Time Series Analysis and Forecasting

Chapter 0: Fundamentals

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

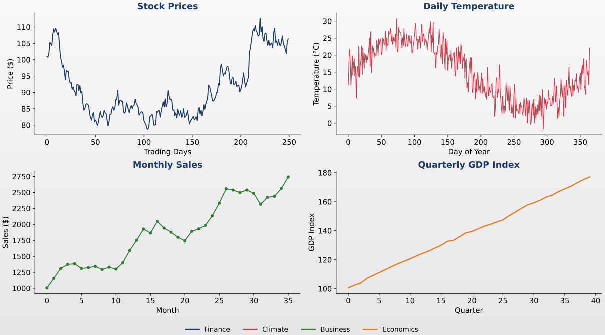
Learning Objectives

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

1. **Define** time series and distinguish from cross-sectional and panel data
2. **Decompose** time series into trend-cycle, seasonal, and residual components
3. **Apply** exponential smoothing methods (SES, Holt, Holt-Winters, ETS)
4. **Evaluate** forecasts using MAE, RMSE, MAPE, sMAPE
5. **Implement** train/validation/test splits and cross-validation
6. **Model** seasonality using dummy variables or Fourier terms
7. **Remove** trend and seasonality through appropriate methods
8. **Distinguish** between deterministic and stochastic trends

Chapter Outline

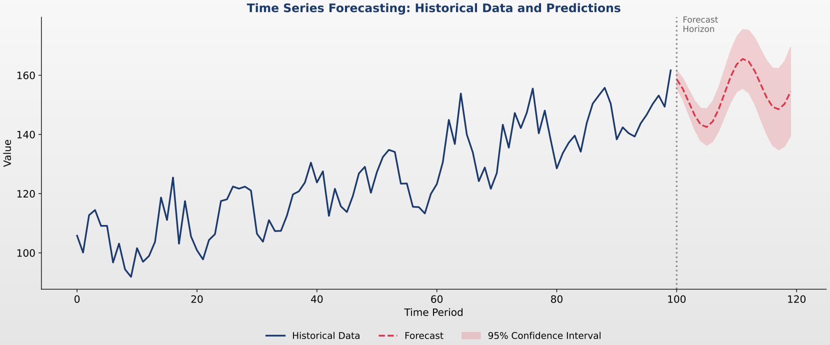
Time Series Are Everywhere



- **Finance**: Stock prices, exchange rates, trading volumes
- **Economics**: GDP, unemployment, inflation rates
- **Business**: Sales, website traffic, customer demand
- **Science**: Temperature, pollution levels, patient vitals

TSA_ch1_motivation

Why Study Time Series?

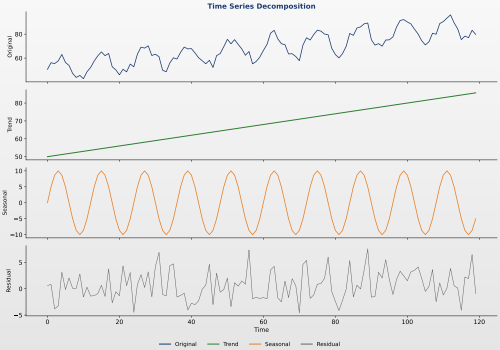


Key Goal: Forecasting

Use historical patterns to predict future values — critical for business planning, risk management, and policy decisions.

TSA_ch1_forecast

Understanding Time Series Structure



Decomposition

Every time series can be decomposed into interpretable components: trend-cycle, seasonality, and noise.

TSA_ch1_components

Definition of a Time Series

Definition 1 (Time Series)

A **time series** is a sequence of observations $\{X_t\}$ indexed by time:

$$\{X_t : t \in \mathcal{T}\}$$

where $\mathcal{T}$ is an index set representing time points.

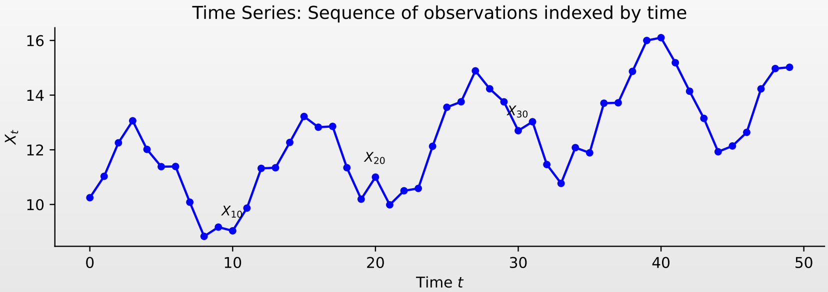
Key Characteristics

- **Ordered**: Natural temporal ordering
- **Dependent**: Consecutive observations correlated
- **Discrete/Continuous**: $t = 1, 2, 3, \ldots$

Notation

- X_t = observation at time t
- $\{X_t\}_{t=1}^{T}$ = series with T observations

Time Series: Visual Illustration

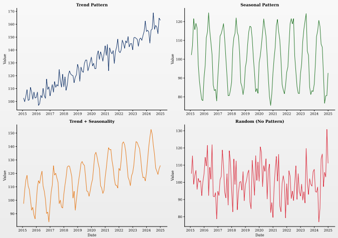


Interpretation

Each point X_t represents an observation at time t. The sequence is ordered and consecutive observations are typically correlated.

TSA_ch1_def_timeseries

Common Time Series Patterns

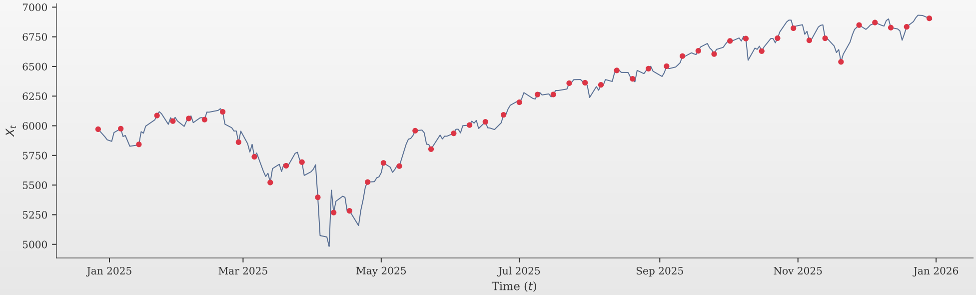


Pattern Types

- **Trend**: Long-term increase or decrease in the data
- **Seasonal**: Regular periodic patterns (e.g., monthly, quarterly)
- **Cyclical**: Medium-term fluctuations (business cycles, 2–10 years)
- **Random**: No systematic pattern – unpredictable fluctuations

