Time Series Analysis and Forecasting

Seminar 0: Fundamentals

Daniel Traian PELE

Academia de Studii Economice din București
IDA Institute Digital Assets
Blockchain Research Center
AI4EFin Artificial Intelligence for Energy Finance
Academia Română, Institutul de Prognoză Economică
MSCA Digital Finance

Seminar Structure

Seminar structure:

1. **Multiple Choice Quiz** – Knowledge check
2. **True/False** – Conceptual checks
3. **Calculation Exercises** – Applied practice
4. **AI-Assisted Exercises** – Human vs. AI analysis
5. **Summary** – Key takeaways

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 usually correlated
C. Observations are independent and identically distributed
D. 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 usually correlated ×

C. **Observations are independent and identically distributed** ✓

D. Data has a natural temporal ordering ×

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

TSA_ch0_definition

Quiz 2: Decomposition

Question

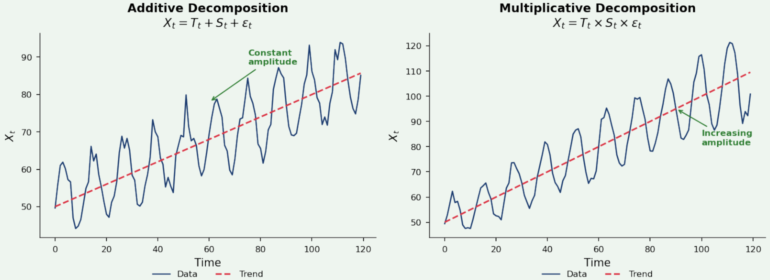
When should you use multiplicative decomposition instead of additive?

A. When the seasonal pattern has constant amplitude
B. When the variance of the series is stable over time
C. When seasonal fluctuations grow proportionally with the level
D. When the 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 the level**
- **Additive**: $X_t = T_t + S_t + \varepsilon_t$ — constant amplitude

TSA_ch0_decomposition

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 $\alpha = 0.9$: $\hat{X}_{t+1} = 0.9X_t + 0.1\hat{X}_t$

- **High** α (e.g. 0.9): 90% weight on the last observation
  - Forecasts very responsive to new data
- **Low** α (e.g. 0.1): smoother, more stable forecasts
  - Averages over more history

TSA_ch0_smoothing

Quiz 4: Moving Averages

Question

A centered moving average of order 5 (MA-5) uses which observations to estimate the trend at time t?

A. $X_t, X_{t+1}, X_{t+2}, X_{t+3}, X_{t+4}$

B. $X_{t-4}, X_{t-3}, X_{t-2}, X_{t-1}, X_t$

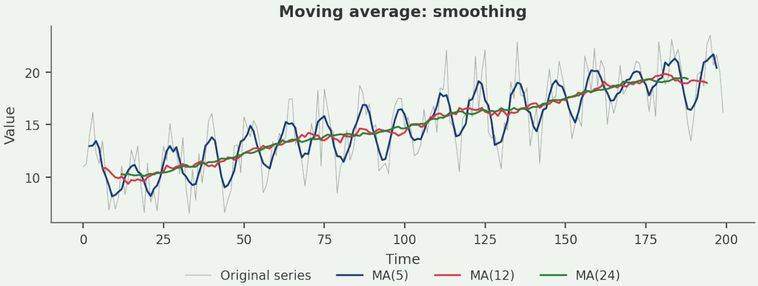
C. $X_{t-2}, X_{t-1}, X_t, X_{t+1}, X_{t+2}$

D. X_{t-1}, X_t, X_{t+1}

Answer on next slide...

Quiz 4: Answer

Answer: C – $X_{t-2}, X_{t-1}, X_t, X_{t+1}, X_{t+2}$



- **Centered MA**: uses $(k-1)/2$ observations on each side of t
- **MA-5**: 2 before + t + 2 after $\Rightarrow$ larger window = smoother

TSA_ch0_definition

Quiz 5: Forecast Evaluation

Question

Which metric is most appropriate for comparing forecast accuracy across 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 5: 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 becomes unstable when $X_t \approx 0$

TSA_ch0_forecast_eval

Quiz 6: Cross-Validation

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)

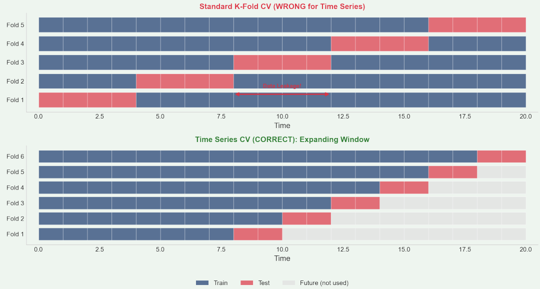
C. Cross-validation is always invalid

D. Time series doesn't need validation

Answer on next slide...

