# Chapter 1: Introduction to Time Series

Fundamentals and Concepts

## 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, seasonal, and residual components

3. **Apply** exponential smoothing (SES, Holt, Holt-Winters, ETS)

4. **Evaluate** forecasts using MAE, RMSE, MAPE; train/validation/test splits

5. **Model** seasonality using dummy variables or Fourier terms

6. **Handle** trend and seasonality through detrending and adjustment

7. **Understand** stochastic processes and stationarity

8. **Compute** ACF/PACF and conduct stationarity tests (ADF, KPSS)

# Chapter Outline

## 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**: Observations have a natural temporal ordering
- **Dependent**: Consecutive observations are typically correlated
- **Discrete** or **Continuous**: Time index can be discrete ($t = 1, 2, 3, \ldots$) or continuous
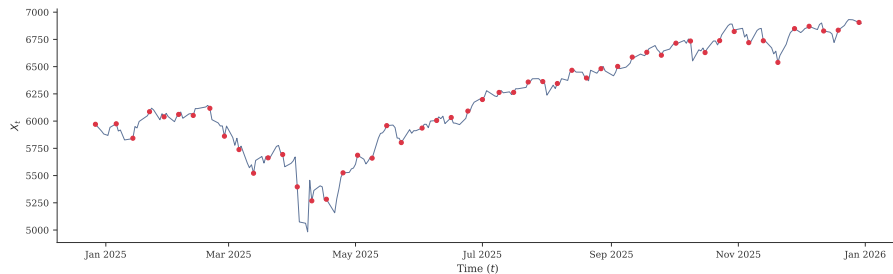
**Notation:**

- $X_t =$ observation at time $t$
- $\{X_t\}_{t=1}^{T} =$ finite time series with $T$ observations
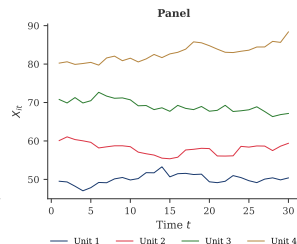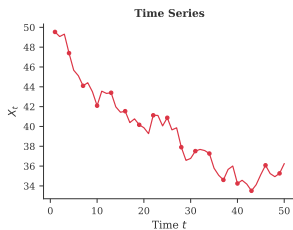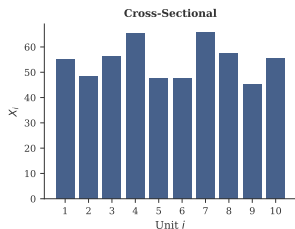
# Common Time Series Patterns
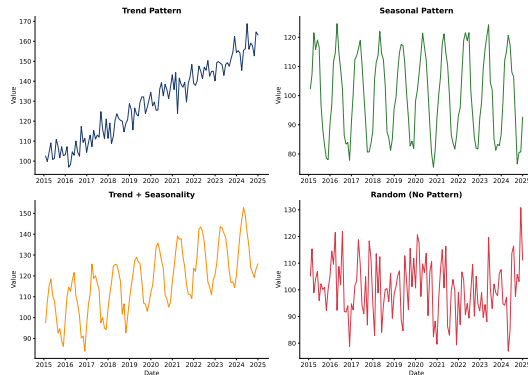


- **Trend**: Long-term increase or decrease in the data
- **Seasonal**: Regular periodic patterns (e.g., monthly, quarterly)
- **Random**: No systematic pattern – unpredictable fluctuations
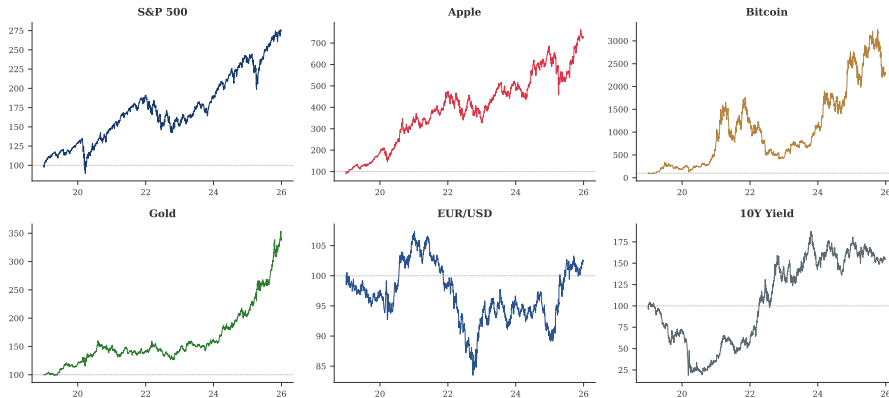
# Time Series: Visual Definition



Each point $X_t$ represents a measurement at discrete time $t$. Data: S&P 500 (2024).

