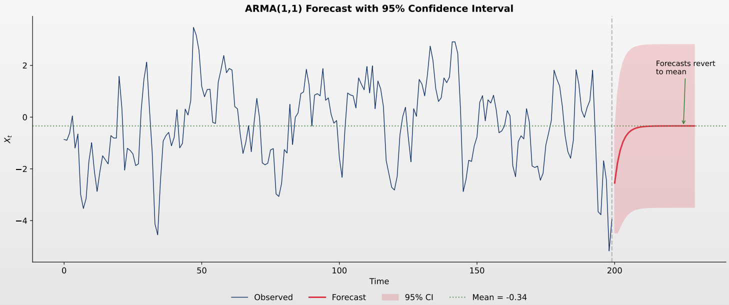
where ψ_j are the MA(∞) coefficients.

For AR(1): $\psi_j = \phi^j$

$$\mathrm{MSFE}(h) = \sigma^2 \frac{1-\phi^{2h}}{1-\phi^2} \to \frac{\sigma^2}{1-\phi^2} = \mathrm{Var}(X_t)$$

Key insight: Forecast uncertainty increases with horizon, eventually reaching unconditional variance

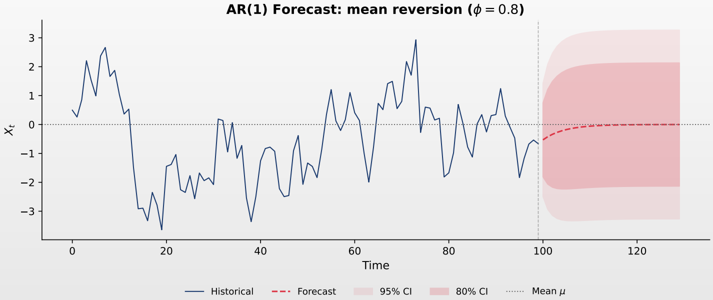
ARMA Forecasting with Confidence Intervals



Forecast Properties

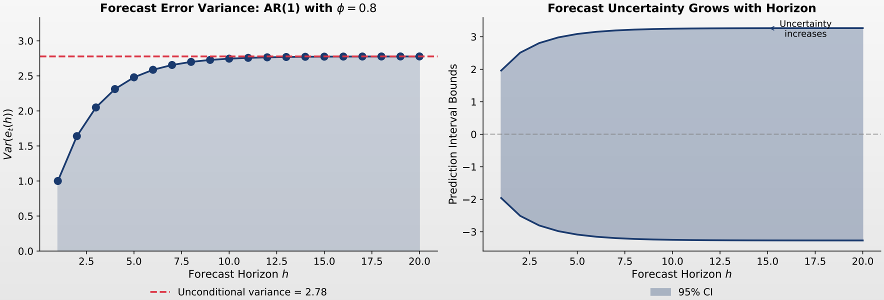
- Forecasts converge to mean; confidence intervals widen with horizon

AR(1) Forecasting: Mean Reversion



- Forecasts converge to μ as horizon increases; rate depends on $|\phi|$
- Confidence intervals widen, eventually reaching unconditional variance

Forecast Error Variance Over Horizon



Variance Decomposition

- Forecast error variance grows with h; converges to $\gamma(0)$ as $h \to \infty$

Confidence Intervals for Forecasts

Assuming Gaussian errors:

$$X_{n+h}|X_n, \ldots \sim N\left(\hat{X}_{n+h|n}, \text{MSFE}(h)\right)$$

$(1-\alpha)$ **confidence interval:**

$$\hat{X}_{n+h|n} \pm z_{\alpha/2} \cdot \sqrt{\text{MSFE}(h)}$$

where $z_{\alpha/2} = 1.96$ for 95% CI.

Properties:

- Intervals widen as horizon increases
- Eventually converge to unconditional interval: $\mu \pm z_{\alpha/2}\sigma_X$
- Width depends on model parameters (AR coefficients, etc.)

