Time Series Analysis and Forecasting

Chapter 1: Introduction to Time Series

Seminar

Seminar Outline

Today's Activities:

1. **Quick Review** – Key concepts recap
2. **Multiple Choice Quizzes** – Test your understanding
3. **True/False Questions** – Conceptual checks
4. **Calculation Exercises** – Hands-on practice
5. **Python Exercises** – Coding practice
6. **Discussion Questions** – Critical thinking

Key Formulas to Remember

Decomposition:

- Additive: $X_t = T_t + S_t + \varepsilon_t$
- Multiplicative: $X_t = T_t \times S_t \times \varepsilon_t$

Exponential Smoothing:

- SES: $\hat{X}_{t+1} = \alpha X_t + (1-\alpha)\hat{X}_t$
- Holt: adds trend b_t
- HW: adds seasonal S_t

Stationarity:

- $\mathbb{E}[X_t] = \mu$ (constant)
- $\text{Var}(X_t) = \sigma^2$ (constant)
- $\text{Cov}(X_t, X_{t+h}) = \gamma(h)$

Random Walk:

- $X_t = X_{t-1} + \varepsilon_t$
- $\text{Var}(X_t) = t\sigma^2$ (grows!)

Key Concepts Summary

Concept	Key Point	When to Use
Additive decomp.	Constant seasonal amplitude	Stable variance
Multiplicative decomp.	Seasonal grows with level	Increasing variance
SES	Level only (α)	No trend, no seasonality
Holt	Level + Trend (α, β)	Trend, no seasonality
Holt-Winters	Level + Trend + Seasonal	Trend and seasonality
ADF Test	H_0: unit root	Test for non-stationarity
KPSS Test	H_0: stationary	Confirm stationarity
Differencing	Remove stochastic trend	Random walk, unit root
Regression	Remove deterministic trend	Linear/polynomial trend

Quiz 1: Time Series Basics

Question

Which of the following is NOT a characteristic of time series data?

A. Observations are ordered in time
B. Consecutive observations are typically correlated
C. Observations are independent and identically distributed
D. The data has a natural temporal ordering

Answer on next slide...

Quiz 1: Answer

Answer: C – Observations are independent and identically distributed

Question: Which is NOT a characteristic of time series data?

- A. Observations are ordered in time ×
- B. Consecutive observations are typically correlated ×
- C. **Observations are independent and identically distributed ✓**
- D. The data has a natural temporal ordering ×

Time series observations are typically **dependent** (autocorrelated), not independent. The assumption of i.i.d. observations is fundamental to cross-sectional analysis but violated in time series. This temporal dependence is what makes time series analysis unique and requires specialized methods.

Quiz 2: Decomposition

Question

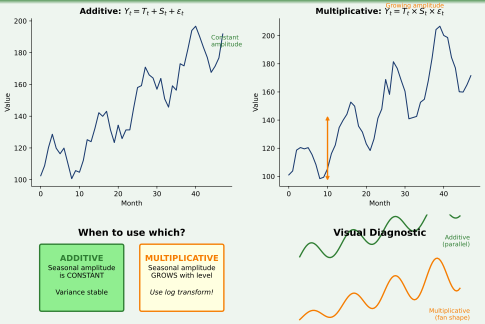
When should you use multiplicative decomposition instead of additive?

A. When the seasonal pattern has constant amplitude
B. When the variance of the series is stable over time
C. When the seasonal fluctuations grow proportionally with the level
D. When the time series has no trend component

Answer on next slide...

Quiz 2: Answer

Answer: C – When seasonal fluctuations grow proportionally with the level



Multiplicative: $X_t = T_t \times S_t \times \varepsilon_t$, seasonal amplitude **scales with level** (fan-shaped pattern)

Quiz 3: Exponential Smoothing

Question

In Simple Exponential Smoothing with $\alpha = 0.9$, what happens?

- A. Forecasts are very smooth and stable
- B. Recent observations have very little weight
- C. Forecasts react quickly to recent changes
- D. The forecast is essentially a long-term average

Answer on next slide...

Quiz 3: Answer

Answer: C – Forecasts react quickly to recent changes

With $\alpha = 0.9$: $\hat{X}_{t+1} = 0.9X_t + 0.1\hat{X}_t$
This means 90% weight on the most recent observation! High α values make forecasts very responsive to new data. Low α (e.g., 0.1) produces smoother, more stable forecasts that average over more history.

