



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

Learning Objectives

By the end of this chapter, you will be able to:

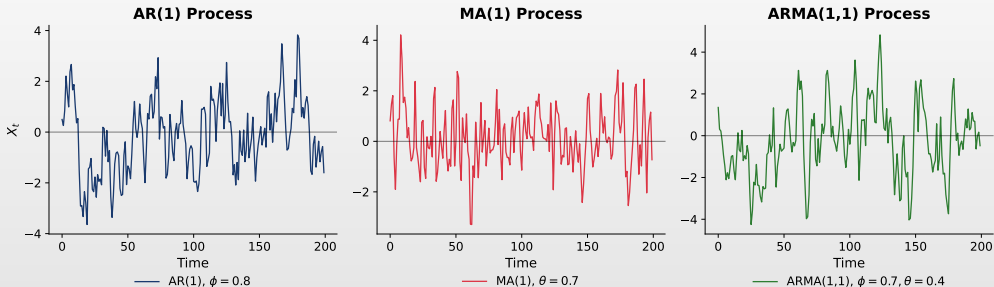
1. **Define** and simulate $AR(p)$, $MA(q)$, and $ARMA(p, q)$ processes
2. **Verify** stationarity and invertibility conditions
3. **Identify** orders p and q through ACF/PACF analysis
4. **Estimate** parameters via Yule-Walker, MLE, and information criteria (AIC, BIC)
5. **Diagnose** the model through residual analysis and the Ljung-Box test
6. **Forecast** using ARMA models with confidence intervals
7. **Apply** the Box-Jenkins methodology to real data (sunspots)

Chapter Structure

- Motivation
- Introduction and the Lag Operator
- Autoregressive (AR) Models
- Moving Average (MA) Models
- ARMA Models
- Model Identification
- Parameter Estimation
- Model Diagnostics
- Forecasting with ARMA
- Practical Implementation
- Case Study: Real Data
- Summary
- Quiz

Why ARMA Models?

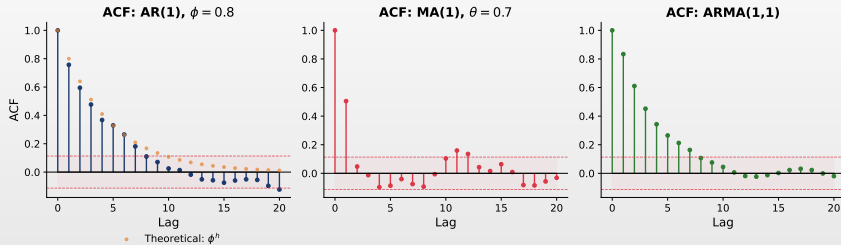
Stationary processes: AR, MA and ARMA



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

Model Identification Through ACF Patterns

Distinct ACF patterns for different models



ACF Reflects Model Structure

- ▣ **Distinct patterns:** AR: exponential decay; MA: sharp cutoff; ARMA: mixed decay
- ▣ **Identification:** Visual analysis of ACF/PACF guides the selection of orders p and q

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 stationarity**
 - ▶ Parameters remain stable over time
 - ▶ Autocorrelation structure is maintained
- **Non-stationary data** \succ difference first (ARIMA, Ch. 3)

Chapter Objective

- Parametric models for stationary series \succ combining dependence on past observations (AR) with the influence of random shocks (MA)

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: $LX_t = X_{t-1}$

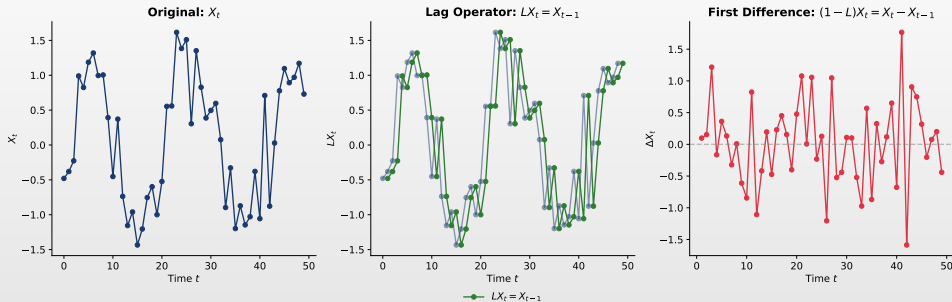
Properties

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

Lag Polynomials

- **AR polynomial:** $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$
- **MA polynomial:** $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q$

The Lag Operator: Visual Illustration



Role of the Lag Operator

- ▣ **Notation foundation:** Enables compact writing of difference equations
- ▣ **Utility:** Facilitates algebraic manipulation of ARMA models

The White Noise Process

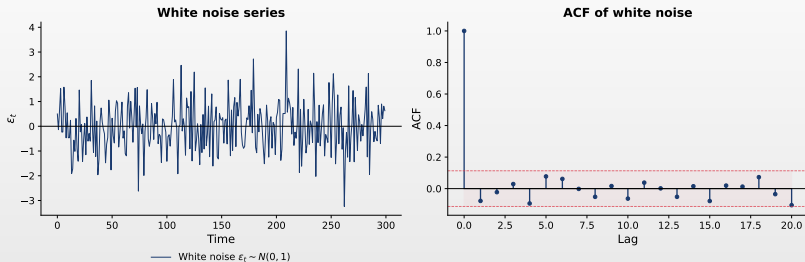
Definition 2 (White Noise)

- 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. $\text{Var}(\varepsilon_t) = \sigma^2$ for all t
 3. $\text{Cov}(\varepsilon_t, \varepsilon_s) = 0$ for all $t \neq s$

Properties

- **Building block:** White noise underlies all ARMA models
- **ACF:** $\rho(0) = 1$, $\rho(h) = 0$ for $h \neq 0$; PACF: same pattern
- **Gaussian white noise:** $\varepsilon_t \sim N(0, \sigma^2)$ i.i.d.
- **Unpredictable:** White noise is *not* predictable \succ it is purely random

White Noise: Visual Illustration



Key Characteristics

- Left: Random fluctuations, no patterns, constant variance
- Right: ACF only a spike at lag 0; others within significance bounds \Rightarrow no linear dependence

The AR(1) Model: Definition

Definition 3 (AR(1) Process)

- An **autoregressive process of order 1** is: $X_t = c + \phi X_{t-1} + \varepsilon_t$
- $\varepsilon_t \sim WN(0, \sigma^2)$ and $|\phi| < 1$ for stationarity

Interpretation

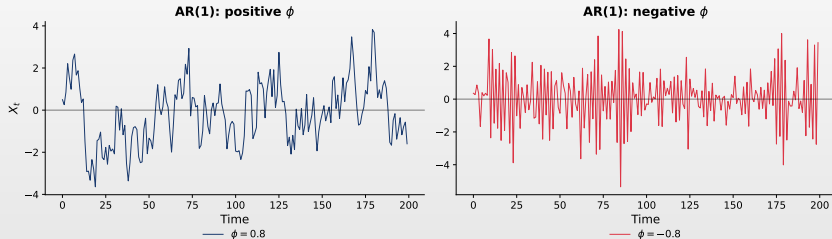
- c : constant (intercept)
- ϕ : autoregressive coefficient
 - ▶ Measures the persistence of the series
- ε_t : innovation (shock)

Lag Operator Notation

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

AR(1): Visual Illustration

AR(1): different behavior for positive vs negative ϕ



Visual Interpretation

- ▣ **Positive ϕ :** Persistent fluctuations, gradual mean reversion
- ▣ **Negative ϕ :** Oscillating behavior, alternating around the mean
- ▣ **Larger $|\phi|$ \succ greater persistence, slower reversion**

AR(1) Stationarity Condition

Necessary and Sufficient Condition: $|\phi| < 1$

- The root of the characteristic equation must lie outside the unit circle

- Shocks diminish over time
 - ▶ Process reverts to the mean
 - ▶ Finite, stable variance

Non-stationary ($|\phi| \geq 1$)

- $|\phi| = 1$: random walk
 - ▶ Unit root, variance $\rightarrow \infty$
- $|\phi| > 1$: explosive process

Characteristic Equation

- $\phi(z) = 1 - \phi z = 0 \implies z = 1/\phi$
- Stationarity \Leftrightarrow root outside the unit circle ($|z| > 1$)

AR(1) Properties

Stationary AR(1) with $|\phi| < 1$

□ Moment properties:

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

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

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

Autocorrelation (ACF): $\rho(h) = \phi^h$

Key Observation

□ **AR(1) signature:** ACF decays exponentially with factor ϕ

- ▶ $\phi > 0$: monotone decay towards zero
- ▶ $\phi < 0$: damped oscillations (alternating signs)

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: $\mathbb{E}[X_t] = c + \phi\mathbb{E}[X_{t-1}] + \mathbb{E}[\varepsilon_t]$
- By stationarity, $\mathbb{E}[X_t] = \mathbb{E}[X_{t-1}] = \mu$, and $\mathbb{E}[\varepsilon_t] = 0$: $\mu = c + \phi\mu$
- Solving: $\mu - \phi\mu = c \implies \mu(1 - \phi) = c \implies \boxed{\mu = \frac{c}{1 - \phi}}$

Requirement

- **Necessary condition:** $\phi \neq 1$ for the mean to be defined
 - ▶ If $\phi = 1$ (unit root), the mean is undefined
 - ▶ The process becomes a random walk (non-stationarity)

Proof: AR(1) Variance

Claim

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

Proof

\square Assume $c = 0$. Take the variance of $X_t = \phi X_{t-1} + \varepsilon_t$:

$$\square \text{Var}(X_t) = \phi^2 \text{Var}(X_{t-1}) + \text{Var}(\varepsilon_t) + 2\phi \underbrace{\text{Cov}(X_{t-1}, \varepsilon_t)}_{=0}$$

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

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

Note

\square Requires $|\phi| < 1$ for positive variance. When $|\phi| \rightarrow 1$, variance $\rightarrow \infty$

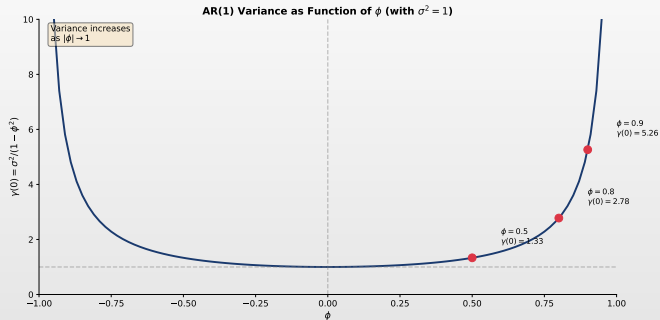
Proof: AR(1) Autocorrelation Function

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

- Find the autocovariance $\gamma(h) = \text{Cov}(X_t, X_{t-h})$

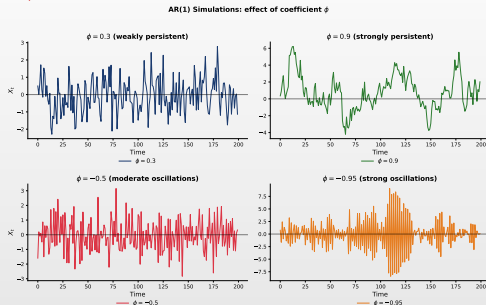
Proof

- 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 $h \geq 1$: $\mathbb{E}[\varepsilon_t X_{t-h}] = 0 \succ \gamma(h) = \phi \gamma(h-1)$
- Recursive relation from $\gamma(0)$: $\gamma(1) = \phi \gamma(0)$, $\gamma(2) = \phi^2 \gamma(0)$, ... $\gamma(h) = \phi^h \gamma(0)$
- ACF: $\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{\phi^h \gamma(0)}{\gamma(0)} = \boxed{\phi^h}$