Quiz 6: Answer

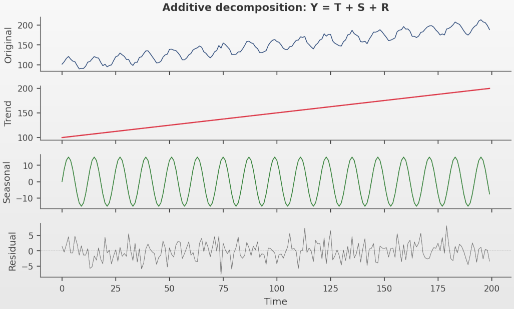
Answer: B – It would violate temporal ordering



Principle: future data cannot be used to predict the past! Rolling/expanding window CV is recommended.

TSA_ch0_forecast_eval

Visual: Time Series Decomposition



Decomposition Components

- **Trend**: long-term movement **Seasonality**: periodic pattern **Residual**: random noise

TSA_ch0_decomposition

True or False? — Questions

Statement	T/F?
1. SES forecasts are flat (constant for all horizons).	?
2. RMSE penalizes large errors more than MAE.	?
3. Multiplicative decomposition requires positive data.	?
4. A larger α means more smoothing.	?
5. The test set is used for hyperparameter tuning.	?
6. Seasonal naive uses the value from one season ago.	?
7. MAPE can be infinite if actual values are zero.	?

True or False? — Answers

Statement	T/F	Explanation
1. SES forecasts are flat (constant for all horizons).	T	**No trend**
2. RMSE penalizes large errors more than MAE.	T	**Squared errors**
3. Multiplicative decomposition requires positive data.	T	**Cannot $\times$ negative**
4. A larger α means more smoothing.	F	**Large α = less smooth**
5. The test set is used for hyperparameter tuning.	F	**Use validation!**
6. Seasonal naive uses the value from one season ago.	T	$\hat{x}_{t+h} = x_{t+h-m}$
7. MAPE can be infinite if actual values are zero.	T	**Division by zero**

Exercise 1: Simple Exponential Smoothing

Problem

- **Data**: $X = [10, 12, 11, 14, 13]$ with $\alpha = 0.3$, $\hat{X}_1 = 10$
- **Calculate**: a) Forecasts $\hat{X}_2$ through $\hat{X}_6$; b) MAE and RMSE
- **Formula**: $\hat{X}_{t+1} = \alpha X_t + (1 - \alpha)\hat{X}_t$

Solution

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**

- **MAE** $= 1.745$ **RMSE** $= 2.04$

Exercise 2: Error Metrics

Problem

- **Data**: $X = [100, 110, 105, 120]$, $\hat{X} = [95, 108, 110, 115]$
- **Calculate**: MAE, MSE, RMSE, MAPE

Solution

- **Errors**: $e = [5, 2, -5, 5]$
- **MAE** $= (|5| + |2| + |-5| + |5|)/4 = 4.25$
- **MSE** $= (25 + 4 + 25 + 25)/4 = 19.75$
- **RMSE** $= \sqrt{19.75} = 4.44$
- **MAPE** $= 25 \times (0.05 + 0.018 + 0.048 + 0.042) = 3.95\%$

Exercise 3: Seasonal Indices

Problem

- **Data**: Seasonal indices: $S = [0.85, 1.05, 0.90, 1.20]$, Trend Q4: $T = 1000$
- **Calculate**: a) Verify normalization. b) Q4 forecast. c) Deseasonalize $X_{Q4} = 1150$

Solution

- **a) Normalization**: $\sum S_i = 0.85 + 1.05 + 0.90 + 1.20 = 4.00$ ✓
- **b) Forecast**: $\hat{X}_{Q4} = 1000 \times 1.20 = \mathbf{1200}$
- **c) Deseasonalization**: $X_{deseasonalized} = 1150/1.20 = \mathbf{958.33}$ (below trend)

AI in Time Series Analysis

Why use AI tools in this course?

- AI assistants (Claude, ChatGPT, GitHub Copilot) can generate code and analysis
- Your job is to **evaluate, interpret, and critique** — skills AI cannot replace

Learning objectives:

- Write precise prompts for econometric tasks
- Identify errors in AI-generated statistical analysis
- Interpret output critically using course concepts
- Compare AI solutions with manual methodology

Important

- AI is a **tool**, not a replacement for understanding
- You must be able to explain *why* each step is correct

AI Exercise 1: Decomposition Audit

AI Output (monthly US retail sales, 2015–2023)

Prompt: "I have a file with monthly US retail sales from 2015 to 2023. What trends and patterns do you see?"