TSA_ch1_patterns

Time Series: Visual Definition

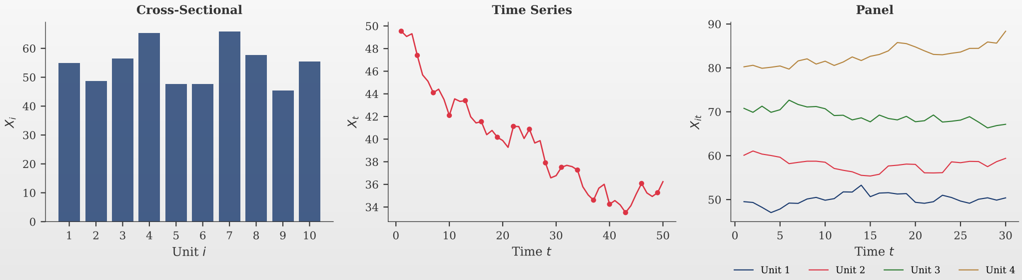


Interpretation

Each point X_t represents a measurement at discrete time t. The temporal ordering creates dependence between observations. Data: S&P 500 (2024).

TSA_ch1_definition

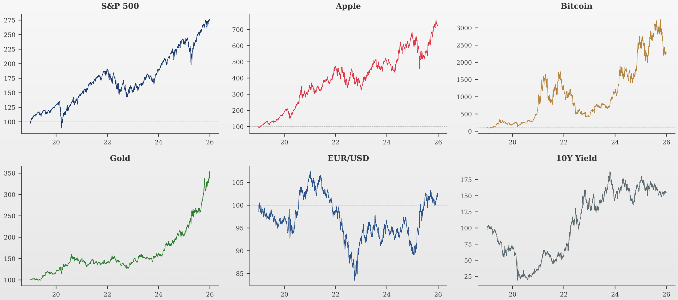
Types of Data: Comparison



Data Type	Units (N)	Time (T)	Example
Cross-sectional	Many	1	Survey of 1000 households
Time series	1	Many	Daily S&P 500 prices
Panel	Many	Many	GDP of 50 countries, 20 years

TSA_ch1_data_types

Examples of Time Series Data



Real Financial Data

Yahoo Finance (2019–2025), normalized to base 100. Notice different volatility patterns: Bitcoin most volatile, Gold most stable.

TSA_ch1_examples

Why Decompose a Time Series?

Decomposition separates a time series into interpretable components:

Goals:

- Understand underlying patterns
- Remove seasonality for modeling
- Identify trend direction
- Isolate irregular fluctuations
- Improve forecasting accuracy

Components:

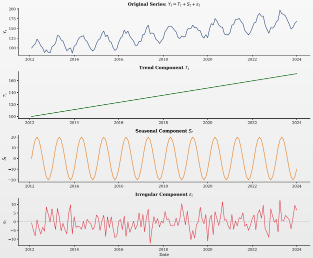
- T_t = **Trend-Cycle**: Long-term movement
- S_t = **Seasonal**: Regular periodic pattern
- ε_t = **Residual**: Random noise

Note: Cyclical component is typically absorbed into T_t

Classical Decomposition Models

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

Time Series Decomposition: Visual Example

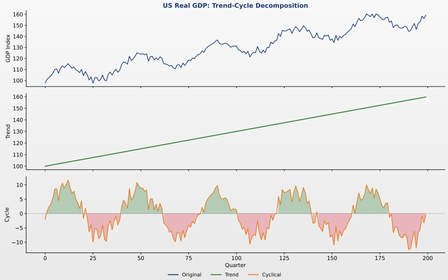


Components Explained

- **Original**: observed series
- **Trend-Cycle**: long-term movement
- **Seasonal**: periodic pattern
- **Residual**: random noise

TSA_ch1_decomposition

The Cyclical Component



Characteristics

- Medium-term fluctuations (2–10 years)
- No fixed period (unlike seasonal)
- Reflects expansions/recessions

In Practice

- Cycle is often combined with trend
- Difficult to identify in short series
- Usually not modeled separately

TSA_ch1_cyclical

Additive Decomposition Model

Model

$$X_t = T_t + S_t + \varepsilon_t \tag{1}$$

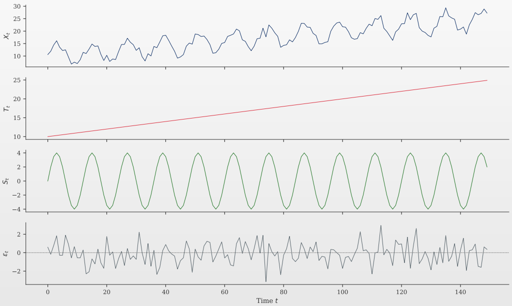
When to Use

- Seasonal fluctuations are **constant** over time
- Variance of the series is **stable**

Properties

- $\mathbb{E}[\varepsilon_t] = 0$ (zero mean)
- $\sum_{j=1}^{s} S_j = 0$ (seasonal sums to zero)
- Units of S_t same as X_t

Additive Decomposition: US Retail Sales (FRED)



Interpretation

Original = Trend + Seasonal + Residual. Seasonal amplitude stays constant. Data: US Retail Sales (RSXFS) from FRED.

TSA_ch0_additive

Multiplicative Decomposition Model

Model

$$X_t = T_t \times S_t \times \varepsilon_t \tag{2}$$

When to Use

- Seasonal fluctuations **grow** with series level
- Variance **increases** over time

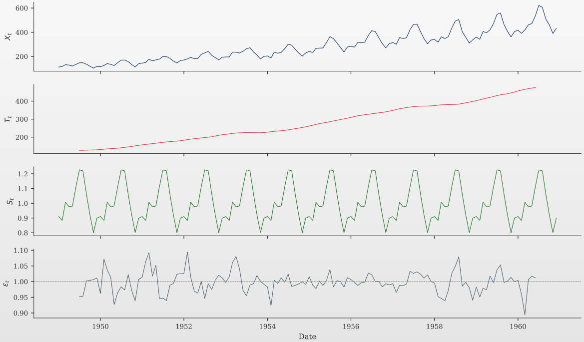
Properties

- $\mathbb{E}[\varepsilon_t] = 1$ (centered at 1)
- $\frac{1}{s} \sum S_j = 1$ (averages to 1)
- S_t is dimensionless ratio