AR(1) Variance as a Function of ϕ 

Observations

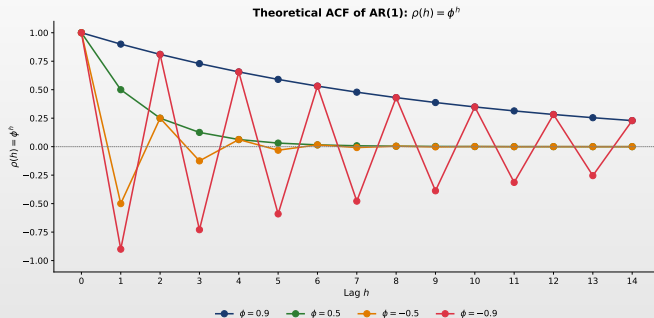
- As $|\phi| \rightarrow 1$, the variance explodes \succ non-stationarity
- For $\phi = 0$: $\gamma(0) = \sigma^2$ (white noise); variance increases monotonically with $|\phi|$

AR(1) Simulations: Effect of ϕ 

Interpretation

- Different values of ϕ produce distinct behaviors: larger $|\phi|$ \succ more persistence
- Positive ϕ creates smooth trajectories; negative ϕ creates oscillations
- As $|\phi| \rightarrow 1$, the process approaches non-stationarity

Theoretical AR(1) ACF



ACF Pattern

- Formula: $\rho(h) = \phi^h$ \succ exponential decay
- $\phi > 0$: monotone decay; $\phi < 0$: alternating signs

Proof: AR(1) Stationarity Condition

Claim

- AR(1) is stationary if and only if $|\phi| < 1$

Proof

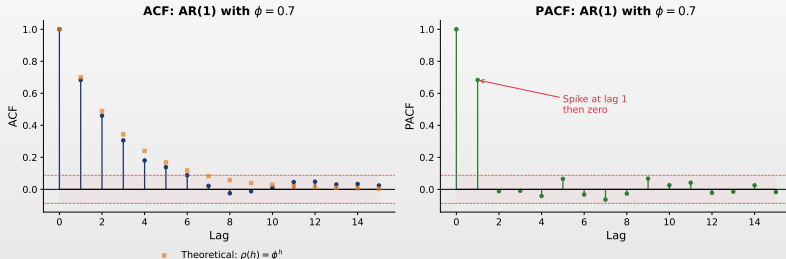
- Recursive substitution: $X_t = \phi X_{t-1} + \varepsilon_t = \phi(\phi X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \dots$
- After n steps: $X_t = \phi^n X_{t-n} + \sum_{j=0}^{n-1} \phi^j \varepsilon_{t-j}$
- If $|\phi| < 1$: $\phi^n \rightarrow 0$ as $n \rightarrow \infty$, so $X_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$
- Finite variance: $\text{Var}(X_t) = \sigma^2 \sum_{j=0}^{\infty} \phi^{2j} = \frac{\sigma^2}{1-\phi^2} < \infty$ (geometric series)

Conclusion

- Converges $\iff |\phi| < 1$. For $|\phi| \geq 1$, the term $\phi^n X_{t-n}$ does not vanish \Rightarrow infinite variance

AR(1) ACF and PACF: Theory vs Sample

ACF and PACF for AR(1): theory vs sample



Interpretation

- ACF: Exponential decay with factor ϕ ; formula: $\rho(h) = \phi^h$
- PACF: A single spike at lag 1, then cuts off \succ identifies AR(1)
- Sample estimates fluctuate around theoretical values

The Partial Autocorrelation Function (PACF)

Definition 4 (PACF)

- The **partial autocorrelation** of order k , denoted π_k , measures the correlation between X_t and X_{t-k} **after removing** the linear effects of the intermediate variables $X_{t-1}, \dots, X_{t-k+1}$

Formal Definition

- $\pi_1 = \rho(1)$
- For $k \geq 2$: π_k is the last coefficient in:
$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_k X_{t-k} + e_t$$
- $\pi_k = \alpha_k$ (coefficient of X_{t-k})

Computation via Yule-Walker

- Solve the Yule-Walker equations of order k
- π_k = last element of the solution vector

Utility

- **Identification:** PACF determines the order p of an AR model
 - PACF cuts off after lag p

AR(1) ACF and PACF Patterns

ACF of AR(1)

- Decays exponentially: $\rho(h) = \phi^h$
 - $\phi > 0$: all positive
 - $\phi < 0$: alternating signs

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

Key Pattern

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

The AR(p) Model: General Form

Definition 5 (AR(p) Process)

- ▣ An **autoregressive process of order p** is: $X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t$
- ▣ **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

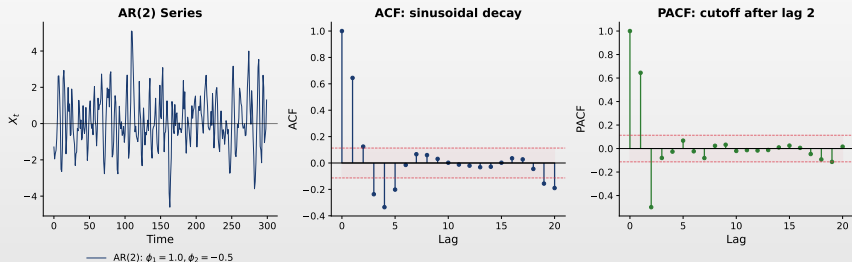
- ▣ All roots of $\phi(z) = 0$ must lie **outside** the unit circle
- ▣ Equivalently: all roots have modulus > 1

PACF Pattern

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

AR(p): Visual Illustration

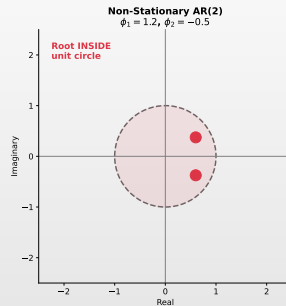
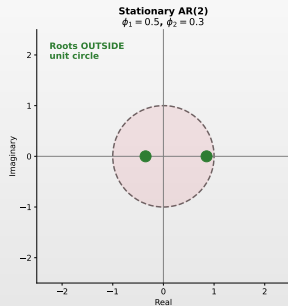
AR(2) Process: pseudo-cyclic behavior



Observations

- AR(2) can exhibit pseudo-cyclic behavior (complex roots); damped sinusoidal ACF
- PACF cuts off after lag 2 \succ key identification pattern

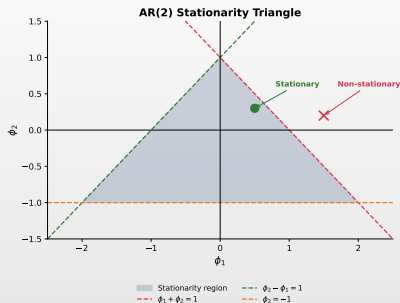
AR(2) Stationarity: Unit Circle Visualization



Characteristic Polynomial and Unit Circle Condition

- **Characteristic polynomial** of an $AR(p)$ process: $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p$
- All roots of $\phi(z) = 0$ must lie **outside** the unit circle ($|z| > 1$)
- Roots on the circle: non-stationary; roots inside: explosive process

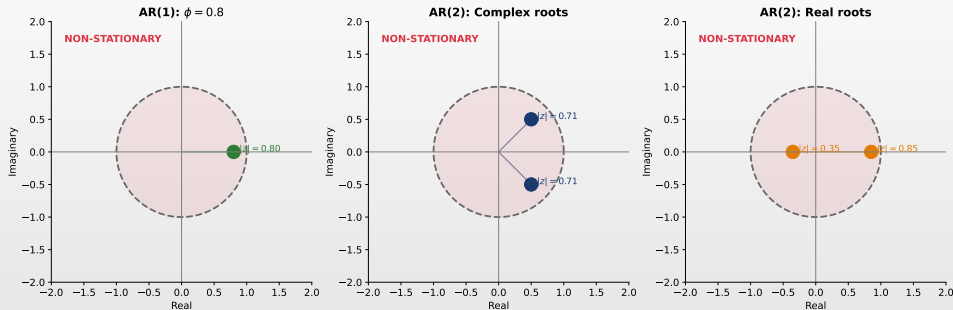
The AR(2) Stationarity Triangle



Stationarity Region

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

Characteristic Polynomial Roots



Types of Roots

- Real roots: exponential decay in ACF
- Complex roots: damped oscillations (pseudo-cycles)
- All roots must lie outside the unit circle

The AR(2) Model

Definition 6 (AR(2) Process)

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

Stationarity Conditions

$$\square \phi_1 + \phi_2 < 1; \quad \phi_2 - \phi_1 < 1; \quad |\phi_2| < 1$$

ACF Behavior

- \square **Real roots:** mixture of two exponential decays
- \square **Complex roots:** damped sinusoidal pattern (pseudo-cycles)
- \square **PACF:** Cuts off after lag 2 ($\pi_k = 0$ for $k > 2$)

The MA(1) Model: Definition

Definition 7 (MA(1) Process)

- ▣ A **moving average process of order 1** is: $X_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$
- ▣ $\varepsilon_t \sim WN(0, \sigma^2)$

Interpretation

- ▣ μ : process mean
- ▣ θ : MA coefficient
 - Measures the impact of the past shock
- ▣ Depends on ε_t and ε_{t-1}

Lag Operator Notation

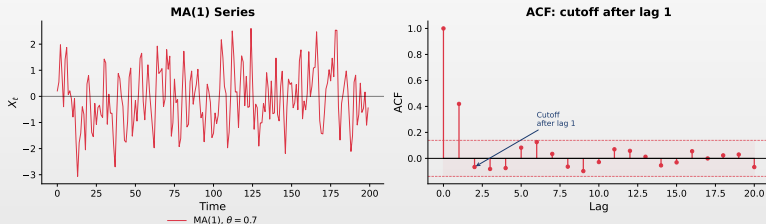
- ▣ $X_t = \mu + \theta(L)\varepsilon_t$
- ▣ $\theta(L) = 1 + \theta L$

Key Property

- ▣ **Guaranteed stationarity:** MA processes are always stationary
 - Does not depend on the value of θ

MA(1): Visual Illustration

MA(1): short memory series with ACF cutoff



Visual Interpretation

- Left panel: MA(1) series \succ rapid mean reversion
- Right panel: ACF with **cutoff after lag 1**; PACF exponential decay