Quiz 4: Stationarity

Question

A random walk process $X_t = X_{t-1} + \varepsilon_t$ is:

- A. Strictly stationary
- B. Weakly stationary
- C. Non-stationary because variance grows with time
- D. Stationary after adding a constant

Answer on next slide...

Quiz 4: Answer

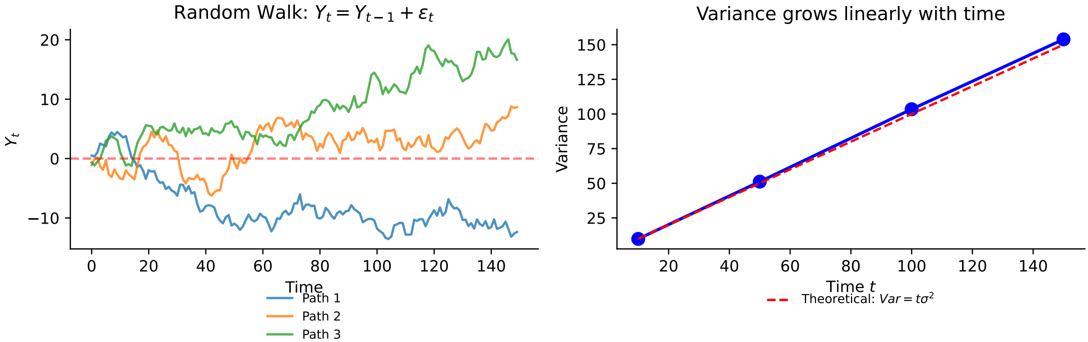
Answer: C – Non-stationary because variance grows with time

For random walk: $X_t = \sum_{i=1}^{t} \varepsilon_i$

- $\mathbb{E}[X_t] = 0$ (constant mean – OK)
- $\text{Var}(X_t) = t\sigma^2$ (variance depends on t – NOT OK!)

Since variance is not constant, the process violates the stationarity condition. Solution: **differencing** gives $\Delta X_t = \varepsilon_t$ which IS stationary.

Visual: Random Walk vs Stationary



Random walk paths wander unpredictably; variance grows linearly with time $\Rightarrow$ non-stationary.

Quiz 5: Unit Root Tests

Question

You run ADF and KPSS tests. ADF fails to reject H_0, and KPSS rejects H_0. What do you conclude?

- A. The series is stationary
- B. The series has a unit root (non-stationary)
- C. The results are inconclusive
- D. You need to run more tests

Answer on next slide...

Quiz 5: Answer

Answer: B – The series has a unit root (non-stationary)

- ADF: H_0 = unit root. Fail to reject $\Rightarrow$ evidence FOR unit root
- KPSS: H_0 = stationary. Reject $\Rightarrow$ evidence AGAINST stationarity

Both tests agree: the series is **non-stationary**. You should difference the series before modeling with ARMA.

Quiz 6: Forecast Evaluation

Question

Which metric is most appropriate for comparing forecast accuracy across different time series with different scales?

- A. Mean Absolute Error (MAE)
- B. Root Mean Squared Error (RMSE)
- C. Mean Absolute Percentage Error (MAPE)
- D. Mean Squared Error (MSE)

Answer on next slide...

Quiz 6: Answer

Answer: C – Mean Absolute Percentage Error (MAPE)

$\text{MAPE} = \frac{100}{n} \sum \left| \frac{e_t}{X_t} \right|$ expresses errors as **percentages**.

- MAE, RMSE, MSE are **scale-dependent** (units of X_t)
- MAPE is **scale-independent** (always in %)
- Caveat: MAPE fails when X_t is near zero

Quiz 7: Trend Types

Question

A deterministic trend can be removed by:

A. Differencing
B. Regression on time
C. Seasonal adjustment
D. Moving average smoothing

Answer on next slide...

Quiz 7: Answer

Answer: B – Regression on time

Deterministic trend: $Y_t = \alpha + \beta t + \varepsilon_t$ where β is fixed.
Removal method: Regress Y_t on t, then analyze residuals $\hat{\varepsilon}_t = Y_t - \hat{\alpha} - \hat{\beta} t$
Why not differencing? Differencing a deterministic trend gives: $\Delta Y_t = \beta + \Delta\varepsilon_t$, which removes the trend but leaves a constant. For *stochastic* trends (unit roots), differencing is correct.