Tip

Log transform converts multiplicative to additive model: $\log X_t = \log T_t + \log S_t + \log \varepsilon_t$

Multiplicative Decomposition: Real Data

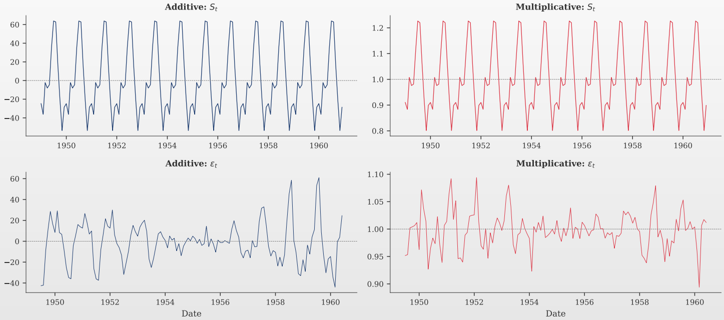


Example

Classic Box-Jenkins airline passengers (1949–1960). Seasonal amplitude grows with level.

TSA_ch0_multiplicative

Additive vs Multiplicative: Comparison



Key Difference

- **Multiplicative**: seasonal is a *ratio* (centered at 1)
- **Additive**: seasonal in *absolute units* (centered at 0)

TSA_ch0_comparison

Trend Estimation: Moving Average

Definition 2 (Centered Moving Average)

The **centered moving average** of order $2q+1$ is:

$$\hat{T}_t = \frac{1}{2q+1} \sum_{j=-q}^{q} X_{t+j} \tag{3}$$

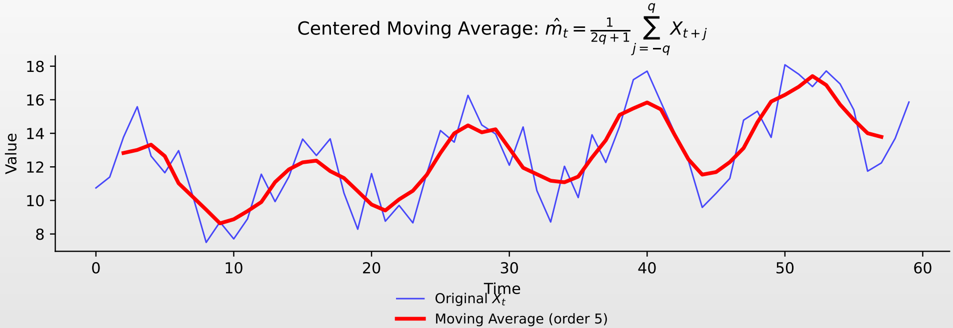
For Seasonal Data

- Period s **odd**: simple average
- Period s **even**: $2 \times s$ MA with half-weights

Properties

- Smooths seasonal & random
- Larger window $\Rightarrow$ smoother
- Trade-off: lose endpoints

Centered Moving Average: Visual Illustration



Interpretation

The moving average smooths out short-term fluctuations, revealing the underlying trend.

TSA_ch0_ma

Classical Decomposition Algorithm

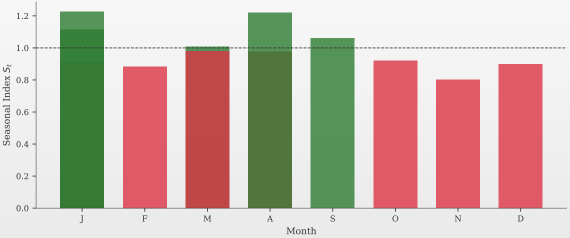
Steps for Multiplicative Decomposition

1. **Estimate Trend**: $\hat{T}_t = MA_s(X_t)$
2. **Detrend**: $D_t = X_t / \hat{T}_t$
3. **Estimate Seasonal**: $\hat{S}_j = \text{mean}(D_t \text{ for season } j)$
4. **Normalize**: Scale so $\frac{1}{s} \sum_{j=1}^{s} \hat{S}_j = 1$
5. **Compute Residuals**: $\hat{\varepsilon}_t = X_t / (\hat{T}_t \times \hat{S}_t)$

Note

For **additive** decomposition: replace division with subtraction and multiplication with addition.

Seasonal Indices: Interpretation



Interpretation

- $S_t > 1$: above-average activity
- $S_t < 1$: below-average activity
- Airline data shows peak travel in July–August

TSA_ch0_seasonal

STL Decomposition: A Modern Approach

Definition 3 (STL - Seasonal-Trend decomposition using LOESS)

STL uses locally weighted regression (LOESS): $X_t = T_t + S_t + R_t$

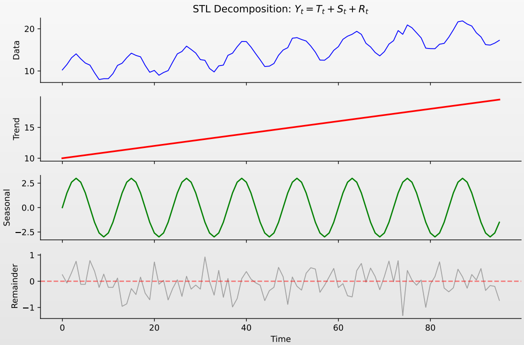
Advantages

- Any seasonal period
- Seasonal can change over time
- Robust to outliers
- Smooth trend estimates

Key Parameters

- `period`: Seasonal period
- `seasonal`: Smoothing window
- `robust`: Downweight outliers

STL Decomposition: Visual Illustration



Key Insight

STL separates the series into trend, seasonal, and remainder using LOESS.

TSA_ch0_stl

Exponential Smoothing: Overview

Definition

Exponential smoothing produces forecasts based on weighted averages of past observations, with weights decaying exponentially.

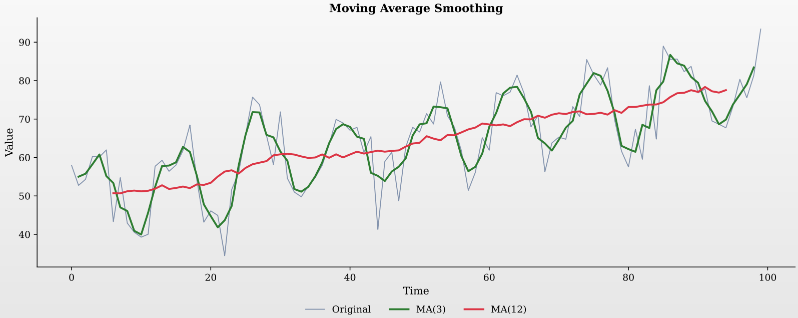
Why Exponential Smoothing?