MA(1) Properties

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

- **Mean:** $\mathbb{E}[X_t] = \mu$; **Variance:** $\gamma(0) = \sigma^2(1 + \theta^2)$
- **Autocovariance:** $\gamma(1) = \theta\sigma^2$, $\gamma(h) = 0$ ($h > 1$)
- **ACF:** $\rho(1) = \frac{\theta}{1+\theta^2}$, $\rho(h) = 0$ ($h > 1$)

Key Observation

- **MA(1) signature:** ACF cuts off after lag 1
 - ▶ $\rho(1) \neq 0$, but $\rho(h) = 0$ for $h > 1$; opposite pattern to AR(1)

Proof: MA(1) Variance and Autocovariance

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

▣ **Variance:**

$$\gamma(0) = \text{Var}(\varepsilon_t + \theta\varepsilon_{t-1}) = \sigma^2 + \theta^2\sigma^2 + 0 = \boxed{\sigma^2(1 + \theta^2)}$$

Autocovariance at lag 1

$$\square \gamma(1) = \text{Cov}(\varepsilon_t + \theta\varepsilon_{t-1}, \varepsilon_{t-1} + \theta\varepsilon_{t-2})$$

$$\square = \text{Cov}(\varepsilon_t, \varepsilon_{t-1}) + \theta\text{Cov}(\varepsilon_t, \varepsilon_{t-2}) + \theta\text{Cov}(\varepsilon_{t-1}, \varepsilon_{t-1}) + \theta^2\text{Cov}(\varepsilon_{t-1}, \varepsilon_{t-2})$$

$$\square = 0 + 0 + \theta\sigma^2 + 0 = \boxed{\theta\sigma^2}$$

Autocovariance at lag $h \geq 2$

$$\square \text{ No common } \varepsilon \text{ terms } \succ \gamma(h) = 0$$

Proof: Maximum ACF for MA(1)

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

- ACF at lag 1: $\rho(1) = \frac{\theta}{1+\theta^2}$
- Differentiate: $\frac{d\rho(1)}{d\theta} = \frac{1-\theta^2}{(1+\theta^2)^2} = 0 \succ \theta = \pm 1$
- At these values: $\rho(1)|_{\theta=1} = \frac{1}{2}$, $\rho(1)|_{\theta=-1} = -\frac{1}{2}$

Implication

- **Practical test:** If $|\hat{\rho}(1)| > 0.5$ from data, the process is **not** MA(1)
 - ▶ The maximum $|\rho(1)| = 0.5$ is reached at $\theta = \pm 1$
 - ▶ Consider AR or ARMA models as alternatives

Proof: ACF for MA(1)

Claim: $\rho(1) = \frac{\theta}{1+\theta^2}$ and $\rho(h) = 0$ for $h > 1$

- MA(1) has non-zero autocorrelation **only** at lag 1

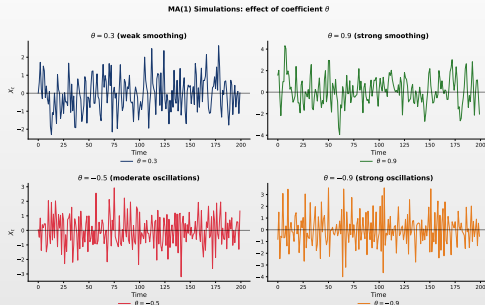
Proof

- Let $X_t = \varepsilon_t + \theta\varepsilon_{t-1}$. Autocorrelation at lag 1:
- $\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{\theta\sigma^2}{\sigma^2(1+\theta^2)} = \frac{\theta}{1+\theta^2}$
- For $h > 1$: $\gamma(h) = \text{Cov}(\varepsilon_t + \theta\varepsilon_{t-1}, \varepsilon_{t-h} + \theta\varepsilon_{t-h-1})$
- The terms $\varepsilon_t, \varepsilon_{t-1}$ do not overlap with $\varepsilon_{t-h}, \varepsilon_{t-h-1}$ when $h > 1$, so $\gamma(h) = 0$

Practical Consequence

- ACF cuts off sharply after lag 1 \Rightarrow distinctive signature of MA(1) processes

MA(1) Simulations: Effect of θ



Interpretation

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

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 $h > 1$
 - ▶ $|\rho(1)| \leq 0.5$ always

PACF of MA(1)

- ▣ **Decays exponentially**
 - ▶ Or with alternating signs
 - ▶ Does *not* cut off

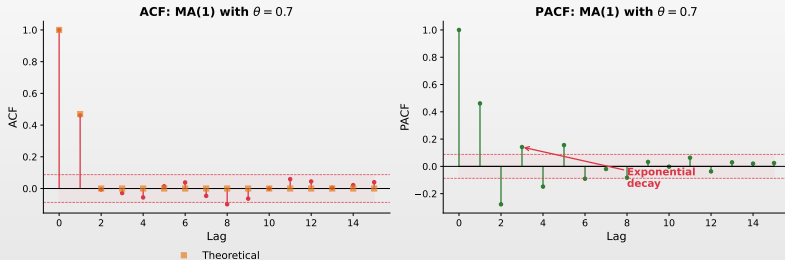
	ACF	PACF
MA(1)	Cuts off at lag 1	Exponential decay

Observation

- ▣ **Opposite pattern to AR(1)!**

MA(1) ACF and PACF: Visual Comparison

ACF and PACF for MA(1): opposite pattern to AR(1)



Interpretation

- ACF: A single spike at lag 1, then cuts off \succ key MA(1) signature
- PACF: Exponential decay \succ opposite pattern to AR(1)
- This reversal differentiates MA processes from AR processes

Invertibility of MA Models

Definition 8 (Invertibility)

- An MA process is **invertible** if it can be written as an infinite AR process:
- $X_t = \mu + \sum_{j=1}^{\infty} \pi_j (X_{t-j} - \mu) + \varepsilon_t$

Invertibility Conditions

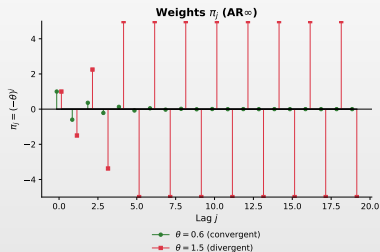
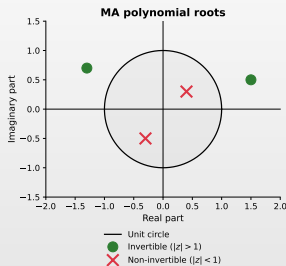
- **MA(1)**: Invertible if $|\theta| < 1$
- **MA(q)**: Roots of $\theta(z) = 0$ outside the unit circle

Why Invertibility Matters

- Ensures unique representation (without invertibility, multiple MA models describe the same data)
- Necessary for forecasting and estimation
- **Stationarity** \succ AR; **Invertibility** \succ MA

Invertibility: Visual Illustration

Invertibility of MA models



Interpretation

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

Proof: MA(1) Invertibility

Claim

- MA(1) is invertible if and only if $|\theta| < 1$

Proof

- From $X_t = \varepsilon_t + \theta\varepsilon_{t-1}$, isolate: $\varepsilon_t = X_t - \theta\varepsilon_{t-1}$
- Recursive back-substitution: $\varepsilon_t = X_t - \theta(X_{t-1} - \theta\varepsilon_{t-2}) = X_t - \theta X_{t-1} + \theta^2\varepsilon_{t-2}$
- Continuing: $\varepsilon_t = \sum_{j=0}^n (-\theta)^j X_{t-j} + (-\theta)^{n+1} \varepsilon_{t-n-1}$
- If $|\theta| < 1$: $(-\theta)^{n+1} \rightarrow 0$, so

$$\varepsilon_t = \sum_{j=0}^{\infty} (-\theta)^j X_{t-j}$$

Conclusion

- Geometric series converges $\iff |\theta| < 1 \Rightarrow$ MA(1) can be written as AR(∞)

The MA(q) Model: General Form

Definition 9 (MA(q) Process)

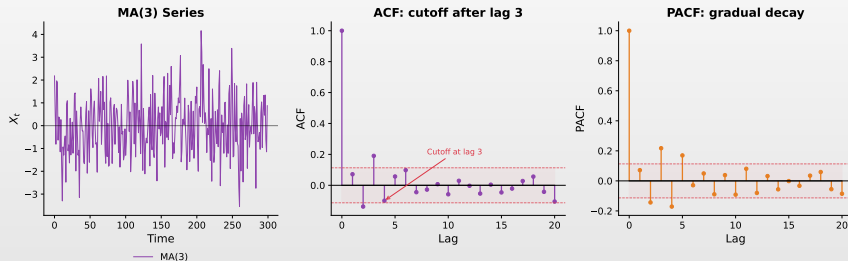
- ▣ A **moving average process of order q**: $X_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q}$
- ▣ **Lag operator**: $X_t = \mu + \theta(L)\varepsilon_t$, where $\theta(L) = 1 + \theta_1L + \cdots + \theta_qL^q$

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 the unit circle

MA(q): Visual Illustration

MA(q) Process: ACF signature cuts off after lag q



Observation

- MA(3) process: key signature \succ ACF cuts off after lag q ($\rho(h) = 0$ for $h > 3$)

The ARMA(p,q) Model: Definition

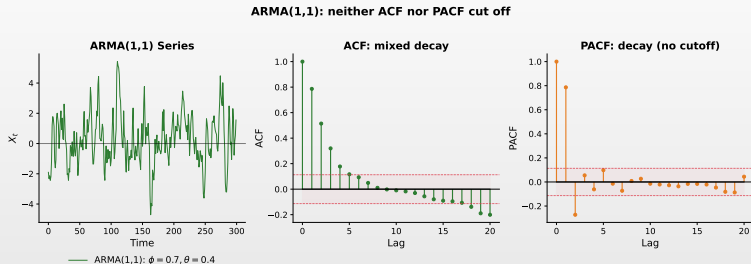
Definition 10 (ARMA(p,q) Process)

- $X_t = c + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$
- **Compact form:** $\phi(L)X_t = c + \theta(L)\varepsilon_t$, where $\mu = \frac{c}{1 - \phi_1 - \cdots - \phi_p}$

Key Idea

- **Flexibility:** Combines AR and MA components
 - ▶ AR captures persistence; MA captures shock response
- **Parsimony:** ARMA(1,1) can be better than AR(5) or MA(5)
 - ▶ Fewer parameters, less risk of overfitting

ARMA: Visual Illustration



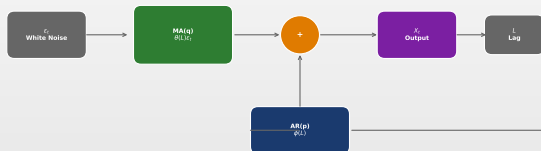
ARMA(1,1) Interpretation

- ▣ **Combines** AR persistence with MA shock response
- ▣ **ACF pattern:** Decay after the first lag (lags decay geometrically)
- ▣ **PACF pattern:** Also decays (no sharp cutoff as in pure AR)
- ▣ Neither ACF nor PACF cuts off \succ key identifier for mixed models