AI: "Applied additive STL, period=12. Trend captures growth from \$420B to \$580B. Seasonal component is stable across the sample."

The data contradicts the AI:

- Dec–Jan seasonal swing: ±\$18B in 2015, ±\$45B in 2023
- Residual std: \$3.2B (2015–2017) vs \$8.7B (2021–2023)

Tasks:

1. Compute $\frac{\text{seasonal amplitude}}{\text{level}}$ for 2015 vs 2023. If the ratio is approximately constant, what decomposition does this imply?
2. If $X_t = T_t \cdot S_t \cdot \varepsilon_t$, show that $\ln X_t = \underbrace{\ln T_t}_{\text{trend}} + \underbrace{\ln S_t}_{\text{seasonal}} + \underbrace{\ln \varepsilon_t}_{\text{residual}}$ is additive.
3. Apply STL to $\ln X_t$. How would you verify constant residual variance? (Hint: Breusch–Pagan or plot $|\hat{\varepsilon}_t|$ vs t.)
4. What are the consequences of using additive STL on multiplicative data for forecasting?

AI Exercise 2: Data Leakage Detection

AI Code (spot the errors)

```
df['ma12'] = df['sales'].rolling(12, center=True).mean()
df['detrended'] = df['sales'] - df['ma12']
model = ExponentialSmoothing(df['sales'], seasonal='add',
                             seasonal_periods=12).fit()
print(f"MAPE = {mape(df['sales'], model.fittedvalues):.1f}%")
# Output: MAPE = 1.1%
```

Find three critical errors:

1. `center=True` means $\text{MA}_{12}(t)$ uses $X_{t-5}, \ldots, X_{t+6}$ — **six future values** leak into the feature. What is the correct setting?
2. MAPE is computed on **in-sample fitted values**, not out-of-sample forecasts. If the true OOS MAPE $= 8.3\%$, what explains the $7\times$ gap?
3. Additive seasonal model on data with growing amplitude. What diagnostic plot reveals this? (Hint: plot $|S_t|$ vs level T_t.)

Fix: rewrite with `center=False`, split 80/20, compute MAPE on the test set, and compare with the seasonal naïve benchmark $\hat{X}_{t+h} = X_{t+h-12}$.

AI Exercise 3: Smoothing Parameter Forensics

AI Model Comparison (Industrial Production Index)

Model	α	β	γ	Train RMSE	Test RMSE
SES	0.98	—	—	1.2	4.8
Holt	0.95	0.89	—	0.9	6.3
Holt-Winters	0.99	0.85	0.01	0.3	9.1

AI conclusion: "Holt-Winters has the lowest RMSE — best model."

Tasks:

1. Show that SES with $\alpha \to 1$ gives $\hat{X}_{t+1|t} = X_t$ (naïve forecast). What does $\alpha = 0.98$ approximate?
2. Test RMSE *increases* with complexity: $4.8 \to 6.3 \to 9.1$. Name this phenomenon. Which RMSE column should guide model selection?
3. $\gamma = 0.01$: the seasonal indices barely update from their initial values. What happens if the seasonal pattern shifts (e.g., post-COVID)?
4. What model would you select, and why? Compute $\text{AIC} = T\ln(\text{SSE}/T) + 2k$ for each.

Summary: Chapter 0

Key Concepts

1. **Time series**: temporally ordered observations, with dependence (autocorrelation)
2. **Decomposition**: additive ($X_t = T_t + S_t + \varepsilon_t$) vs multiplicative ($X_t = T_t \times S_t \times \varepsilon_t$)
3. **Exponential smoothing**: SES, Holt, Holt-Winters — parameter α controls reactivity
4. **Forecast evaluation**: MAE, RMSE, MAPE — the choice depends on context
5. **Seasonality**: seasonal indices, forecasting and deseasonalization

Questions?

Bibliography I

Time Series Fundamentals

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

Financial Time Series

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

Bibliography II

Modern Approaches and Machine Learning

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

Online Resources and Code

- **Quantlet**: https://quantlet.com — Code repository for statistics
- **Quantinar**: https://quantinar.com — Quantitative methods learning platform
- **GitHub TSA**: https://github.com/QuantLet/TSA/tree/main/TSA_ch0 — Python code for this seminar

Thank You!

Questions?

Seminar materials are available at: `https://danpele.github.io/Time-Series-Analysis/`

Quantlet Quantinar