# 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 |

# Examples of Time Series Data



Real financial data from Yahoo Finance (2019–2025). Normalized to base 100.

# 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
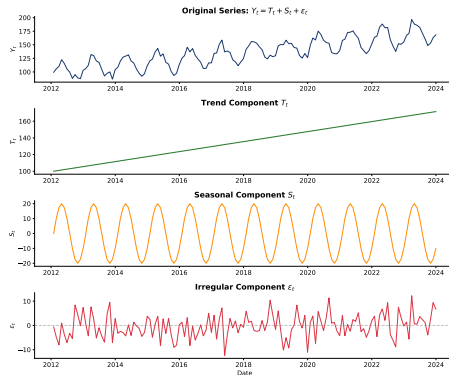- Improve forecasting accuracy

**Components:**

- $T_t = $ **Trend**: Long-term movement
- $S_t = $ **Seasonal**: Regular periodic pattern
- $C_t = $ **Cyclical**: Business cycle fluctuations
- $\varepsilon_t = $ **Residual**: Random noise

## 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



- **Original**: The observed time series with all components
- **Trend**: Underlying long-term movement extracted via smoothing
- **Seasonal**: Regular periodic pattern that repeats each cycle
- **Residual**: Random noise after removing trend and seasonality

# The Cyclical Component

**Cyclical component** $C_t$: Medium-term fluctuations (2–10 years)

**Characteristics:**

- Business cycle fluctuations
- No fixed period (unlike seasonal)
- Duration varies: 2–10 years
- Amplitude varies over time

**Examples:**

- Economic expansions/recessions
- Credit cycles
- Real estate cycles
- Commodity price cycles

## Practical Note

Often combined with trend as **trend-cycle** component because:

- Difficult to separate from trend with short data
- Many decomposition methods estimate $T_t + C_t$ together

# Additive Decomposition Model

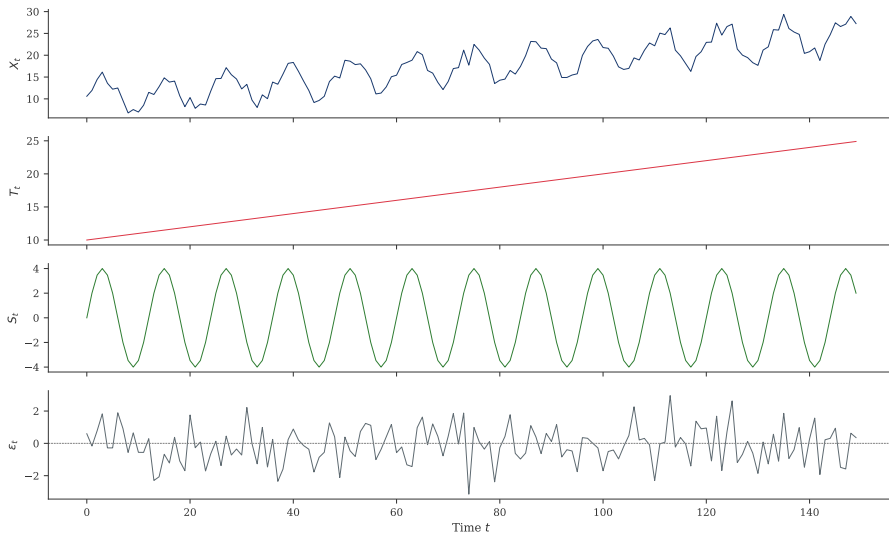**Formula:** $X_t = T_t + S_t + \varepsilon_t$