In Python: `model.get_forecast(h).conf_int()`

Forecast Evaluation

Out-of-sample testing:

1. Split data: training set (fit model) and test set (evaluate)
2. Generate forecasts for test period
3. Compare forecasts to actual values

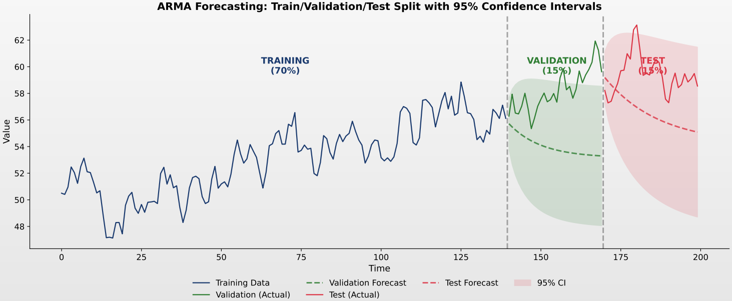
Metrics (from Chapter 1):

- $\text{MAE} = \frac{1}{n}\sum |e_t|$
- $\text{RMSE} = \sqrt{\frac{1}{n}\sum e_t^2}$
- $\text{MAPE} = \frac{100}{n}\sum \left|\frac{e_t}{X_t}\right|$

Rolling/expanding window:

- Re-estimate model as new data arrives
- More realistic assessment of forecast performance

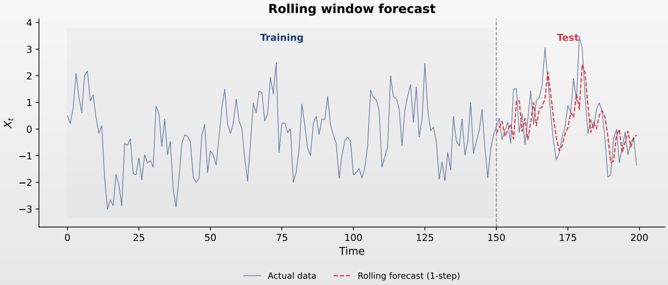
Train/Validation/Test Forecasting Example



Best Practice

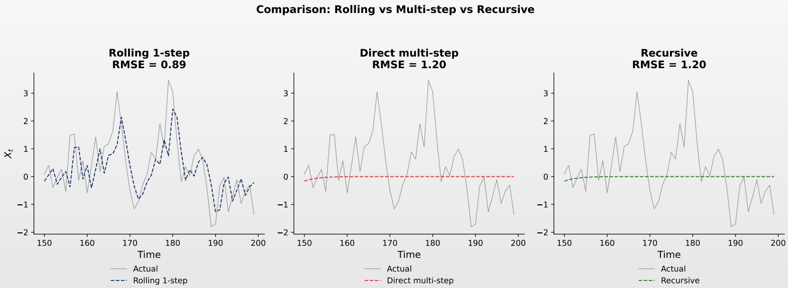
- Evaluate forecasts on held-out data: train for fitting, validation for selection, test for assessment

Rolling Window Forecasting



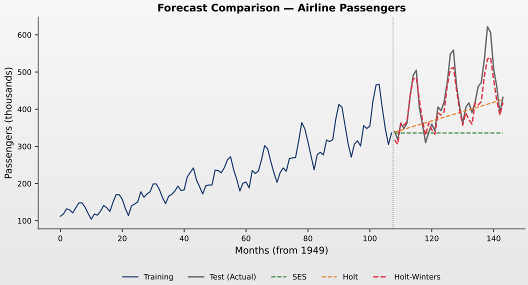
- **Fixed window:** Re-estimate using most recent w observations
- **Expanding window:** Use all data up to forecast origin
- Generate 1-step forecast, move window forward, repeat

Rolling vs Multi-Step Forecasting



- **Rolling 1-step:** More accurate, requires frequent re-estimation
- **Direct multi-step:** Estimate separate model for each horizon h
- **Recursive multi-step:** Iterate 1-step forecasts (error accumulation)