Quiz 8: ACF Interpretation

Question

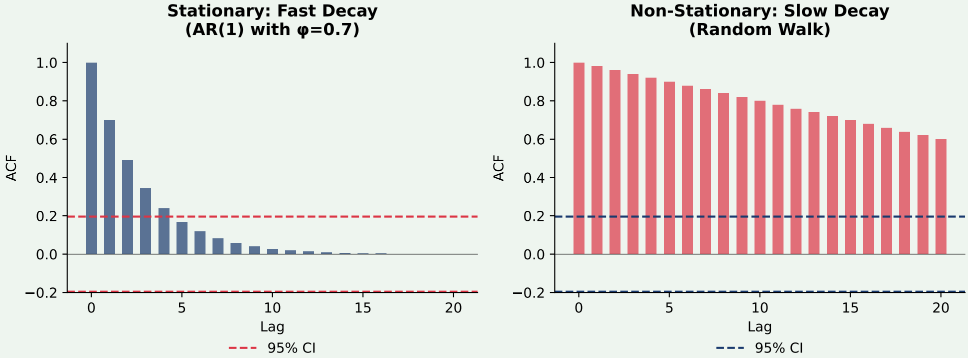
If the ACF of a time series decays very slowly (remains significant for many lags), this suggests:

A. The series is white noise
B. The series is likely non-stationary
C. The series has no autocorrelation
D. The series is perfectly predictable

Answer on next slide...

Quiz 8: Answer

Answer: B – The series is likely non-stationary



Stationary: ACF decays quickly ($\rho_k = \phi^k \to 0$) **Non-stationary**: ACF stays near 1 $\Rightarrow$ differencing needed

Quiz 9: Holt's Method

Question

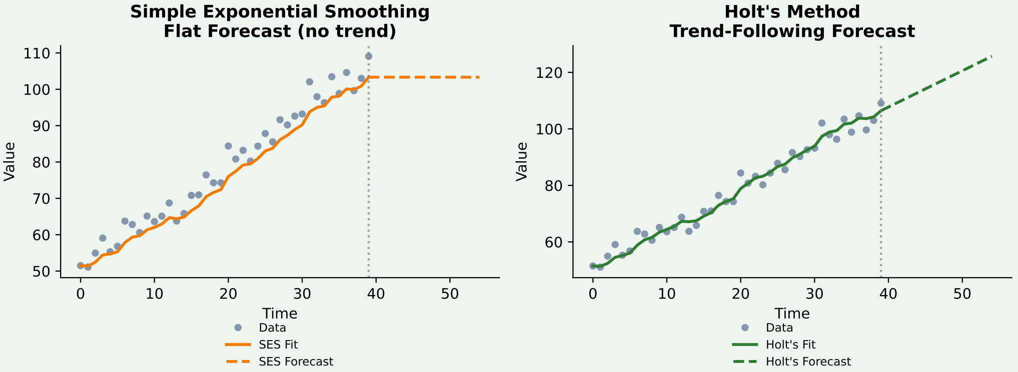
Holt's exponential smoothing differs from SES by adding:

A. A seasonal component
B. A trend component
C. A cyclical component
D. An irregular component

Answer on next slide...

Quiz 9: Answer

Answer: B – A trend component



Holt: $L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + b_{t-1})$; $b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}$ **Forecast**: $\hat{Y}_{t+h} = L_t + h \cdot b_t$

Quiz 10: White Noise

Question

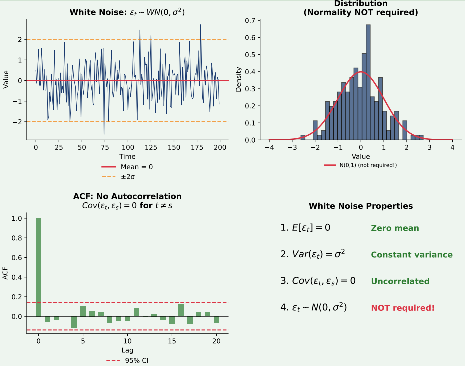
Which property is NOT required for a process to be white noise?

- A. $\mathbb{E}[\varepsilon_t] = 0$
- B. $\text{Var}(\varepsilon_t) = \sigma^2$ (constant)
- C. $\text{Cov}(\varepsilon_t, \varepsilon_s) = 0$ for $t \neq s$
- D. $\varepsilon_t \sim N(0, \sigma^2)$

Answer on next slide...