- Simple yet effective
- Recent obs. get higher weights
- Handles trend & seasonality
- Foundation for ETS models

Three Main Methods

1. **SES**: Level only
2. **Holt**: Level + Trend
3. **Holt-Winters**: + Seasonality

Moving Average Smoothing



Window Size Trade-off

- **Small window**: Responsive but noisy
- **Large window**: Smoother but slower to react

TSA_ch1_moving_average

Simple Exponential Smoothing (SES)

Model

$$\hat{X}_{t+1|t} = \alpha X_t + (1-\alpha)\hat{X}_{t|t-1} \tag{4}$$

where $\alpha \in (0,1)$ is the **smoothing parameter**.

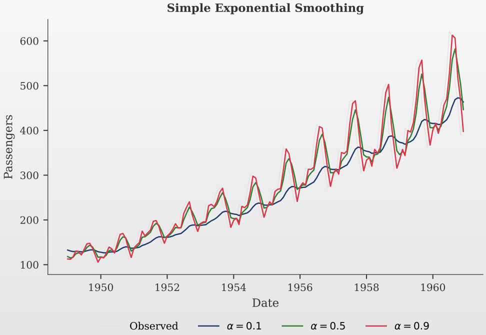
How It Works

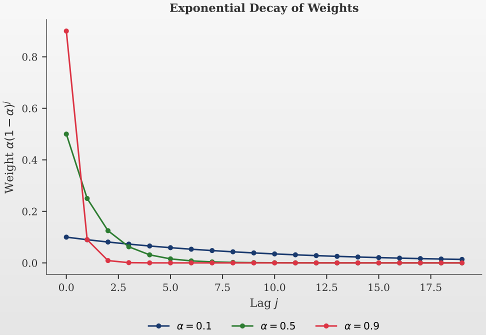
- Weights decay exponentially
- Large α: responsive
- Small α: smoother

Level Form

$$\ell_t = \alpha X_t + (1-\alpha)\ell_{t-1}$$

Simple Exponential Smoothing: Effect of α





Trade-off

Smaller α produces smoother forecasts; larger α follows data more closely.

TSA_ch0_ses

Holt's Linear Trend Method

Equations

- **Level:** $\ell_t = \alpha X_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$
- **Trend:** $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$
- **Forecast:** $\hat{X}_{t+h|t} = \ell_t + h \cdot b_t$

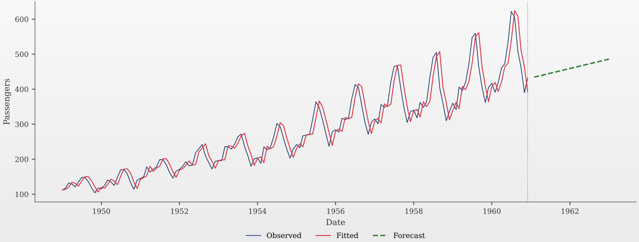
Parameters

- α: Level smoothing
- β^*: Trend smoothing

Components

- ℓ_t: Estimated level
- b_t: Estimated trend (slope)

Holt's Method: Visualization



Interpretation

- Holt's method captures both level and trend
- Projects them into the forecast horizon
- α controls responsiveness to level changes
- β^* controls responsiveness to trend changes

TSA_ch0_holt

Holt-Winters Seasonal Method

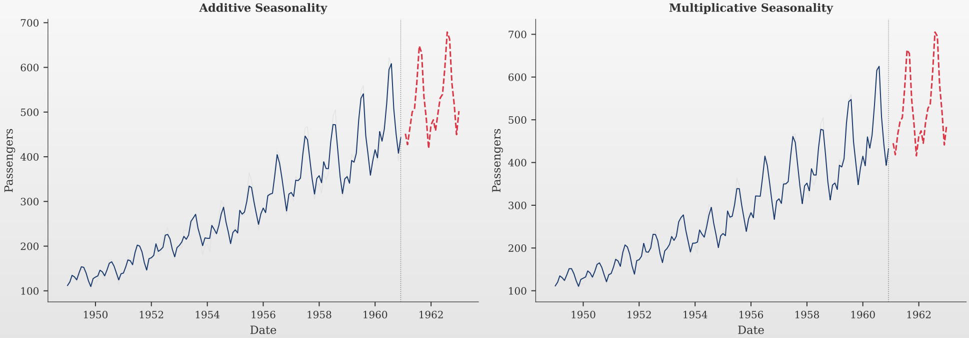
Equations (Additive Seasonality)

- **Level:** $\ell_t = \alpha(X_t - S_{t-s}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$
- **Trend:** $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$
- **Seasonal:** $S_t = \gamma(X_t - \ell_t) + (1-\gamma)S_{t-s}$
- **Forecast:** $\hat{X}_{t+h|t} = \ell_t + h \cdot b_t + S_{t+h-s(k+1)}$

Parameters

- α: Level smoothing
- β^*: Trend smoothing
- γ: Seasonal smoothing
- s: Seasonal period

Holt-Winters: Capturing Seasonality



Key Feature

Holt-Winters decomposes the series and produces seasonal forecasts with trend.

ETS Framework: Error-Trend-Seasonal

Definition 4 (ETS Models)

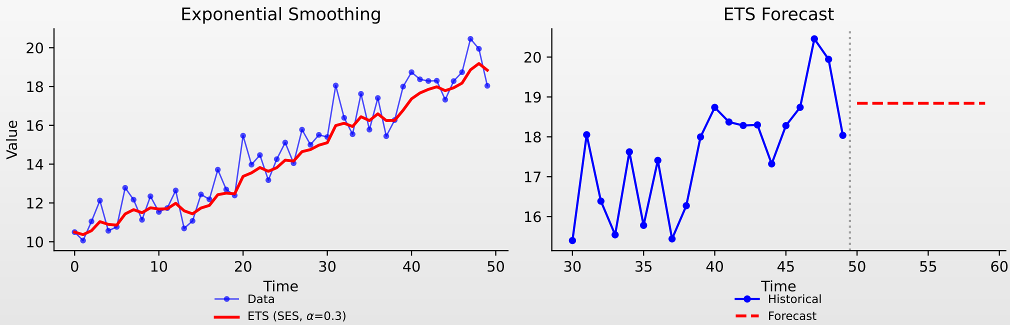
The **ETS framework** generalizes exponential smoothing: ETS(E, T, S)