Real Data Application: Forecasting Comparison



- Real data often exhibits non-stationarity, structural breaks
- Compare multiple models: ARMA, exponential smoothing, naive
- Use cross-validation or rolling evaluation for robust assessment

Quiz: Forecast Properties

Question

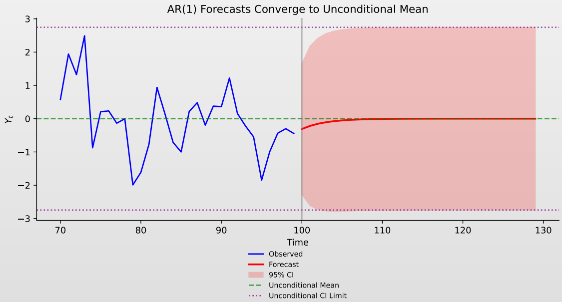
For a stationary AR(1) model, what happens to forecasts as horizon $h \to \infty$?

(A) Forecasts grow without bound
(B) Forecasts oscillate forever
(C) Forecasts converge to the unconditional mean μ
(D) Forecasts become more accurate

Quiz: Forecast Properties – Answer

Correct Answer: (C) Forecasts converge to μ

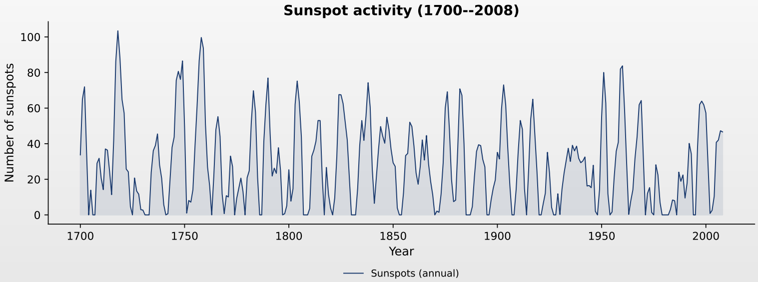
- $\hat{X}_{n+h|n} = \mu + \phi^h(X_n - \mu) \to \mu$ as $h \to \infty$ (since $|\phi| < 1$)



Workflow Summary

1. **Data preparation**
   - Check for missing values, outliers
   - Transform if necessary (log, differencing)
2. **Stationarity check**
   - Visual inspection: time plot, ACF
   - Formal tests: ADF, KPSS
   - Difference if non-stationary
3. **Model identification**
   - ACF/PACF patterns
   - Information criteria grid search
4. **Estimation and diagnostics**
   - Fit model, check significance
   - Residual analysis, Ljung-Box test
5. **Forecasting**
   - Point forecasts with confidence intervals
   - Out-of-sample validation

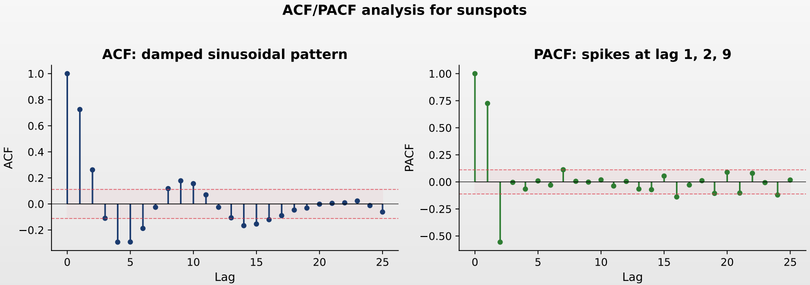
Case Study: Sunspot Numbers



Data Description

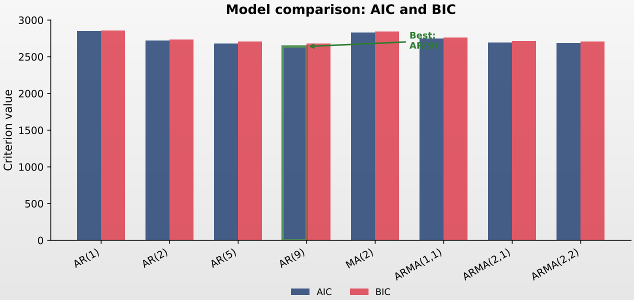
- **Yearly sunspot numbers (1700–2008)**: Classic dataset, stationary with 11-year cycles
- We will apply the complete Box-Jenkins methodology