ARMA Model Structure

ARMA(p,q) Model Structure

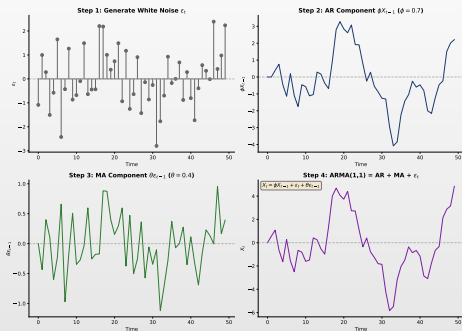
$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$



Components

- ▣ **AR component:** influence of past values of the series
- ▣ **MA component:** impact of past random shocks

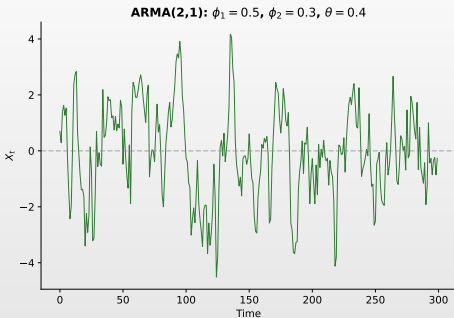
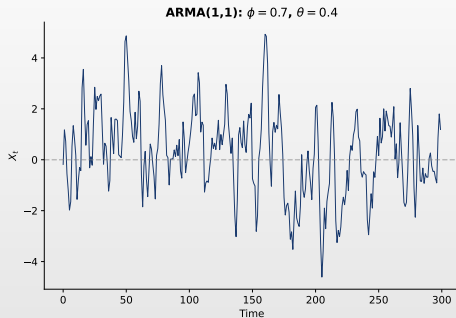
How ARMA Simulation Works



Steps

- Generate white noise, apply the ARMA equation recursively, obtain simulated series

ARMA Examples



Observation

- Different combinations of orders (p, q) produce distinct behaviors

The ARMA(1,1) Model

Definition 11 (ARMA(1,1) Process)

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

Properties (stationarity and invertibility)

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

ACF

$$\boxed{\cdot} \quad \rho(1) = \frac{(1+\phi\theta)(\phi+\theta)}{1+2\phi\theta+\theta^2}; \quad \rho(h) = \phi \cdot \rho(h-1) \text{ for } h \geq 2$$

$\boxed{\cdot}$ ACF decays exponentially after lag 1 (starting point depends on ϕ and θ)

Proof: ARMA(1,1) Variance

Claim

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

Proof

- Let $Y_t = X_t - \mu$: $Y_t = \phi Y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$
- Square: $Y_t^2 = \phi^2 Y_{t-1}^2 + \varepsilon_t^2 + \theta^2 \varepsilon_{t-1}^2 + 2\phi Y_{t-1} \varepsilon_t + 2\phi\theta Y_{t-1} \varepsilon_{t-1} + 2\theta \varepsilon_t \varepsilon_{t-1}$
- Take expectations; $\mathbb{E}[\varepsilon_t Y_{t-1}] = 0$, $\mathbb{E}[\varepsilon_t \varepsilon_{t-1}] = 0$:
- $\gamma(0) = \phi^2 \gamma(0) + \sigma^2 + \theta^2 \sigma^2 + 2\phi\theta \mathbb{E}[\varepsilon_{t-1} Y_{t-1}]$
- From $Y_{t-1} = \phi Y_{t-2} + \varepsilon_{t-1} + \theta \varepsilon_{t-2}$: only ε_{t-1}^2 contributes $\Rightarrow \mathbb{E}[\varepsilon_{t-1} Y_{t-1}] = \sigma^2$
- $\gamma(0)(1 - \phi^2) = (1 + 2\phi\theta + \theta^2)\sigma^2 \implies \boxed{\gamma(0) = \frac{(1 + 2\phi\theta + \theta^2)\sigma^2}{1 - \phi^2}}$

Proof: ARMA(1,1) ACF at Lag 1

Claim

$$\square \quad \rho(1) = \frac{(1+\phi\theta)(\phi+\theta)}{1+2\phi\theta+\theta^2}; \quad \rho(h) = \phi \rho(h-1) \text{ for } h \geq 2$$

Proof

□ Multiply Y_t by Y_{t-1} and take expectations:

$$\square \quad \gamma(1) = \phi\gamma(0) + \underbrace{\mathbb{E}[\varepsilon_t Y_{t-1}]}_{=0} + \theta \underbrace{\mathbb{E}[\varepsilon_{t-1} Y_{t-1}]}_{=\sigma^2} = \phi\gamma(0) + \theta\sigma^2$$

□ Divide by $\gamma(0)$: $\rho(1) = \phi + \frac{\theta\sigma^2}{\gamma(0)}$. Substitute $\gamma(0)$:

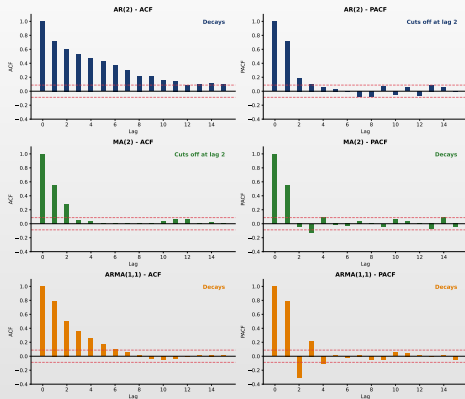
$$\square \quad \rho(1) = \phi + \frac{\theta(1-\phi^2)}{1+2\phi\theta+\theta^2} = \frac{\phi(1+2\phi\theta+\theta^2)+\theta(1-\phi^2)}{1+2\phi\theta+\theta^2}$$

$$\square \quad \text{Numerator: } \phi + \theta + \phi^2\theta + \phi\theta^2 = (\phi + \theta)(1 + \phi\theta), \text{ so } \boxed{\rho(1) = \frac{(1 + \phi\theta)(\phi + \theta)}{1 + 2\phi\theta + \theta^2}}$$

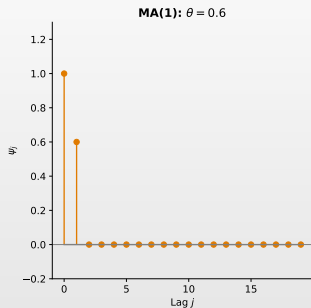
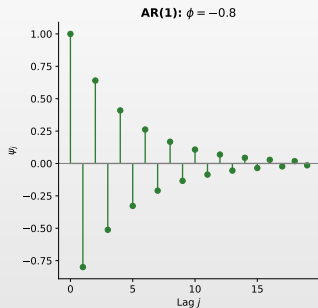
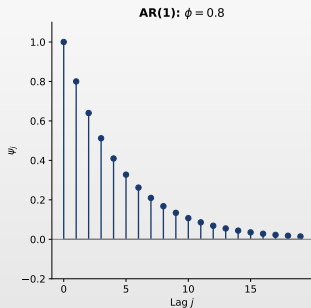
Recursion

□ For $h \geq 2$: $\gamma(h) = \phi\gamma(h-1)$, so $\rho(h) = \phi \rho(h-1) \Rightarrow$ exponential decay from lag 1

ACF/PACF Patterns: AR vs MA vs ARMA



Impulse Response Functions



Shock Propagation

- ▣ Shows how a unit shock propagates through the system over time
- ▣ **AR**: exponential or oscillating decay; **MA**: effect limited to q periods

Stationarity and Invertibility Summary

Conditions for a Valid ARMA(p,q) Model

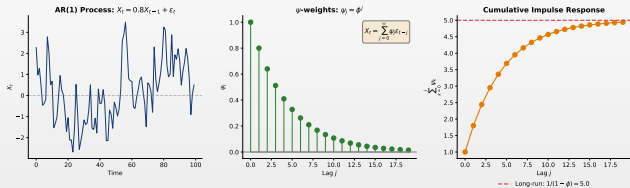
▣ Requirements summary:

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

Implications

- ▣ **Stationarity:** Can be written as $\text{MA}(\infty)$: $X_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$
- ▣ **Invertibility:** Can be written as $\text{AR}(\infty)$: $X_t = \mu + \sum_{j=1}^{\infty} \pi_j (X_{t-j} - \mu) + \varepsilon_t$
- ▣ **Causal representation:** X_t depends only on *past* shocks \succ necessary for forecasting

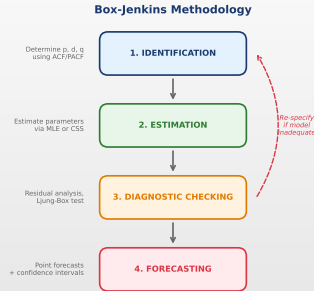
Wold's Decomposition Theorem



Wold's Theorem

- Any purely non-deterministic stationary process can be written as MA(∞):
- $X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ with $\sum \psi_j^2 < \infty$
- Theoretical justification for ARMA modeling

The Box-Jenkins Methodology



Iterative Approach

- Identification \succ estimation \succ validation; repeat until residuals are white noise



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

Parsimony Principle

- Start simple (small p, q), increase order if checks are not satisfied
- Simpler models are preferred

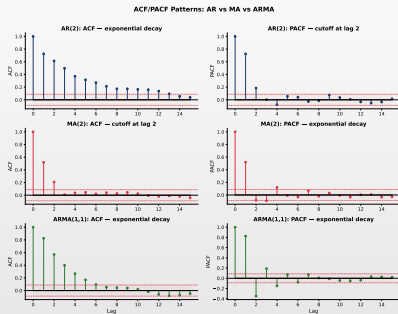
ACF/PACF Identification Rules

Theoretical Patterns for Stationary Processes

- The table summarizes ACF/PACF patterns for model identification:

Model	ACF Pattern	PACF Pattern
AR(1)	Exponential decay	Spike at lag 1, then 0
AR(2)	Exp./damped sinusoid	Spikes at lags 1-2, then 0
AR(p)	Gradual decay	Cuts off after lag p
MA(1)	Spike at lag 1, then 0	Exponential decay
MA(2)	Spikes at lags 1-2, then 0	Exp./damped sinusoid
MA(q)	Cuts off after lag q	Gradual decay
ARMA(p,q)	Decays	Decays

ACF/PACF Patterns: Visual Guide



Identification Rules

- **AR**: ACF decays, PACF cuts off \succ use PACF for order p
- **MA**: ACF cuts off, PACF decays \succ use ACF for order q
- **ARMA**: Both decay \succ requires information criteria

Information Criteria

AIC (Akaike)

- $AIC = -2 \ln(\hat{L}) + 2k$
- Moderate penalty
 - ▶ Tends to select larger models
 - ▶ Optimal for forecasting

BIC (Bayesian)