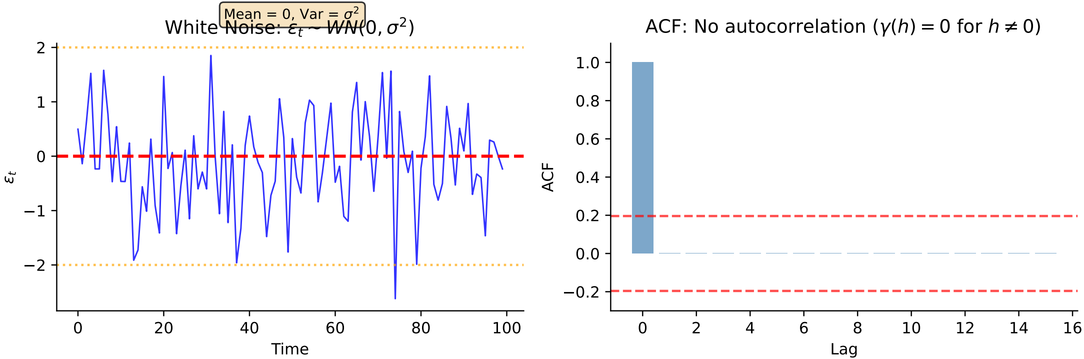
Quiz 10: Answer

Answer: D – Normality is NOT required



White noise: Zero mean, constant variance, uncorrelated. **Gaussian WN**: Adds normality $\Rightarrow$ also independent (not just uncorrelated)

Visual: White Noise Properties



Left: white noise fluctuates around zero. Right: ACF shows no autocorrelation (all values near zero after lag 0).

Quiz 11: Forecast Horizon

Question

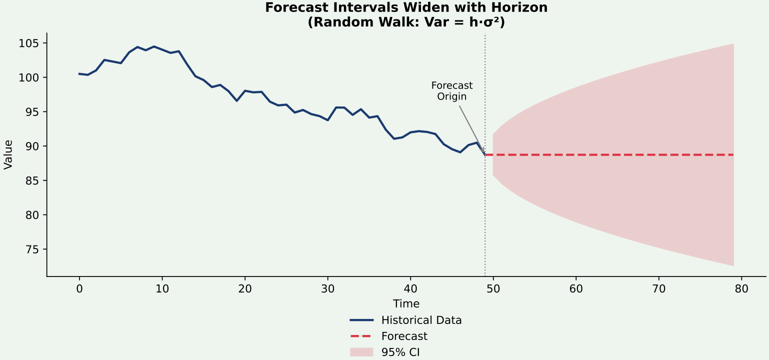
As forecast horizon h increases, what typically happens to forecast intervals?

A. They become narrower
B. They stay the same width
C. They become wider
D. They disappear

Answer on next slide...

Quiz 11: Answer

Answer: C – They become wider



Random walk: Var $= h\sigma^2$ (grows linearly) **95% CI**: $\hat{Y}_{t+h} \pm 1.96\sqrt{h}\sigma$ (widens with $\sqrt{h}$)

Quiz 12: Seasonality Detection

Question

The ACF shows significant spikes at lags 12, 24, and 36 for monthly data. This suggests:

A. No seasonality
B. Annual seasonality
C. Weekly seasonality
D. Random noise

Answer on next slide...

Quiz 12: Answer

Answer: B – Annual seasonality

Pattern recognition:

- Lag 12: correlation with same month last year
- Lag 24: correlation with same month two years ago
- Lag 36: correlation with same month three years ago

Seasonal period: $s = 12$ (monthly data with annual cycle)
Common patterns: Retail sales (December peaks), energy consumption (summer/winter), tourism data

Quiz 13: Cross-Validation in Time Series

Question

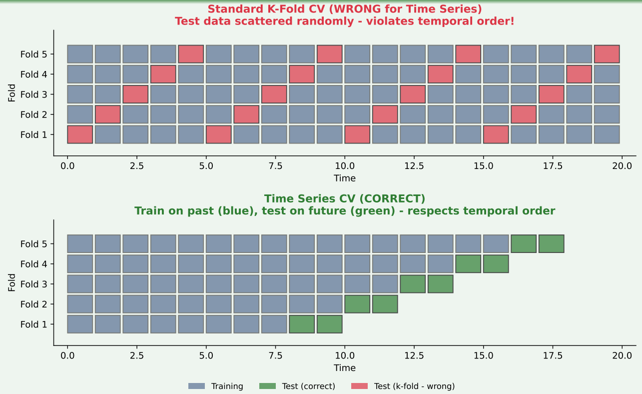
Why can't we use standard k-fold cross-validation for time series?