Step 1: ACF/PACF Analysis



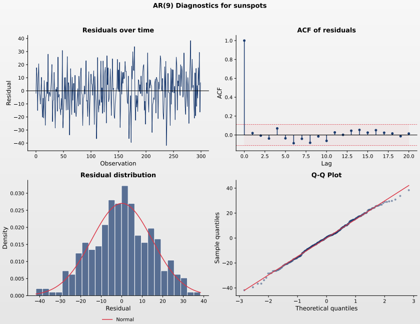
- **ACF**: Slow, sinusoidal decay — suggests AR process
- **PACF**: Significant spikes at lags 1, 2, 9 — suggests AR(9) or AR(2)
- Series appears stationary (no differencing needed, $d = 0$)

Step 2: Model Comparison



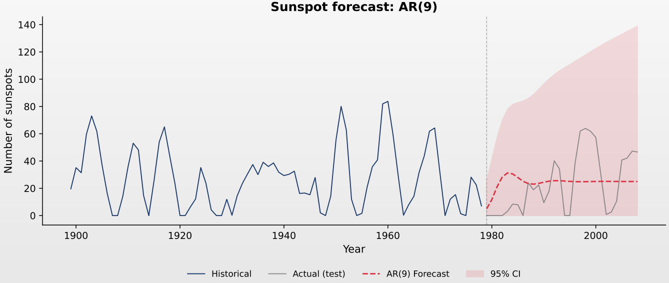
We compare candidate models using AIC. The **AR(9)** model has the lowest AIC, capturing the 11-year solar cycle.

Step 3: Diagnostic Checking



Residuals resemble white noise: zero mean, constant variance, no significant ACF, approximately normal.

Step 4: Forecasting



- AR(9) captures the cyclical nature of sunspots
- 95% CI covers most actual values; RMSE $\approx$ 30 (reasonable)

Key Takeaways

1. **AR(p):** Depends on p past values; stationarity requires roots outside unit circle; PACF cuts off at lag p
2. **MA(q):** Depends on q past shocks; always stationary; ACF cuts off at lag q
3. **ARMA(p,q):** Combines AR and MA; both ACF and PACF decay
4. **Box-Jenkins:** Identify $\rightarrow$ Estimate $\rightarrow$ Diagnose $\rightarrow$ Forecast
5. **Diagnostics:** Residuals must be white noise
6. **Forecasts:** Converge to mean; uncertainty increases with horizon

Next Chapter Preview

Chapter 3: ARIMA and Seasonal Models

- ARIMA(p,d,q): Integrated models for non-stationary data
- Seasonal ARIMA: SARIMA(p,d,q)(P,D,Q)$_s$
- Seasonal differencing
- Real-world applications with seasonal patterns

Reading:

- Hyndman & Athanasopoulos, *Forecasting: Principles and Practice*, Ch. 9
- Box, Jenkins, Reinsel & Ljung, *Time Series Analysis*, Ch. 3-4

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.
- Shumway, R.H., & Stoffer, D.S. (2017). *Time Series Analysis and Its Applications*. 4th ed., Springer.

Data Sources

Real Data – Case Studies

- **Sunspot Numbers**: Monthly (1749–2024)
  - Source: SIDC – Royal Observatory of Belgium
- **All illustrations**: Reproducible via QuantLet links

Software & Tools

- **Python**: statsmodels (ARIMA), numpy, matplotlib
  - Key functions: ARIMA(), acf(), pacf()
- **R**: forecast, tseries packages
  - Key functions: auto.arima(), Arima(), adf.test()

Thank You!

Questions?