Component	**N**	**A**	**M**
Error (E)	–	Additive	Multiplicative
Trend (T)	None	Additive	Multiplicative
Seasonal (S)	None	Additive	Multiplicative

Examples

- ETS(A,N,N) = Simple Exponential Smoothing
- ETS(A,A,N) = Holt's Linear Method
- ETS(A,A,A) = Holt-Winters Additive

ETS: Exponential Smoothing Illustration

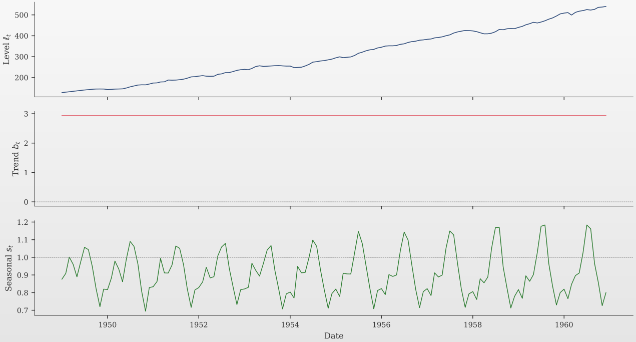


Interpretation

ETS models use exponentially weighted observations for forecasting. Weights decay as observations get older.

TSA_ch0_ets

ETS Model Selection



Interpretation

The ETS framework provides a systematic way to choose the best model using AIC/BIC.

TSA_ch0_ets_select

Damped Trend Methods

Damping Parameter

Introduces $\phi \in (0, 1)$ to prevent over-projection

Equations

- **Level:** $\ell_t = \alpha X_t + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$
- **Trend:** $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}$
- **Forecast:** $\hat{X}_{t+h|t} = \ell_t + \phi \frac{1-\phi^h}{1-\phi} b_t$

Key Insight

- As $h \to \infty$: forecast $\to$ constant
- Prevents unrealistic long-term extrapolation
- Often best for longer horizons

Forecast Accuracy Metrics

Forecast Error

- $e_t = X_t - \hat{X}_t$ (actual minus predicted)

Scale-Dependent

- $\text{MAE} = \frac{1}{n} \sum |e_t|$
- $\text{MSE} = \frac{1}{n} \sum e_t^2$
- $\text{RMSE} = \sqrt{\text{MSE}}$

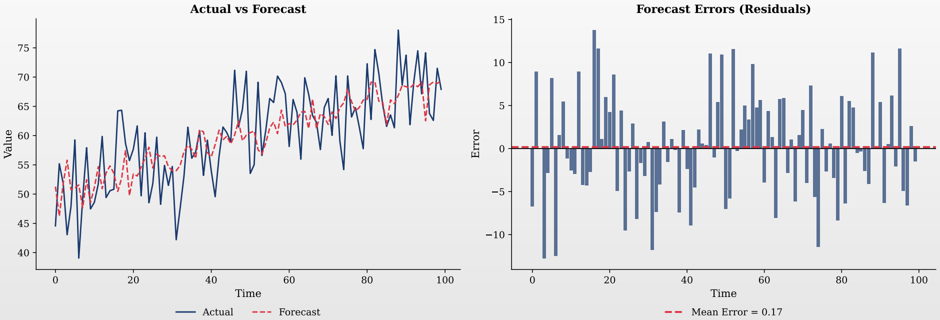
Scale-Independent

- $\text{MAPE} = \frac{100}{n} \sum \left| \frac{e_t}{X_t} \right|$
- $\text{sMAPE} = \frac{100}{n} \sum \frac{|e_t|}{(|X_t| + |\hat{X}_t|)/2}$

Which to use?

- Same series: RMSE, MAE
- Compare across series: MAPE, sMAPE

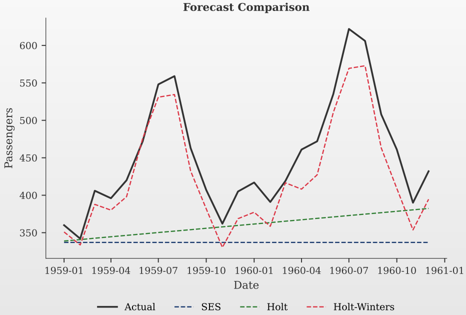
Forecast Evaluation: Visual Example

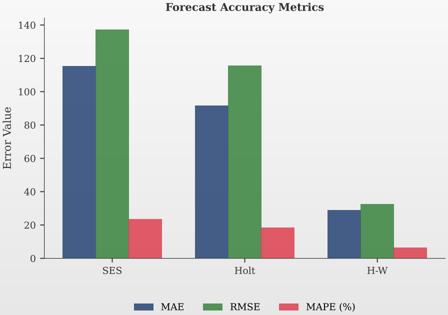


- **Top**: Actual values vs. forecasted values – visual assessment of fit
- **Bottom**: Residuals should be centered around zero with no pattern
- Good forecasts have small, random residuals with constant variance

TSA_ch1_forecast_eval

Comparing Forecast Methods





Interpretation

- **Left:** Comparing SES, Holt, and Holt-Winters forecasts
- **Right:** Error metrics for each method

TSA_ch0_forecast_compare

Residual Diagnostics

Residual Properties

Good forecasts should have residuals that are:

1. **Zero mean**: $\mathbb{E}[e_t] = 0$
2. **Uncorrelated**: $\text{Cov}(e_t, e_{t-k}) = 0$
3. **Constant variance**: $\text{Var}(e_t) = \sigma^2$
4. **Normally distributed**

Diagnostic Tests

Ljung-Box test (autocorrelation):

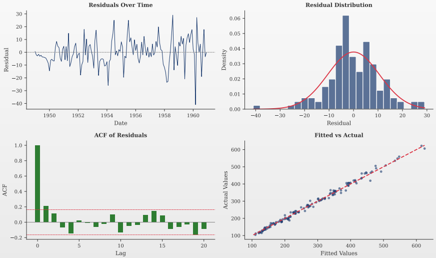
$$Q = T(T+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{T-k} \sim \chi^2_h$$

Jarque-Bera test (normality):

$$JB = \frac{T}{6}\left(S^2 + \frac{(K-3)^2}{4}\right) \sim \chi^2_2$$

S = skewness, K = kurtosis

Residual Diagnostics: Visualization



What to Check

- Time plot (no patterns)
- Histogram (normality)
- ACF (no autocorrelation)
- Q-Q plot (normality)

TSA_ch0_residuals

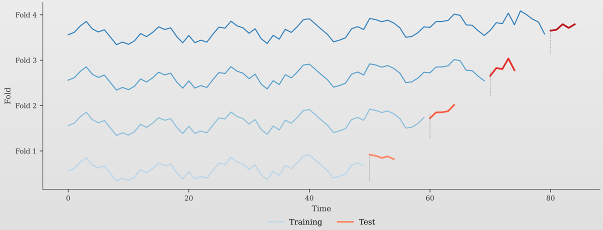
Time Series Cross-Validation

Why Not Standard CV?