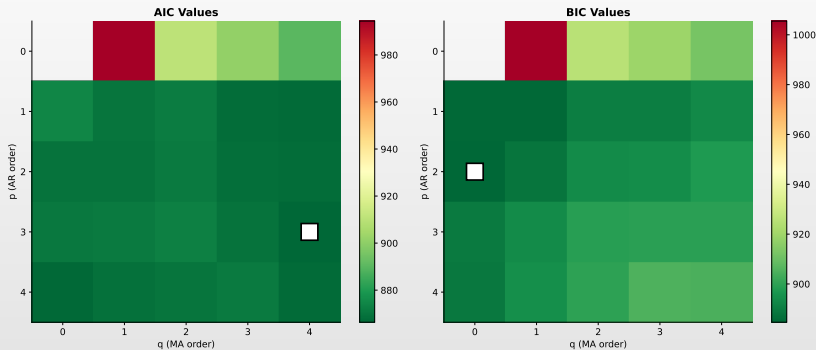
- $BIC = -2 \ln(\hat{L}) + k \ln(n)$
- Stronger penalty
 - ▶ Prefers parsimonious models
 - ▶ Consistent for identification

where: \hat{L} = maximum of the likelihood function, k = number of estimated parameters, n = sample size

Rules

- Lower values = better model. Compare models on the *same data*

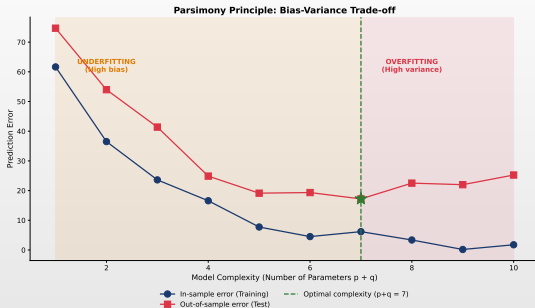
AIC vs BIC: Model Selection



Interpretation

- White square marks the best model; lower values (green) are better

Parsimony Principle: Bias-Variance Trade-off



Bias-Variance Trade-off

- Too simple model \succ high bias (underfitting)
- Too complex model \succ high variance (overfitting)
- The optimum lies at the intersection of the two curves

Automatic Model Selection

Grid Search Approach

- ▣ Estimate ARMA(p, q) for $p = 0, \dots, p_{max}$ and $q = 0, \dots, q_{max}$
- ▣ Select the model with the lowest AIC or BIC; verify with validation tests

In Python

- ▣ `pm.auto_arima()` from the `pmdarima` package
- ▣ Automatically tests stationarity, iterates over orders (p, q) , returns the best model

Caution

- ▣ Automatic selection is not the final answer \succ verify model validity
- ▣ Full Auto-ARIMA (including selection of d) \succ Chapter 3

Estimation Methods Overview

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

- ▣ Equates sample autocorrelations with theoretical values
- ▣ Simple, closed-form for AR models; not efficient for MA

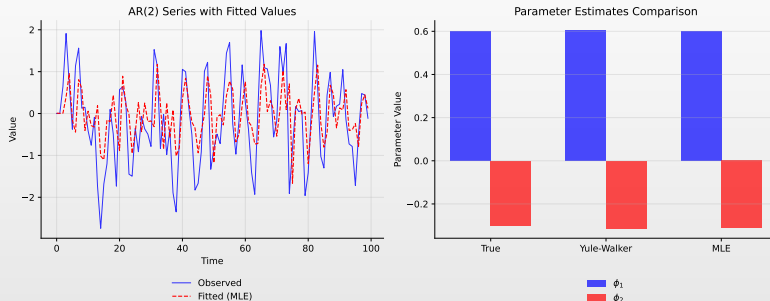
2. Maximum Likelihood Estimation (MLE)

- ▣ Most common approach; requires distributional assumption (Gaussian)
- ▣ Efficient and consistent

3. Conditional Least Squares

- ▣ Minimizes the sum of squared residuals
- ▣ Conditional on initial observations; algorithmically simpler than exact MLE

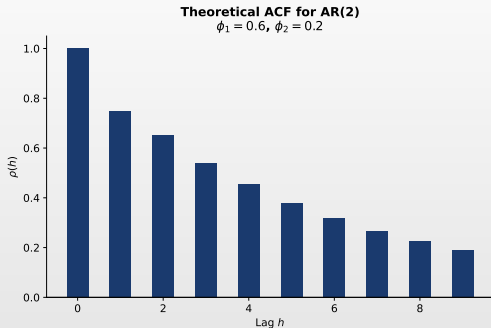
Estimation Methods Comparison



Comparison

- **MLE:** most efficient, but requires distributional assumption
- **Yule-Walker:** closed-form, only for AR models
- **CLS:** compromise between MLE and Yule-Walker

The 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

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

Main Idea

- Linear relationship between autocorrelations and AR parameters
- Allows closed-form estimation (no numerical optimization)

The Yule-Walker Equations: Matrix Form

Yule-Walker Equations for AR(p)

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

Matrix Form

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

\square **Estimation:** Replace $\rho(k)$ with $\hat{\rho}(k)$; the Toeplitz matrix is symmetric and positive definite

Numerical Example: Yule-Walker for AR(2)

Sample Data ($T = 100$)

▣ **Estimated autocorrelations:** $\hat{\rho}(1) = 0.75$, $\hat{\rho}(2) = 0.65$

▶ Estimated variance: $\hat{\gamma}(0) = 4.0$

Step 1: Matrix System

▣ **Yule-Walker:** $R\hat{\phi} = \rho$

▶
$$\begin{pmatrix} 1 & 0.75 \\ 0.75 & 1 \end{pmatrix} \begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{pmatrix} = \begin{pmatrix} 0.75 \\ 0.65 \end{pmatrix}$$

Step 2: Solution (Cramer's Rule)

▣ $\det(R) = 1 - 0.75^2 = 0.4375$

▣ $\hat{\phi}_1 = \frac{0.75 \times 1 - 0.75 \times 0.65}{0.4375} = \frac{0.2625}{0.4375} = \boxed{0.600}$ $\hat{\phi}_2 = \frac{0.65 \times 1 - 0.75 \times 0.75}{0.4375} = \frac{0.0875}{0.4375} = \boxed{0.200}$

Step 3: Noise Variance

▣ $\hat{\sigma}^2 = \hat{\gamma}(0)(1 - \hat{\phi}_1\hat{\rho}(1) - \hat{\phi}_2\hat{\rho}(2)) = 4.0(1 - 0.45 - 0.13) = \boxed{1.68}$

Stationarity check: $\hat{\phi}_1 + \hat{\phi}_2 = 0.8 < 1 \checkmark$ $|\hat{\phi}_2| = 0.2 < 1 \checkmark$ $\hat{\phi}_2 - \hat{\phi}_1 = -0.4 > -1 \checkmark$

Proof: The Yule-Walker Equations

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

- Start from AR(p): $X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t$
- Multiply 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 $k \geq 1$: $\mathbb{E}[\varepsilon_t X_{t-k}] = 0 \succ \gamma(k) = \phi_1 \gamma(k-1) + \cdots + \phi_p \gamma(k-p)$
- Dividing by $\gamma(0)$: $\rho(k) = \phi_1 \rho(k-1) + \phi_2 \rho(k-2) + \cdots + \phi_p \rho(k-p)$

Special Case AR(1)

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

Maximum Likelihood Estimation

ARMA(p,q) Log-Likelihood (Gaussian errors: $\varepsilon_t \sim N(0, \sigma^2)$)

- ▣ $\ell(\phi, \theta, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^n \varepsilon_t^2$
- ▣ ε_t are the innovations computed recursively

Estimation Procedure

- ▣ Initialization: use method of moments or OLS for starting values
- ▣ Optimization: numerical methods (BFGS, Newton-Raphson)
- ▣ Iterate until convergence

In Practice

- ▣ `statsmodels.tsa.arima.model.ARIMA` implements exact MLE with automatic initialization

Standard Errors and Inference

Asymptotic Distribution of MLE

- ▣ $\hat{\theta} \xrightarrow{d} N(\theta_0, \frac{1}{n}I(\theta_0)^{-1})$, where $I(\theta)$ is the **Fisher information matrix**
- ▣ $I(\theta) = -E\left[\frac{\partial^2 \ln L(\theta)}{\partial \theta \partial \theta'}\right] \succ$ average curvature of the log-likelihood
- ▣ Estimated variance-covariance matrix: $\hat{V} = \frac{1}{n}\hat{I}^{-1}$

What is the Standard Error (SE)?

- ▣ $SE(\hat{\theta}_j) = \sqrt{\hat{V}_{jj}} = \sqrt{\text{diag}_j\left(\frac{1}{n}\hat{I}^{-1}\right)} \succ$ measures estimation uncertainty
- ▣ **Example AR(1):** $SE(\hat{\phi}) \approx \sqrt{(1 - \hat{\phi}^2)/n}$; for $\hat{\phi} = 0.8$, $n = 100$: $SE \approx 0.06$
- ▣ **Interpretation:** small SE \Rightarrow parameter is estimated with high precision

Testing Parameter Significance

- ▣ $H_0 : \theta_j = 0$ Statistic: $z = \frac{\hat{\theta}_j}{SE(\hat{\theta}_j)} \sim N(0, 1)$ asymptotically
- ▣ Reject if $|z| > 1.96$ at 5% \Rightarrow **CI:** $\hat{\theta}_j \pm 1.96 \cdot SE(\hat{\theta}_j)$

Residual Analysis

If the model is correctly specified, residuals must be white noise

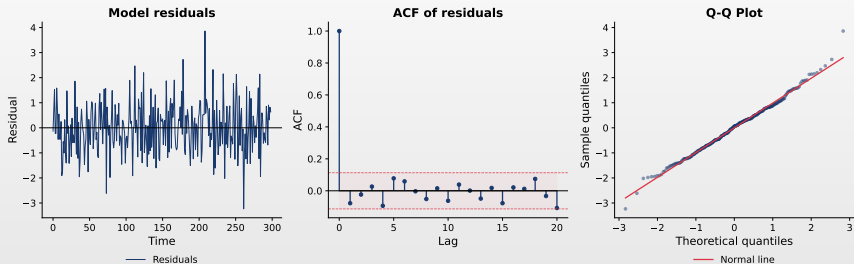
- ▣ **Residual time plot**
 - ▶ Fluctuates around zero, no obvious patterns; constant variance
- ▣ **Residual ACF**
 - ▶ All correlations within significance bounds; no significant spikes \succ white noise
- ▣ **Histogram / Q-Q plot**
 - ▶ Approximately normal distribution; heavy tails \succ non-normal errors

Decision

- ▣ ✓ All checks OK \succ adequate model
- ▣ × Not satisfied \succ return to identification

Residual Diagnostics: Example

AR(1) Model Diagnostics: white noise residuals



Interpretation

- **Residual plot:** random fluctuations around zero, constant variance
- **Residual ACF:** no significant spikes \leadsto white noise
- **Q-Q plot:** points on the diagonal \leadsto normally distributed residuals

The Ljung-Box Test

Definition 12 (Ljung-Box Test)