- A. Time series data is too small
- B. It would violate temporal ordering (future predicting past)
- C. Cross-validation is always invalid
- D. Time series don't need validation

Answer on next slide...

Quiz 13: Answer

Answer: B – It would violate temporal ordering



Rule: Never use future data to predict past! Use rolling/expanding window CV.

Quiz 14: MAPE Limitation

Question

MAPE (Mean Absolute Percentage Error) should NOT be used when:

- A. Comparing models on the same dataset
- B. The actual values can be zero or near zero
- C. Forecasting stock prices
- D. The data has a trend

Answer on next slide...

Quiz 14: Answer

Answer: B – When actual values can be zero or near zero

MAPE formula: $\text{MAPE} = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right|$

Problem: When $Y_t \approx 0$, division causes MAPE $\to \infty$

Alternatives:

- **SMAPE**: $\frac{200\%}{n}\sum\frac{|Y_t - \hat{Y}_t|}{|Y_t| + |\hat{Y}_t|}$ (bounded 0–200%)
- **MASE**: $\frac{1}{n}\sum\frac{|e_t|}{\frac{1}{n-1}\sum|Y_t - Y_{t-1}|}$ (scale-free)

True or False? (Set 1)

Question

Mark each statement as True (T) or False (F):

1. A time series with constant mean is always stationary. ________
2. The variance of a random walk increases linearly with time. ________
3. SES forecasts are always flat (constant for all horizons). ________
4. ADF and KPSS tests have the same null hypothesis. ________
5. Lower RMSE always means better forecasts. ________
6. Autocorrelation at lag 0 is always equal to 1. ________

Answer on next slide...

True or False: Answers (Set 1)

Answers

1. A time series with constant mean is always stationary. **FALSE**
   Also need constant variance and covariance depending only on lag.

2. The variance of a random walk increases linearly with time. **TRUE**
   $\text{Var}(X_t) = t\sigma^2$ for random walk starting at X_0.

3. SES forecasts are always flat (constant for all horizons). **TRUE**
   SES has no trend component, so $\hat{X}_{t+h} = L_t$ for all h.

4. ADF and KPSS tests have the same null hypothesis. **FALSE**
   ADF: H_0 = unit root. KPSS: H_0 = stationary. Opposite nulls!

5. Lower RMSE always means better forecasts. **FALSE**
   Depends on context. RMSE is scale-dependent; may overfit to outliers.

6. Autocorrelation at lag 0 is always equal to 1. **TRUE**
   $\rho(0) = \gamma(0)/\gamma(0) = 1$ by definition.

True or False? (Set 2)

Question

Mark each statement as True (T) or False (F):

1. The ACF of a stationary AR(1) process decays exponentially. _______
2. White noise is always normally distributed. _______
3. Differencing can make a non-stationary series stationary. _______
4. The PACF of a MA(1) process cuts off after lag 1. _______
5. You should always use the test set for hyperparameter tuning. _______
6. Holt-Winters is appropriate for data with no seasonality. _______

Answer on next slide...

True or False: Answers (Set 2)

Answers

1. The ACF of a stationary AR(1) decays exponentially. **TRUE**
   For AR(1): $\rho(h) = \phi^h$, which decays exponentially.
2. White noise is always normally distributed. **FALSE**
   White noise only requires zero mean, constant variance, no autocorrelation. Gaussian white noise is a special case.
3. Differencing can make a non-stationary series stationary. **TRUE**
   Differencing removes stochastic trends (unit roots).
4. The PACF of a MA(1) cuts off after lag 1. **FALSE**
   It's the ACF that cuts off for MA. PACF decays for MA processes.
5. You should always use the test set for hyperparameter tuning. **FALSE**
   Use validation set for tuning. Test set is for final evaluation only!
6. Holt-Winters is appropriate for data with no seasonality. **FALSE**
   Use Holt's method (no seasonal component) or SES for non-seasonal data.