- Time series have temporal dependence
- Future data cannot predict the past
- Standard k-fold causes data leakage

Rolling Origin CV

1. Train on $\{X_1, \ldots, X_t\}$
2. Forecast $\hat{X}_{t+h}$
3. Increment t, repeat



TSA_ch0_cv

Train / Validation / Test Split

Three-way split for model development:

Training Set

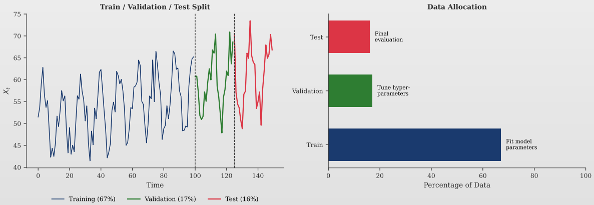
- Fit model parameters
- Largest portion (60–80%)
- Used for estimation

Validation Set

- Tune hyperparameters
- Compare models
- Select best approach

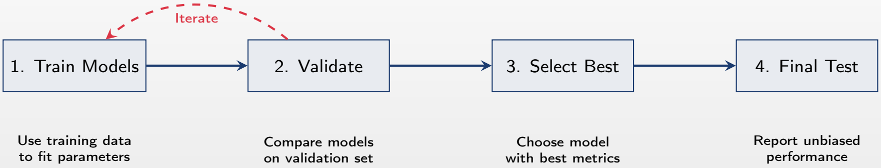
Test Set

- Final evaluation only
- Never used for tuning
- Unbiased performance



TSA_ch0_train_split

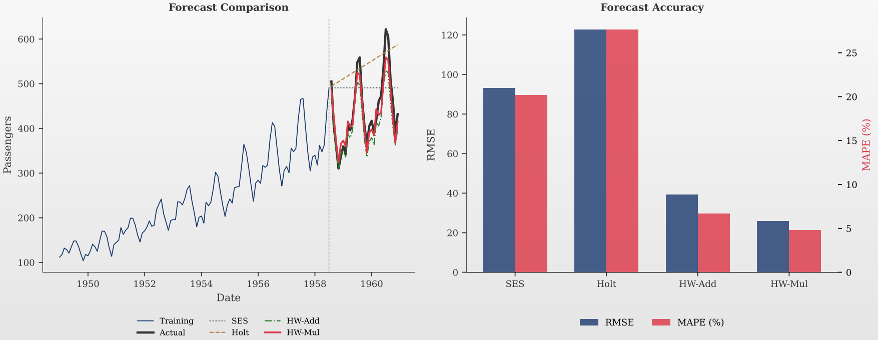
Model Development Workflow



Critical Rule

Never use test set for model selection! This causes *data leakage* and overly optimistic performance estimates.

Real Data: Forecast Comparison

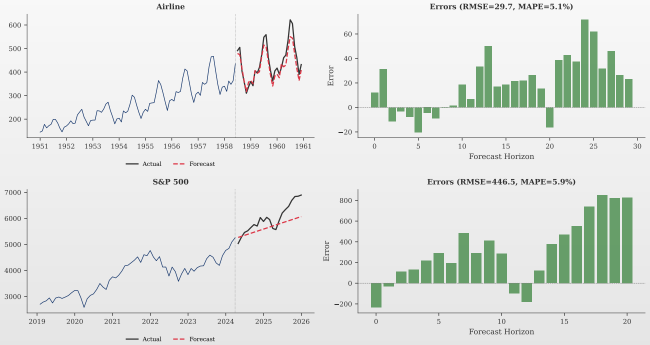


Interpretation

Airline passengers data: Holt-Winters Multiplicative performs best for seasonal data.

TSA_ch0_real_data

Forecast Performance Across Datasets



Interpretation

Different series require different models. Seasonal data needs seasonal methods.

TSA_ch0_multi_series

Modeling Seasonality: Two Approaches

1. Dummy Variables:

$X_t = \mu + \sum_{j=1}^{s-1} \gamma_j D_{jt} + \varepsilon_t$

- $D_{jt} = 1$ if t in season j
- $s - 1$ parameters
- Any seasonal pattern

2. Fourier Terms:

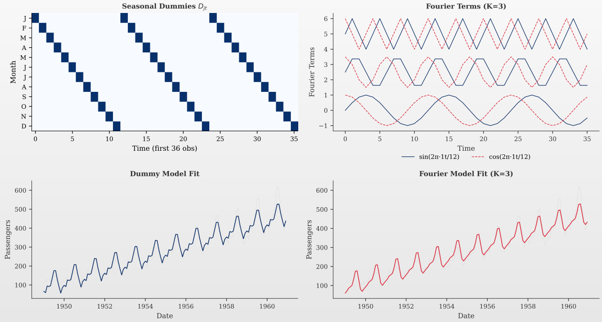
$X_t = \mu + \sum_{k=1}^{K} [\alpha_k \sin(\cdot) + \beta_k \cos(\cdot)]$

- Sinusoidal functions
- $2K$ parameters
- Smooth patterns

Trade-off

- **Dummies**: any pattern, more parameters
- **Fourier**: smooth, fewer parameters

Dummy Variables vs Fourier Terms



Comparison

- **Dummies**: capture any shape but need $s-1$ parameters
- **Fourier**: uses $2K$ parameters for smooth patterns

TSA_ch0_fourier

Choosing Between Dummies and Fourier

Criterion	Dummies	Fourier
Parameters (monthly)	11	$2K$ (often 4–6)
Seasonal pattern	Any shape	Smooth/sinusoidal
Interpretation	Direct (month effects)	Frequency components
High-frequency seasons	Many parameters	Efficient
Multiple seasonality	Complex	Easy (add terms)

Guidelines

- Use **dummies**: irregular patterns, interpretable coefficients
- Use **Fourier**: smooth patterns, high-frequency seasonality, multiple periods
- **Fourier terms** are used in TBATS and Facebook Prophet

Why Remove Trend and Seasonality?