- ▣ Tests whether residuals are independently distributed (no autocorrelation)
- ▣ **Statistic:** $Q(m) = n(n+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{n-k}$

Hypotheses and Distribution

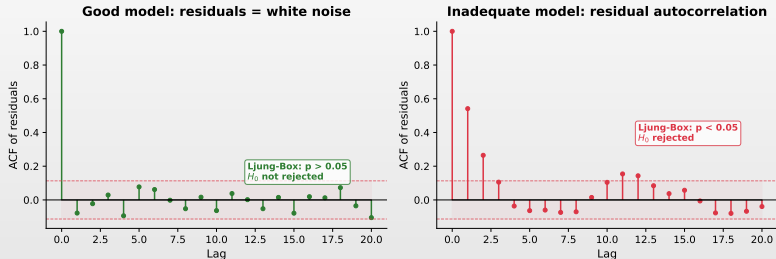
- ▣ H_0 : Residuals are white noise; H_1 : Residuals are autocorrelated
- ▣ Under H_0 , $Q(m) \sim \chi^2(m-p-q)$ approximately

Decision

- ▣ **p-value** > 0.05 \succ do not reject H_0 \succ residuals are white noise
- ▣ **p-value** < 0.05 \succ residual autocorrelation \succ inadequate model

The Ljung-Box Test: Visual Illustration

Ljung-Box Test: good model vs inadequate model



Interpretation

- Left: good model γ white noise residuals
- Right: inadequate model γ residual autocorrelation γ re-specification needed

Model Checklist

A Good ARMA Model Should Satisfy

- ▣ **Stationarity:** AR roots outside the unit circle (arroots)
- ▣ **Invertibility:** MA roots outside the unit circle (maroots)
- ▣ **White noise residuals:** No significant ACF (Ljung-Box test)
- ▣ **Normal residuals:** Q-Q plot, Jarque-Bera test
- ▣ **No heteroscedasticity:** Constant variance (ARCH test)
- ▣ **Simple:** Lowest AIC/BIC among adequate models

If Checks Are Not Satisfied

- ▣ Return to identification, try different orders

Point Forecasts

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

- The conditional expectation minimizes MSE

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

- $\hat{X}_{n+1|n} = c + \phi X_n$; $\hat{X}_{n+h|n} = \mu + \phi^h(X_n - \mu)$
- Forecasts converge to the mean μ as $h \rightarrow \infty$ (mean reversion)

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

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

Forecast Uncertainty

Mean Square Forecast Error (MSFE)

- **Error:** $e_{n+h|n} = X_{n+h} - \hat{X}_{n+h|n}$
- **MSFE:** $\text{MSFE}(h) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2$, where ψ_j are the $\text{MA}(\infty)$ coefficients

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

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

Key Observation

- Forecast uncertainty increases with the horizon
- Converges to the unconditional variance $\text{Var}(X_t)$

Proof: MSFE for AR(1)

Claim

$$\square \text{ MSFE}(h) = \sigma^2 \frac{1 - \phi^{2h}}{1 - \phi^2} \quad \text{and} \quad \text{MSFE}(\infty) = \gamma(0)$$

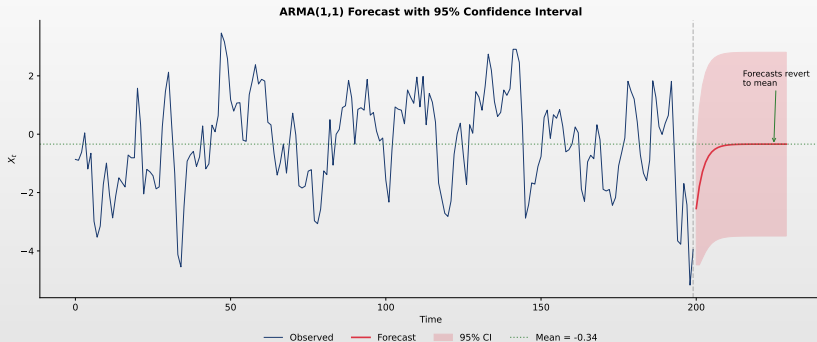
Proof

- \square Forecast error at horizon h : $e_{n+h|n} = X_{n+h} - \hat{X}_{n+h|n}$
- \square By recursive substitution: $e_{n+h|n} = \sum_{j=0}^{h-1} \phi^j \varepsilon_{n+h-j}$
- \square $\text{MSFE}(h) = \mathbb{E}[e_{n+h|n}^2] = \sigma^2 \sum_{j=0}^{h-1} \phi^{2j} = \boxed{\sigma^2 \frac{1 - \phi^{2h}}{1 - \phi^2}}$
- \square Limit: $\text{MSFE}(\infty) = \frac{\sigma^2}{1 - \phi^2} = \gamma(0) \Rightarrow$ forecast converges to unconditional mean

Interpretation

- \square At long horizons, we do no better than the unconditional mean: $\text{CI} \rightarrow 2 \times 1.96 \sqrt{\gamma(0)}$

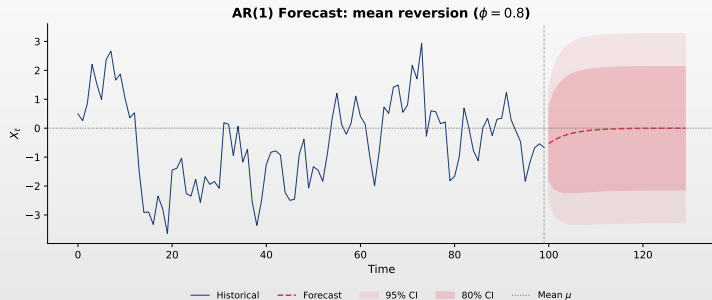
ARMA Forecast with Confidence Intervals



Observation

- The confidence band widens with the horizon \nearrow convergence to the unconditional interval

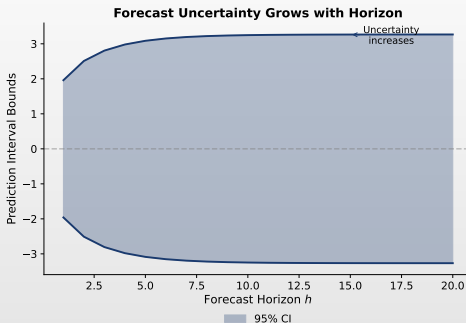
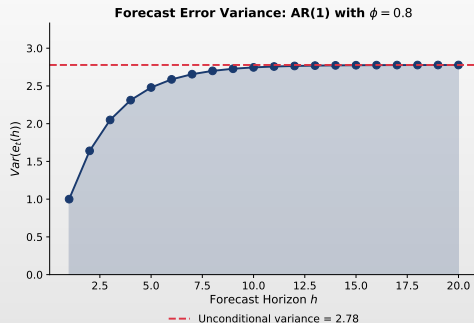
AR(1) Forecast: Mean Reversion



Properties

- ▣ Forecasts converge to the unconditional mean μ as the horizon increases
- ▣ Larger $|\phi|$ \succ slower reversion; CIs widen with the horizon

Forecast Error Variance by Horizon



Observation

- MSFE increases monotonically with horizon h \rightarrow convergence to $\text{Var}(X_t)$ (predictability limit)

Confidence Intervals for Forecasts

Formulas

- ▣ $X_{n+h}|X_n, \dots \sim N(\hat{X}_{n+h|n}, \text{MSFE}(h))$
- ▣ **CI** $(1 - \alpha)$: $\hat{X}_{n+h|n} \pm z_{\alpha/2} \cdot \sqrt{\text{MSFE}(h)}$, where $z_{\alpha/2} = 1.96$ for 95%

Properties

- ▣ Intervals widen as the horizon increases
 - ▶ Converge to the unconditional interval: $\mu \pm z_{\alpha/2} \sigma_X$
- ▣ Width depends on model parameters
 - ▶ Larger AR coefficients \succ wider intervals
- ▣ **Python**: `model.get_forecast(h).conf_int()`

Forecast Evaluation

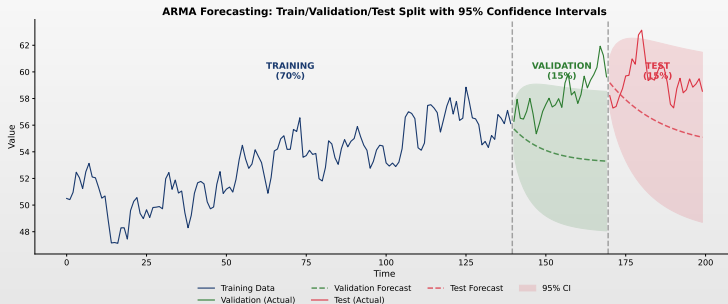
Out-of-Sample Testing

- Split data: training + test
- Generate forecasts on test
- Compare with actual values
- **Rolling window**: re-estimate as new data arrives

Error Metrics

- **MAE** = $\frac{1}{n} \sum |e_t|$
 - ▶ Robust to outliers
- **RMSE** = $\sqrt{\frac{1}{n} \sum e_t^2}$
 - ▶ Penalizes large errors
- **MAPE** = $\frac{100}{n} \sum \left| \frac{e_t}{X_t} \right|$
 - ▶ Percentage-based, interpretable

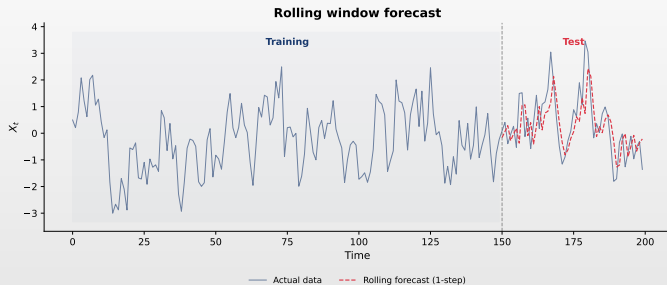
Train/Validation/Test Forecast Example



Best Practice

- Always evaluate forecasts on data not used for estimation (train/validation/test split)

Rolling Window Forecast

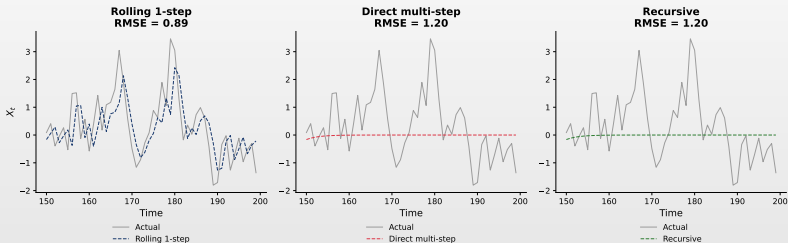


Rolling Forecast Methodology

- **Fixed window** (last w obs.) vs **expanding** (all data); generate 1-step forecast, repeat