Exercise 1: Simple Exponential Smoothing

Problem: Given the following data and $\alpha = 0.3$:

t	1	2	3	4	5
X_t	10	12	11	14	13

Starting with $\hat{X}_1 = X_1 = 10$, calculate:

a) The forecasts $\hat{X}_2, \hat{X}_3, \hat{X}_4, \hat{X}_5$
b) The forecast for $t = 6$: $\hat{X}_6$
c) The forecast errors $e_t = X_t - \hat{X}_t$ for $t = 2, 3, 4, 5$
d) The MAE and RMSE

Formula: $\hat{X}_{t+1} = \alpha X_t + (1 - \alpha)\hat{X}_t$

Exercise 1: Solution

Using $\hat{X}_{t+1} = 0.3X_t + 0.7\hat{X}_t$:

t	1	2	3	4	5	6
X_t	10	12	11	14	13	?
$\hat{X}_t$	10	10	10.6	10.72	11.70	**12.09**
e_t	–	2	0.4	3.28	1.30	–

Calculations:

- $\hat{X}_2 = 0.3(10) + 0.7(10) = 10$
- $\hat{X}_3 = 0.3(12) + 0.7(10) = 10.6$
- $\hat{X}_4 = 0.3(11) + 0.7(10.6) = 10.72$
- $\hat{X}_5 = 0.3(14) + 0.7(10.72) = 11.70$
- $\hat{X}_6 = 0.3(13) + 0.7(11.70) = \mathbf{12.09}$

$\textbf{MAE} = \frac{|2|+|0.4|+|3.28|+|1.30|}{4} = 1.745 \quad \textbf{RMSE} = \sqrt{\frac{4+0.16+10.76+1.69}{4}} = 2.04$

Exercise 2: Autocovariance

Problem: For a stationary process with:

- $\mathbb{E}[X_t] = 5$
- $\gamma(0) = 4$ (variance)
- $\gamma(1) = 2$
- $\gamma(2) = 1$

Calculate:

a) The autocorrelation function $\rho(0), \rho(1), \rho(2)$

b) $\text{Cov}(X_t, X_{t-1})$

c) $\text{Corr}(X_5, X_7)$

d) If $X_t = 6$, what is $\mathbb{E}[X_{t+1}|X_t = 6]$ assuming AR(1)?

Exercise 2: Solution

a) Autocorrelations:

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

- $\rho(0) = \gamma(0)/\gamma(0) = 1$
- $\rho(1) = \gamma(1)/\gamma(0) = 2/4 = 0.5$
- $\rho(2) = \gamma(2)/\gamma(0) = 1/4 = 0.25$

b) $\text{Cov}(X_t, X_{t-1}) = \gamma(1) = 2$ (by stationarity, lag 1 covariance)

c) $\text{Corr}(X_5, X_7) = \rho(|7-5|) = \rho(2) = 0.25$

d) For AR(1) with $\phi = \rho(1) = 0.5$:

$$\mathbb{E}[X_{t+1}|X_t] = \mu + \phi(X_t - \mu) = 5 + 0.5(6-5) = 5.5$$

Exercise 3: Random Walk Properties

Problem: Consider a random walk $X_t = X_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim WN(0, 4)$ and $X_0 = 100$.

Calculate:

a) $\mathbb{E}[X_{10}]$
b) $\text{Var}(X_{10})$
c) $\text{Cov}(X_5, X_{10})$
d) The 95% confidence interval for X_{100}
e) After observing $X_5 = 108$, what is your best forecast for X_6?

Exercise 3: Solution

Random walk: $X_t = X_0 + \sum_{i=1}^{t} \varepsilon_i$ with $\sigma^2 = 4$

a) $\mathbb{E}[X_{10}] = X_0 = 100$ (mean stays at starting value)

b) $\text{Var}(X_{10}) = 10 \times \sigma^2 = 10 \times 4 = 40$

c) $\text{Cov}(X_5, X_{10}) = \min(5, 10) \times \sigma^2 = 5 \times 4 = 20$

d) For X_{100}:

- $\mathbb{E}[X_{100}] = 100$, $\text{Var}(X_{100}) = 400$, $SD = 20$
- 95% CI: $100 \pm 1.96 \times 20 = [60.8, 139.2]$

e) Best forecast: $\hat{X}_6 = X_5 = 108$
(Random walk: best forecast is the last observed value)

Python Exercise 1: Load and Plot

Task: Load S&P 500 data and create a basic time series plot.

Starter Code

```
import yfinance as yf
import matplotlib.pyplot as plt

# Download S&P 500 data
sp500 = yf.download('^GSPC', start='2020-01-01', end='2025-01-01')

# TODO: Plot the closing prices
# TODO: Add title, labels, and grid
# TODO: Calculate and print basic statistics
```

Questions:

1. What is the mean and standard deviation of returns?
2. Does the series appear stationary? Why or why not?

Python Exercise 2: Decomposition

Task: Perform STL decomposition on airline passengers data.