Before modeling, we often need to make series stationary:

Reasons to detrend:

- Stationarity requirement
- Focus on fluctuations
- Avoid spurious regression
- Enable valid inference

Reasons to deseasonalize:

- Reveal underlying trend
- Compare across seasons
- Simplify modeling
- Focus on irregular component

Important

After modeling the detrended/deseasonalized series, we must **reverse the transformation** for forecasting.

Trend Removal Methods

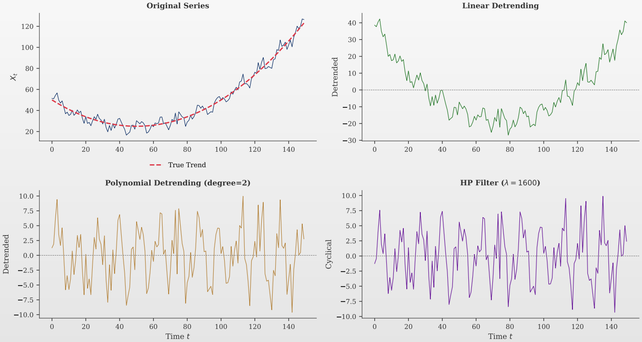
Six Common Detrending Approaches

1. **Differencing**: $\Delta X_t = X_t - X_{t-1}$
2. **Linear regression**: $\hat{T}_t = \hat{\beta}_0 + \hat{\beta}_1 t$
3. **Polynomial**: Higher-order polynomial
4. **HP Filter**: Balance fit vs smoothness
5. **Moving average**: $\hat{T}_t = MA_q(X_t)$
6. **LOESS**: Local polynomial regression

Choice Depends On

- Nature of trend (deterministic vs stochastic)
- Purpose (forecasting vs analysis)

Detrending Methods: Comparison

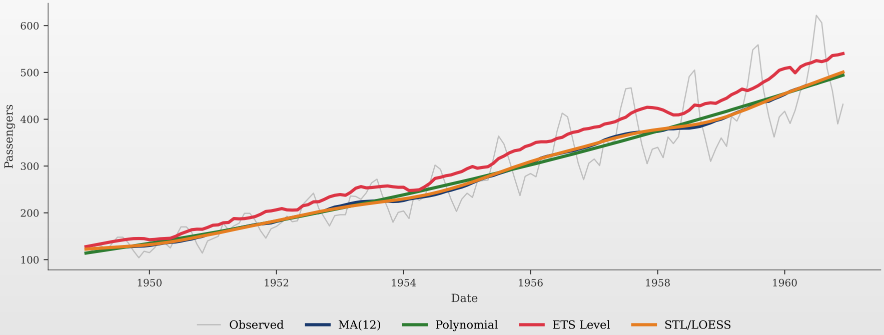


Key Insight

Different methods produce different residuals. Choose based on trend type and analysis goals.

TSA_ch0_detrending

Trend Estimation: Multiple Approaches



Interpretation

Different methods capture trend at varying levels of smoothness.

TSA_ch0_trend

Hodrick-Prescott (HP) Filter

Definition 5 (HP Filter)

The **HP filter** decomposes X_t into trend τ_t and cycle c_t: $X_t = \tau_t + c_t$, by minimizing:

$$\min_{\{\tau_t\}} \left\{ \sum_{t=1}^{T} (X_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} [(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2 \right\}$$

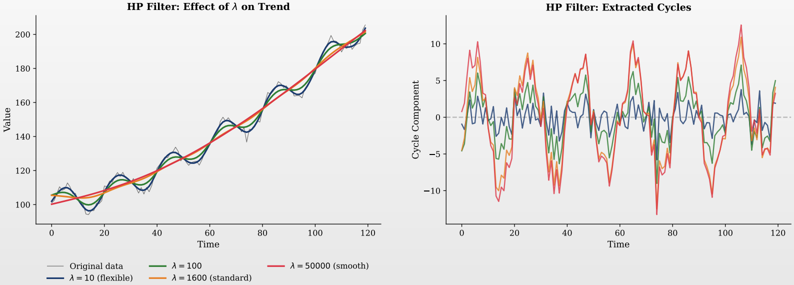
Interpretation

- First term: fit to data
- Second term: smoothness penalty
- λ: trade-off parameter

Standard λ Values

- Annual: $\lambda = 6.25$
- Quarterly: $\lambda = 1600$
- Monthly: $\lambda = 129600$

HP Filter: Effect of λ

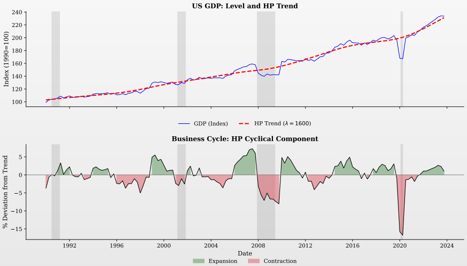


Trade-off

- **Small** λ: Trend follows data closely (more flexible)
- **Large** λ: Trend becomes smoother (approaches linear trend)

TSA_ch1_hp_filter

HP Filter: Business Cycle Extraction



Application

HP filter is widely used in macroeconomics to extract business cycles from GDP and other economic series.

TSA_ch0_hp_cycle

HP Filter: Limitations

Known Issues

- **End-point problem**: Trend estimates unreliable at endpoints
- **Spurious cycles**: Can create artificial dynamics
- λ **choice**: Results sensitive to parameter
- **Non-stationary**: Assumes trend is smooth

Alternatives

- **Band-pass filters**: Baxter-King, Christiano-Fitzgerald
- **Hamilton filter**: Regression-based
- **Unobserved components**: State-space models

Hamilton (2018) Critique

"Why You Should Never Use the Hodrick-Prescott Filter" — suggests using regression on lagged values instead.