Rolling vs Multi-Step Forecast

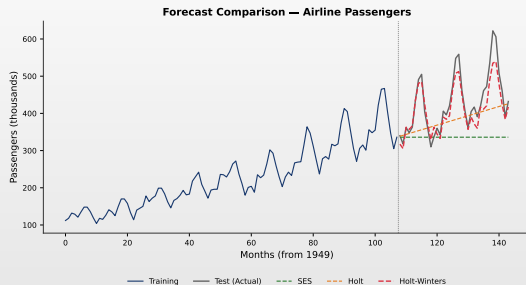
Comparison: Rolling vs Multi-step vs Recursive



Key Differences

- Rolling 1-step (accurate); Multi-step direct (separate model/horizon); Recursive (error accumulation)

Real Data Application: Forecast Comparison



Practical Considerations

- Real data: non-stationarity, structural breaks; compare models; use rolling window validation

Python Implementation: Estimating ARMA Models

Using statsmodels

```
from statsmodels.tsa.arima.model import ARIMA
model = ARIMA(data, order=(2, 0, 1)) # ARMA(2,1)
results = model.fit()
print(results.summary())
```

Forecasting

```
forecast = results.get_forecast(steps=10)
print(forecast.predicted_mean)
print(forecast.conf_int())
```

Note

ARIMA with $d = 0$ is equivalent to ARMA

Python: Model Selection with pmdarima

Automatic ARIMA Selection

```
import pmdarima as pm
model = pm.auto_arima(data,
    start_p=0, max_p=5, start_q=0, max_q=5,
    d=0, seasonal=False,
    information_criterion='aic', trace=True)
print(model.summary())
```

Result

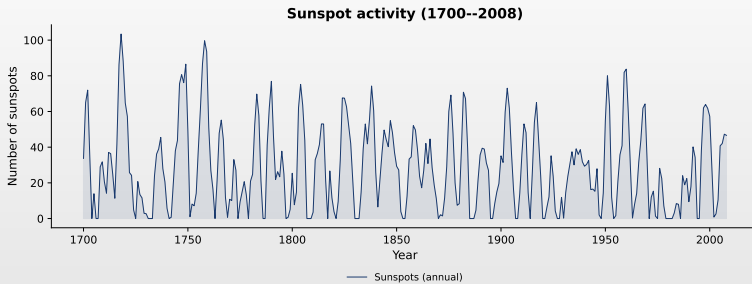
- Best model order and estimated parameters

Workflow Summary

Box-Jenkins Methodology Steps

- 1. **Data preparation:** Check for missing values, outliers; transform if necessary
- 2. **Stationarity check:** Visual inspection, formal tests (ADF, KPSS); difference if non-stationary
- 3. **Model identification:** ACF/PACF patterns; grid search with information criteria
- 4. **Estimation and validation:** Estimate model, check significance; residual analysis, Ljung-Box test
- 5. **Forecasting:** Point forecasts with confidence intervals; out-of-sample validation

Case Study: Sunspots

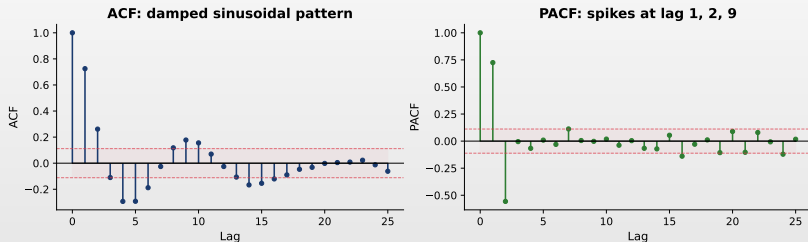


Data Description

- Annual sunspots (1700–2008): stationary series with ~ 11 -year cycles; Box-Jenkins methodology

Step 1: ACF/PACF Analysis

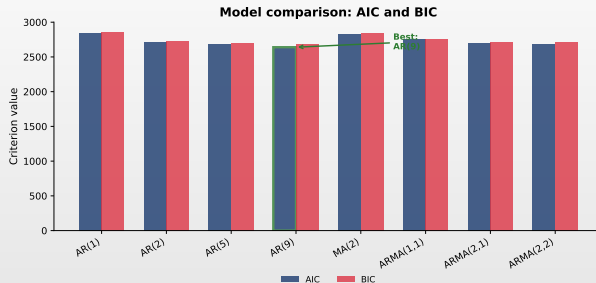
ACF/PACF analysis for sunspots



Identification

- Sinusoidal ACF (AR); PACF with spikes at lags 1, 2, 9 \succ AR(2) or AR(9); stationary series ($d = 0$)

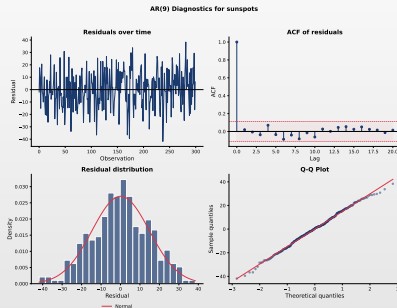
Step 2: Model Comparison



Model Selection

- Compare multiple candidate models using the AIC criterion
- The **AR(9)** model has the lowest AIC, capturing the 11-year solar cycle

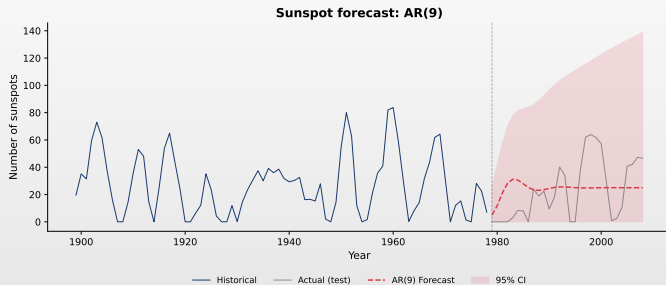
Step 3: Model Diagnostics



AR(9) Diagnostics

- Residuals: white noise, zero mean, constant variance, ACF without structure, \approx normal

Step 4: Forecasting



Results

- AR(9) captures the cyclical; 95% CI covers actual values; RMSE ≈ 30

Key Takeaways

Chapter Summary

- ▣ **AR(p)**: Depends on p past values; stationarity: roots outside the unit circle; PACF cuts off at lag p
- ▣ **MA(q)**: Depends on q past shocks; always stationary; ACF cuts off at lag q
- ▣ **ARMA(p,q)**: Combines AR and MA; both ACF and PACF decay
- ▣ **Box-Jenkins**: Identification \succ Estimation \succ Validation \succ Forecasting
- ▣ **Validation**: Residuals must be white noise
- ▣ **Forecasts**: Converge to the mean; uncertainty increases with the horizon

Next Chapter Preview

Chapter 3: ARIMA Models for Non-Stationary Data

- ▣ Non-stationarity: types, unit root tests (ADF, PP, KPSS)
- ▣ Differencing and the difference operator
- ▣ ARIMA(p,d,q): integrated models for non-stationary data
- ▣ The Auto-ARIMA algorithm: automatic model selection
- ▣ Case study: US GDP Forecasting

Reading

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

Question 1

Question

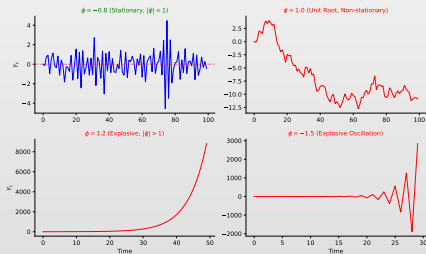
□ For which value of ϕ is the AR(1) process $X_t = c + \phi X_{t-1} + \varepsilon_t$ stationary?

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

Question 1: Answer

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

- ☐ AR(1) is stationary if and only if $|\phi| < 1$
- ☐ Only $|-0.8| = 0.8 < 1$



Question 2

Question

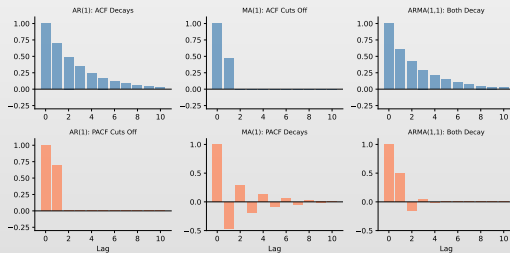
☐ You observe: ACF has a spike at lag 1, then cuts off. PACF decays gradually. What model?

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

Question 2: Answer

Correct Answer: (B) MA(1)

- ACF cuts off \succ MA process
- PACF decays \succ confirms MA(1)



Question 3

Question

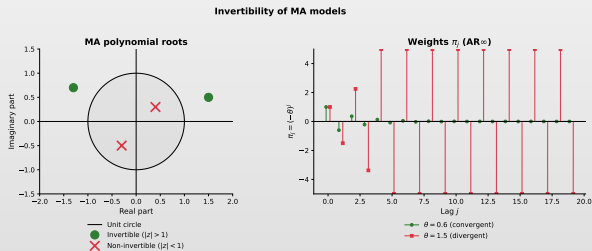
□ Is the MA(1) $X_t = \varepsilon_t + 1.5\varepsilon_{t-1}$ invertible?

- (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

Question 3: Answer

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

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



Question 4

Question

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

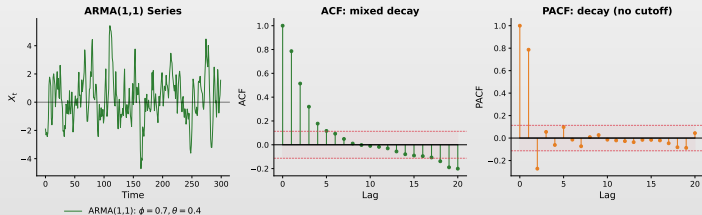
- (A) Pure AR model
- (B) Pure MA model
- (C) ARMA model
- (D) None of the above

Question 4: Answer

Correct Answer: (C) ARMA model

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

ARMA(1,1): neither ACF nor PACF cut off



Question 5

Question

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

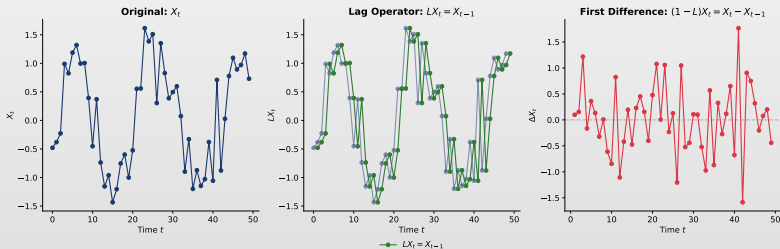
- (A) $X_t - X_{t-1}$
- (B) $X_t - 2X_{t-1} + X_{t-2}$
- (C) $X_t + X_{t-1} + X_{t-2}$
- (D) $X_t - X_{t-2}$

Question 5: Answer

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

☐ $(1 - L)^2 = 1 - 2L + L^2$

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



Question 6

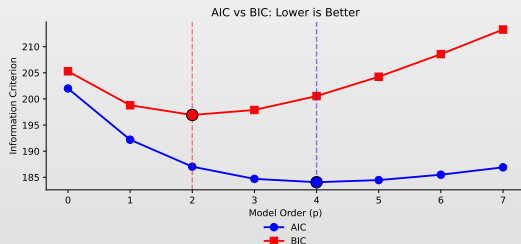
Question

- ☐ 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 the same number of parameters