**When to use:**

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

**Properties:**

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

# Additive Decomposition: Visualization

# Multiplicative Decomposition Model

**Formula:** $\quad X_t = T_t \times S_t \times \varepsilon_t$
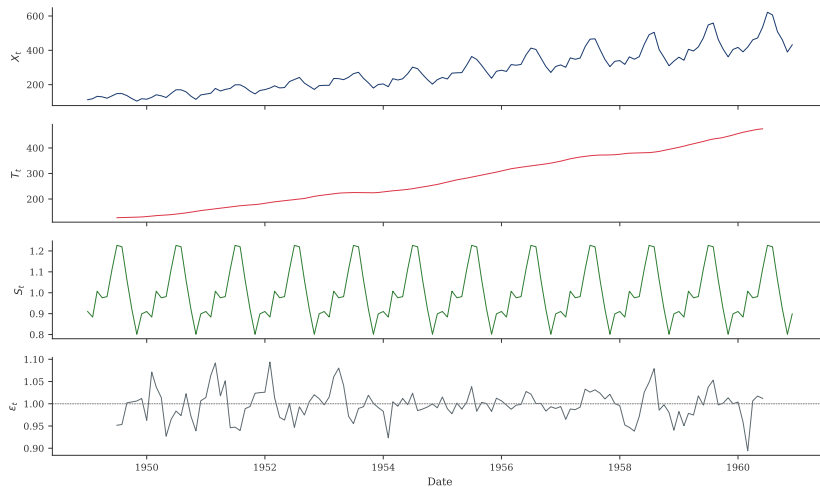
**When to use:**

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

**Properties:**

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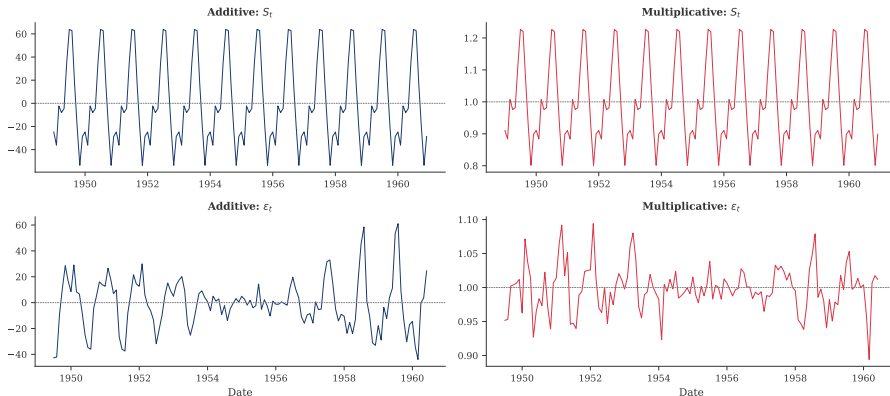- $S_t$ is a ratio (dimensionless)

**Tip:** Log transform converts to additive model.

# Multiplicative Decomposition: Real Data



Classic Box-Jenkins airline passengers dataset (1949–1960).

# Additive vs Multiplicative: Comparison



**Key difference:** In multiplicative model, seasonal component is a *ratio* (centered at 1), while in additive model it's in *absolute units* (centered at 0).

# Trend Estimation: Moving Average

## Definition 2 (Centered Moving Average)

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

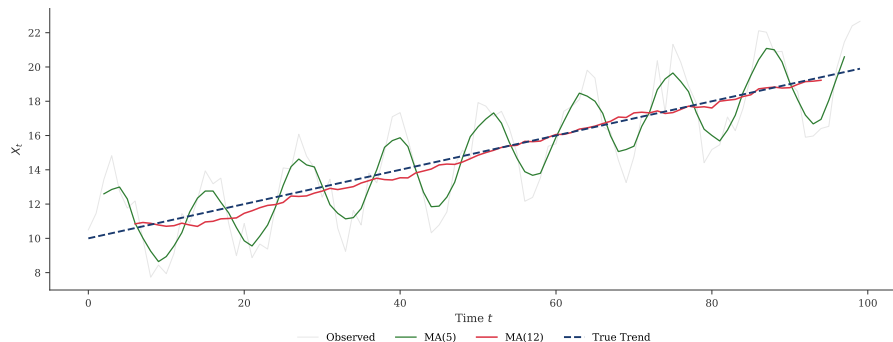$$\hat{T}_t = \frac{1}{2q + 1} \sum_{j=-q}^{q} X_{t+j}$$

**For seasonal data:**

- If period $s$ is **odd**: simple average over $s$ observations
- If period $s$ is **even** (e.g., 12): use $2 \times s$ MA with half-weights at ends

**Properties:**

- Smooths out seasonal and random fluctuations
- Larger window $\Rightarrow$ smoother trend
- Trade-off: lose data at endpoints

# Moving Average: Effect of Window Size



Larger window sizes produce smoother trends but lose more data at boundaries.