Starter Code

```
from statsmodels.tsa.seasonal import STL
import pandas as pd

# Load airline passengers
url = 'https://raw.githubusercontent.com/..../airline.csv'
airline = pd.read_csv(url, parse_dates=['Month'],
                      index_col='Month')

# TODO: Apply STL decomposition with period=12
# TODO: Plot all components
# TODO: What percentage of variance is explained by trend?
```

Hint: Use `STL(data, period=12).fit()`

Python Exercise 3: Exponential Smoothing

Task: Compare SES, Holt, and Holt-Winters on real data.

Starter Code

```
from statsmodels.tsa.holtwinters import (SimpleExpSmoothing,
    ExponentialSmoothing)

# Split data: 80% train, 20% test
train = airline[:'1958']
test = airline['1959':]

# TODO: Fit SES, Holt, and Holt-Winters
# TODO: Generate forecasts for test period
# TODO: Calculate RMSE for each method
# TODO: Which method performs best? Why?
```

Python Exercise 4: Stationarity Testing

Task: Test for stationarity using ADF and KPSS tests.

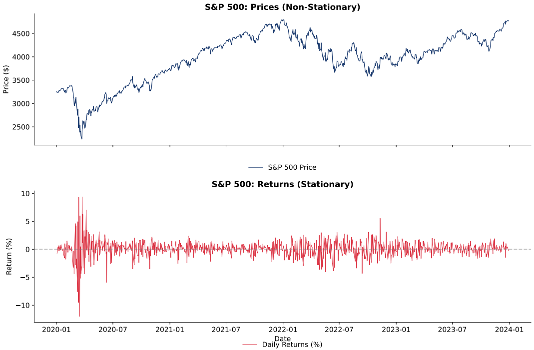
Starter Code

```
from statsmodels.tsa.stattools import adfuller, kpss

# Test S&P 500 prices
prices = sp500['Close']
returns = prices.pct_change().dropna()

# TODO: Run ADF test on prices and returns
# TODO: Run KPSS test on prices and returns
# TODO: Interpret the results

# ADF: adfuller(series)
# KPSS: kpss(series, regression='c')
```

Questions:

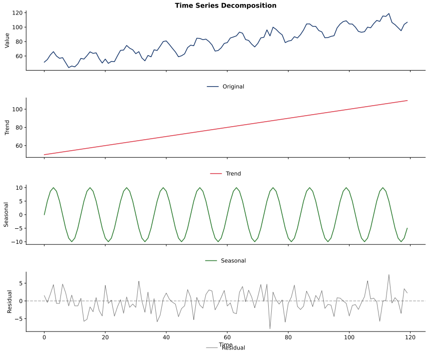
1. Are prices stationary? Are returns stationary?
2. Do ADF and KPSS agree?

Case Study: S&P 500 Index



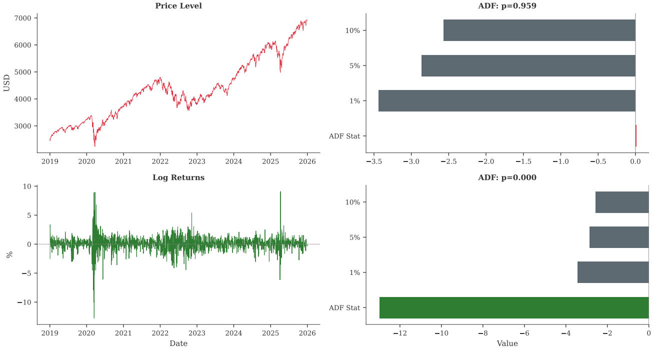
- **Top**: S&P 500 price level – clear upward trend (non-stationary)
- **Bottom**: Daily returns $r_t = \log(P_t/P_{t-1})$ – stationary
- Returns fluctuate around zero mean with no trend
- Volatility clustering visible – periods of high/low volatility

Time Series Decomposition: Real Example



- **Trend**: Long-term direction of the series
- **Seasonal**: Regular periodic patterns
- **Residual**: What remains after removing trend and seasonality
- Decomposition helps understand data structure before modeling

Stationarity Testing: ADF Results



- ADF test compares test statistic to critical values
- If test statistic $<$ critical value $\Rightarrow$ reject unit root (series is stationary)
- Prices: ADF statistic $> -2.86 \Rightarrow$ non-stationary
- Returns: ADF statistic $< -2.86 \Rightarrow$ stationary