Question 6: Answer

Correct Answer: (C) The model with lower BIC is preferred

- ☐ Lower BIC indicates a better balance between estimation quality and complexity
- ☐ BIC penalizes complexity *more* than AIC



Question 7

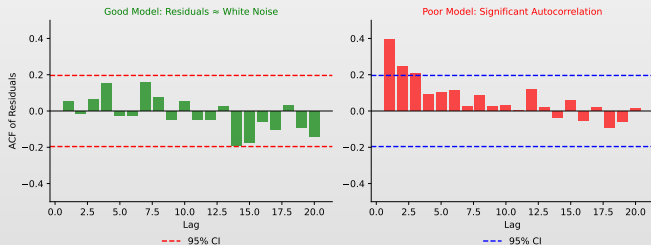
Question

- ☐ After estimating an ARMA model, you run the Ljung-Box test on residuals and obtain $p\text{-value} = 0.03$. What does this mean?
- (A) The model is adequate, residuals are white noise
 - (B) The model is inadequate, residuals have autocorrelation
 - (C) You need to increase the sample size
 - (D) The test is inconclusive

Question 7: Answer

Correct Answer: (B) The model is inadequate

- ☐ p-value < 0.05 rejects H_0 (white noise)
- ☐ Indicates remaining residual autocorrelation



Question 8

Question

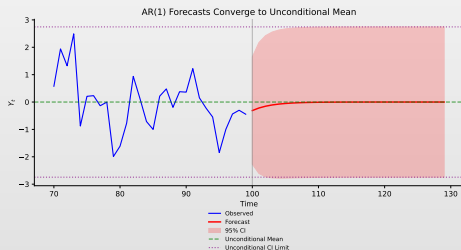
□ For a stationary AR(1) model, what happens to forecasts as the horizon $h \rightarrow \infty$?

- (A) Forecasts increase without bound
- (B) Forecasts oscillate indefinitely
- (C) Forecasts converge to the unconditional mean μ
- (D) Forecasts become more precise

Question 8: Answer

Correct Answer: (C) Forecasts converge to μ

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



Question 9

Question

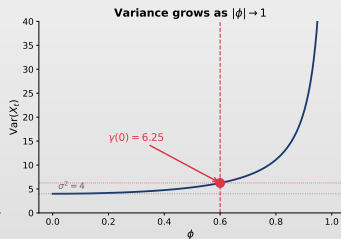
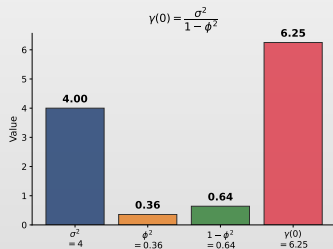
□ Consider an AR(1) process with $\phi = 0.6$ and $\sigma^2 = 4$. What is $\text{Var}(X_t)$?

- (A) 4.0
- (B) 5.56
- (C) 6.25
- (D) 10.0

Question 9: Answer

Correct Answer: (C) 6.25

- ☐ $\text{Var}(X_t) = \frac{\sigma^2}{1-\phi^2} = \frac{4}{1-0.36} = \frac{4}{0.64} = 6.25$
- ☐ The process variance exceeds σ^2 due to persistence



Question 10

Question

□ Consider an MA(1) process with $\theta = 0.5$. What is $\rho(1)$?

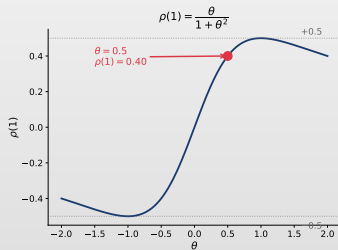
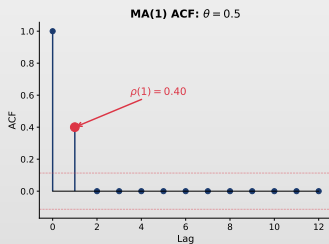
- (A) 0.50
- (B) 0.40
- (C) 0.25
- (D) 0.33

Question 10: Answer

Correct Answer: (B) 0.40

☐ $\rho(1) = \frac{\theta}{1+\theta^2} = \frac{0.5}{1+0.25} = \frac{0.5}{1.25} = 0.40$

☐ Note that $\rho(1) < \theta$ — the autocorrelation is **always** attenuated



Question 11

Question

□ Which statement about the ACF of an ARMA(1,1) process is **true**?

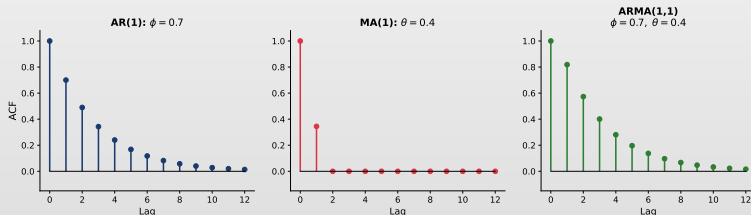
- (A) It cuts off after lag 1
- (B) Exponential decay starting from lag 1, with $\rho(1) \neq \phi$
- (C) It is zero for all lags
- (D) It exactly follows the pattern ϕ^h for all $h \geq 0$

Question 11: Answer

Correct Answer: (B) Exponential decay from lag 1, with $\rho(1) \neq \phi$

- $\rho(1) = \frac{(1+\phi\theta)(\phi+\theta)}{1+2\phi\theta+\theta^2} \neq \phi$ (the MA component modifies lag 1)
- For $h \geq 2$: $\rho(h) = \phi \rho(h-1)$ — exponential decay as in AR(1)

ACF Comparison: AR(1) vs MA(1) vs ARMA(1,1)



Question 12

Question

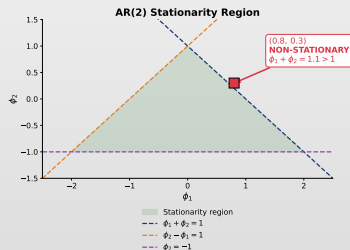
□ An AR(2) process has $\phi_1 = 0.8$ and $\phi_2 = 0.3$. Is it stationary?

- (A) Yes, it is stationary
- (B) No, because $\phi_1 + \phi_2 = 1.1 > 1$
- (C) Cannot be determined without data
- (D) Depends on the value of σ^2

Question 12: Answer

Correct Answer: (B) No, because $\phi_1 + \phi_2 = 1.1 > 1$

- Necessary conditions for AR(2) stationarity:
- $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$, $|\phi_2| < 1$
- Here $0.8 + 0.3 = 1.1 > 1 \Rightarrow$ the first condition is violated



Question 13

Question

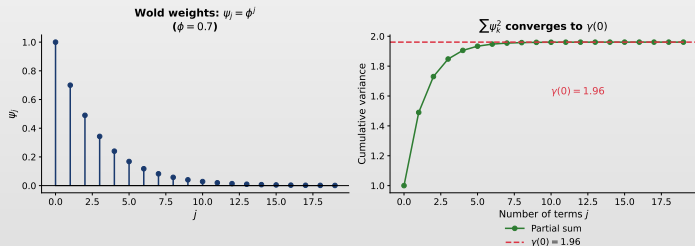
□ What does the Wold decomposition theorem guarantee?

- (A) Any time series is an AR process
- (B) Any stationary process can be written as $MA(\infty)$: $X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$
- (C) Any process has finite variance
- (D) ARMA models are always invertible

Question 13: Answer

Correct Answer: (B) Any stationary process = $MA(\infty)$

- Wold's theorem: $X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + D_t$, where D_t is the deterministic component
- This justifies ARMA models: they are parsimonious approximations of $MA(\infty)$



Question 14

Question

□ AR(1) with $\phi = 0.9$, $\sigma^2 = 1$. What happens to the CI width as $h \rightarrow \infty$?

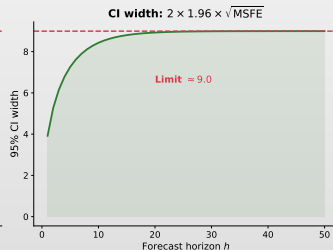
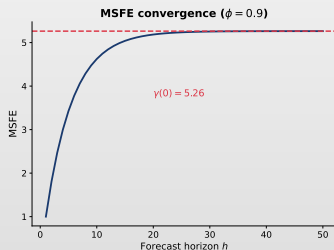
- (A) It remains constant
- (B) It decreases to zero
- (C) It grows toward $2 \times 1.96 \times \sqrt{1/(1 - 0.81)} \approx 9.0$
- (D) It grows to infinity

Question 14: Answer

Correct Answer: (C) Grows toward ≈ 9.0

□ $MSFE(\infty) = \frac{\sigma^2}{1-\phi^2} = \frac{1}{1-0.81} = \frac{1}{0.19} \approx 5.26$

□ $CI \text{ width} = 2 \times 1.96 \sqrt{5.26} \approx 2 \times 1.96 \times 2.29 \approx 9.0$



Data Sources and Software

Software Packages

- ▣ `statsmodels` > Statistical models for Python, including ARIMA
- ▣ `pmdarima` > Automatic ARIMA selection for Python
- ▣ `scipy` > Optimization and statistical functions
- ▣ `numpy`, `pandas` > Data manipulation
- ▣ `matplotlib` > Visualization

Data and Examples

- ▣ Simulated time series for illustrations
- ▣ Examples based on Hyndman & Athanasopoulos (2021)

Bibliography I

Fundamental ARMA Works

- Box, G.E.P., & Jenkins, G.M. (1970). *Time Series Analysis: Forecasting and Control*, Holden-Day.
- Akaike, H. (1974). A New Look at the Statistical Model Identification, *IEEE Transactions on Automatic Control*, 19(6), 716–723.
- Schwarz, G. (1978). Estimating the Dimension of a Model, *The Annals of Statistics*, 6(2), 461–464.

Diagnostics and Validation

- Ljung, G.M., & Box, G.E.P. (1978). On a Measure of Lack of Fit in Time Series Models, *Biometrika*, 65(2), 297–303.
- Brockwell, P.J., & Davis, R.A. (2016). *Introduction to Time Series and Forecasting*, 3rd ed., Springer.

Bibliography II

Textbooks and Additional References

- ▣ Hamilton, J.D. (1994). *Time Series Analysis*, Princeton University Press.
- ▣ Shumway, R.H., & Stoffer, D.S. (2017). *Time Series Analysis and Its Applications*, 4th ed., Springer.
- ▣ Hyndman, R.J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*, 3rd ed., OTexts.

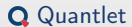
Online Resources and Code

- ▣ **Quantlet**: <https://quantlet.com> → Code repository for statistics
- ▣ **Quantinar**: <https://quantinar.com> → Learning platform for quantitative methods
- ▣ **GitHub TSA**: <https://github.com/QuantLet/TSA> → Python code for this course

Thank You!

Questions?

Course materials available at: <https://danpele.github.io/Time-Series-Analysis/>



Quantlet



Quantinar