## Classical Decomposition Algorithm

**Steps for Multiplicative Decomposition:**

1. **Estimate Trend**: $\hat{T}_t = MA_s(X_t)$

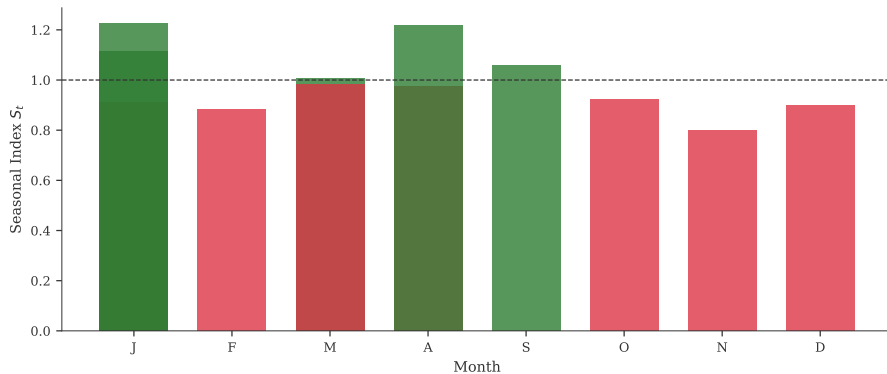2. **Detrend**: $D_t = X_t/\hat{T}_t$

3. **Estimate Seasonal**: Average $D_t$ for each season $j$

$$\hat{S}_j = \text{mean}(D_t \text{ for all } t \text{ in season } j)$$

4. **Normalize**: Scale so $\frac{1}{s}\sum_{j=1}^{s}\hat{S}_j = 1$

5. **Compute Residuals**: $\hat{\varepsilon}_t = X_t/(\hat{T}_t \times \hat{S}_t)$

# Seasonal Indices: Interpretation



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

## STL Decomposition: A Modern Approach

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

**STL** uses locally weighted regression (LOESS) to estimate components:

$$X_t = T_t + S_t + R_t$$

**Advantages over classical decomposition:**

- Handles **any seasonal period** (not just 4 or 12)
- Seasonal component can **change over time**
- **Robust** to outliers (with robust=True option)
- Provides **smooth** trend estimates

**Key parameters:**

- `period`: Seasonal period (e.g., 12 for monthly)
- `seasonal`: Window for seasonal smoothing (odd integer)
- `robust`: Use robust fitting to downweight outliers