Stationarity Comparison: Prices vs Returns

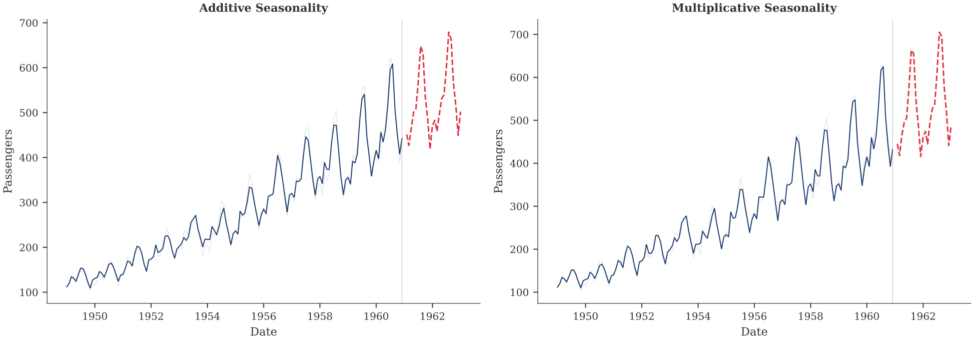
ADF Test Results

Series	ADF Statistic	p-value	Conclusion
S&P 500 Prices	-0.82	0.812	Non-stationary
S&P 500 Returns	-45.3	< 0.001	Stationary

Key Insight

Financial prices are typically $I(1)$ – integrated of order 1.
Taking first differences (returns) achieves stationarity.
This is why we model **returns**, not prices!

Exponential Smoothing Forecast



- Holt-Winters method for data with trend and seasonality
- Smoothing parameters α, β, γ control adaptiveness
- Forecasts capture both trend continuation and seasonal pattern
- Simple yet effective for many business applications

Discussion Question 1

Scenario

You are analyzing monthly sales data for a retail company. The data shows clear seasonality (high sales in December) and an upward trend. The seasonal peaks have been getting larger over time.

Discuss:

1. Should you use additive or multiplicative decomposition? Why?
2. Which exponential smoothing method would you recommend?
3. How would you evaluate your forecast model?
4. What could go wrong if you used the wrong decomposition?

Discussion Question 2

Scenario

A colleague claims: "I ran the ADF test on my stock price data and got a p-value of 0.65, so my data is stationary and I can fit an ARMA model directly."

Discuss:

1. What is wrong with this interpretation?
2. What do the ADF hypotheses actually test?
3. What should the colleague do before fitting an ARMA model?
4. How could the KPSS test help clarify the situation?

Discussion Question 3

Scenario

You're building a forecasting model and achieve excellent results: MAPE of 2% on your dataset. Your manager is impressed and wants to deploy the model immediately.

Discuss:

1. What questions should you ask before deployment?
2. Did you use proper train/validation/test splits?
3. Could there be data leakage in your evaluation?
4. What additional diagnostics would you run?
5. How would you monitor the model in production?

Discussion Question 4

Scenario

You need to forecast daily electricity demand for the next week. The data shows: (1) strong daily patterns (peaks at 6pm), (2) weekly patterns (lower on weekends), and (3) annual patterns (higher in summer/winter).

Discuss:

1. How would you handle multiple seasonal patterns?
2. Would Holt-Winters work here? Why or why not?
3. What's the advantage of Fourier terms in this case?
4. How would you set up your train/validation/test split?

Key Takeaways from Today

1. **Time series are dependent** – not i.i.d. like cross-sectional data
2. **Choose decomposition wisely** – multiplicative when seasonal amplitude grows
3. **Understand smoothing parameters** – high α = reactive, low α = smooth
4. **Test for stationarity** – use both ADF and KPSS together
5. **Proper evaluation** – never tune on test set!
6. **Random walk is non-stationary** – variance grows with time

Next Seminar

ARMA/ARIMA model identification, estimation, and forecasting

References

- Hyndman, R.J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*. 3rd ed., OTexts. `https://otexts.com/fpp3/`
- 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.
- Brockwell, P.J., & Davis, R.A. (2016). *Introduction to Time Series and Forecasting*. 3rd ed., Springer.

Data Sources and Software

Software Tools:

- `statsmodels` – Statistical models for Python
- `pandas` – Data manipulation and time series
- `matplotlib`, `seaborn` – Visualization
- `scipy` – Statistical functions

Data and Examples:

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

Thank You!

Questions?

Good luck with the exercises!