Seasonality Removal Methods

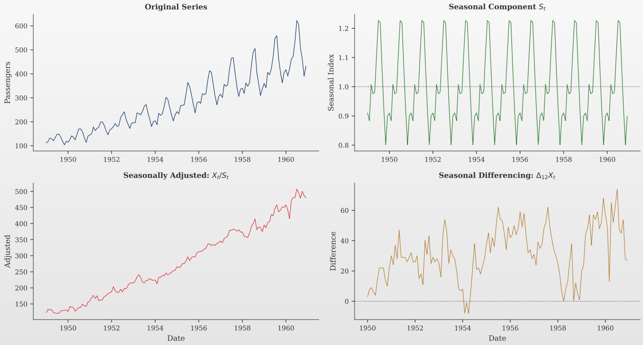
Four Approaches to Remove Seasonality

1. **Seasonal differencing**: $\Delta_s X_t = X_t - X_{t-s}$
2. **Division** (multiplicative): $X_t^{adj} = X_t / \hat{S}_t$
3. **Subtraction** (additive): $X_t^{adj} = X_t - \hat{S}_t$
4. **X-13ARIMA-SEATS**: Government statistical method

Seasonal Period s

- Monthly $\Rightarrow s = 12$
- Quarterly $\Rightarrow s = 4$

Seasonal Adjustment: Visualization



Result

Seasonally adjusted series reveals underlying trend without periodic fluctuations.

TSA_ch0_seasonal_adj

Deterministic vs Stochastic Trend

Deterministic Trend:

$$X_t = \beta_0 + \beta_1 t + \varepsilon_t$$

- Trend is a function of time
- Detrend by regression
- ε_t is stationary

Stochastic Trend:

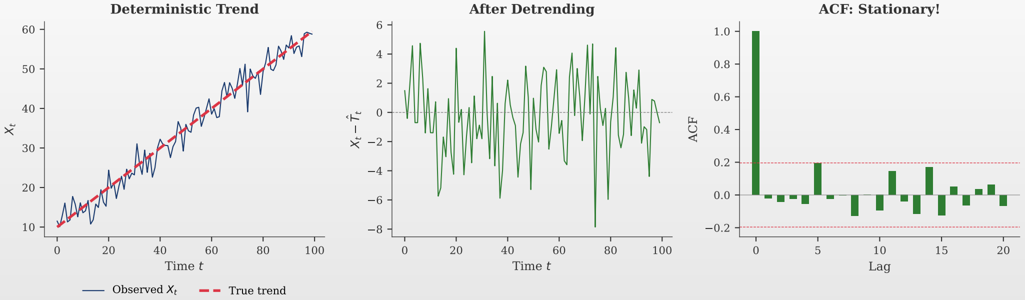
$$X_t = X_{t-1} + \varepsilon_t$$

- Random walk component
- Detrend by differencing
- ΔX_t is stationary

Wrong Method = Problems

- Differencing deterministic trend $\Rightarrow$ over-differencing
- Regression on stochastic trend $\Rightarrow$ spurious regression

Example: Deterministic Trend

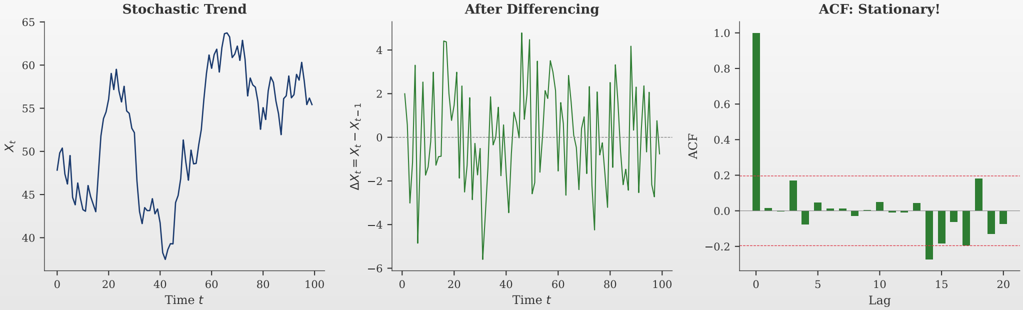


Key

Use regression to remove trend → residuals are stationary (ACF decays quickly).

TSA_ch0_det_trend

Example: Stochastic Trend (Random Walk)

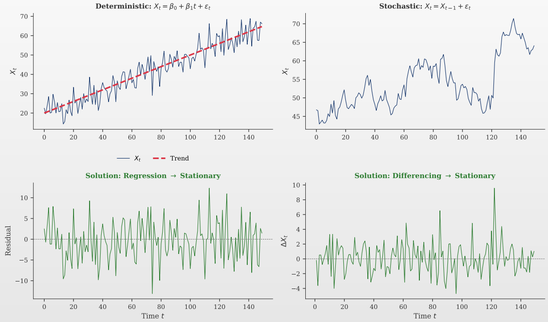


Key

Use differencing to remove trend → differences are stationary (white noise).

TSA_ch0_stoch_trend

Side-by-Side Comparison



Remember

- Deterministic trend → regression
- Stochastic trend → differencing

TSA_ch0_trend_compare

Summary

What We Learned

- **Time Series Definition**: Sequence of observations indexed by time
- **Decomposition**: Trend-Cycle + Seasonal + Residual components
- **Exponential Smoothing**: SES, Holt, Holt-Winters, ETS framework
- **Forecast Evaluation**: MAE, RMSE, MAPE; train/validation/test splits

Key Takeaway

- **Understand Before Modeling**:
  - Always visualize and decompose your data first
  - Choose additive vs multiplicative based on variance behavior

Quick Quiz

1. What is the difference between additive and multiplicative decomposition?
2. When should you use Holt-Winters instead of simple exponential smoothing?
3. Why can't we use standard k-fold cross-validation for time series?
4. What does $\alpha = 0.9$ mean in exponential smoothing?
5. How do you distinguish between deterministic and stochastic trend?

Quiz Answers

1. **Additive vs Multiplicative**: Additive when seasonal amplitude is constant; multiplicative when it grows with the level.
2. **Holt-Winters**: When data has both trend AND seasonality. SES only handles level.
3. **Time Series CV**: Standard k-fold ignores temporal order — would use future data to predict the past (data leakage).
4. $\alpha = 0.9$: High weight on recent observations, forecast reacts quickly to changes but is more volatile.
5. **Trend type**: Deterministic — predictable function of time (use regression). Stochastic — random walk component (use differencing).

What Comes Next?

Chapter 1: Stochastic Processes and Stationarity

- **Stochastic Processes**: Mathematical foundation for time series
  - Random variables indexed by time
  - Strict vs weak (covariance) stationarity
- **Key Processes**: White noise and random walk
  - Building blocks for ARIMA models
  - Understanding mean reversion vs unit roots
- **ACF and PACF**: Tools for model identification
  - Detecting autocorrelation structure
  - Choosing AR and MA orders

Questions?