# STL Decomposition: Example
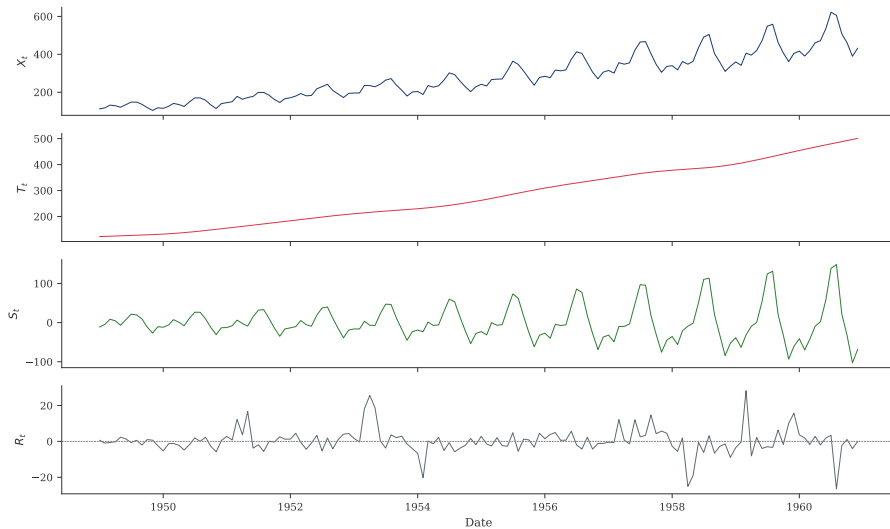
# Exponential Smoothing: Overview

**Exponential smoothing** methods produce forecasts based on weighted averages of past observations, with weights decaying exponentially.
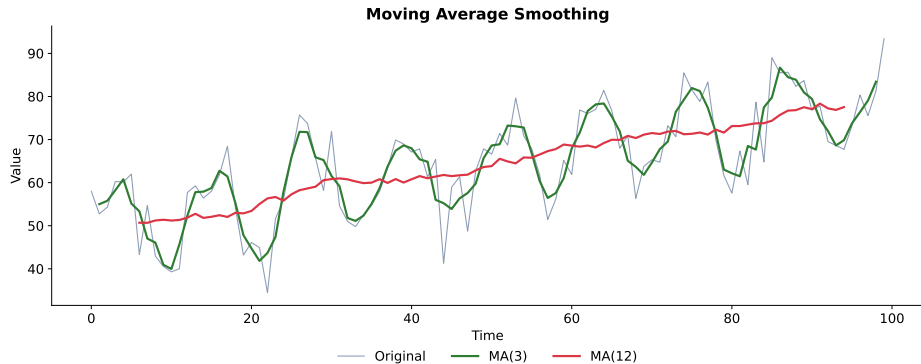
## Why Exponential Smoothing?

- Simple yet effective forecasting methods
- More recent observations get higher weights
- Handles trend and seasonality
- Foundation for ETS models

**Three main methods:**

1. **Simple Exponential Smoothing (SES)**: Level only
2. **Holt's Method**: Level + Trend
3. **Holt-Winters**: Level + Trend + Seasonality

# Moving Average Smoothing



Moving Average Smoothing

- **Small window** (e.g., 5): Responsive to changes but noisy
- **Large window** (e.g., 30): Smoother but slower to react
- Trade-off between noise reduction and lag in detecting changes

# Simple Exponential Smoothing (SES)

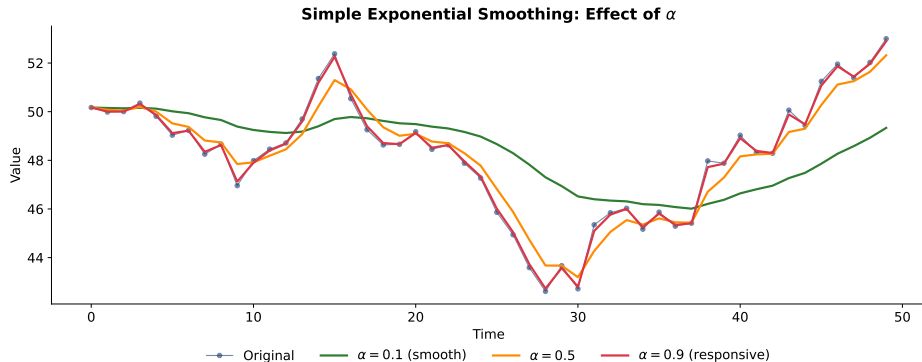**Forecast:** $\quad \hat{X}_{t+1|t} = \alpha X_t + (1-\alpha)\hat{X}_{t|t-1}$

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

**How it works:**

- Weights decay exponentially into the past
- Large $\alpha$: responsive to recent changes
- Small $\alpha$: smoother, more stable forecasts

**Level form:** $\quad \ell_t = \alpha X_t + (1-\alpha)\ell_{t-1}$

# Exponential Smoothing: Effect of Alpha



Simple Exponential Smoothing: Effect of $\alpha$

Legend: Original — $\alpha = 0.1$ (smooth) — $\alpha = 0.5$ — $\alpha = 0.9$ (responsive)

- **Low** $\alpha$ (e.g., 0.1): More weight on past – smoother, slower adaptation
- **High** $\alpha$ (e.g., 0.9): More weight on recent – responsive, more volatile
- Choose $\alpha$ based on how quickly the underlying process changes

# Simple Exponential Smoothing: Effect of $\alpha$
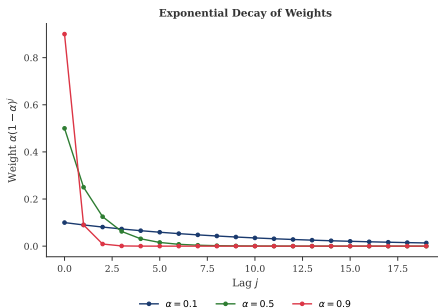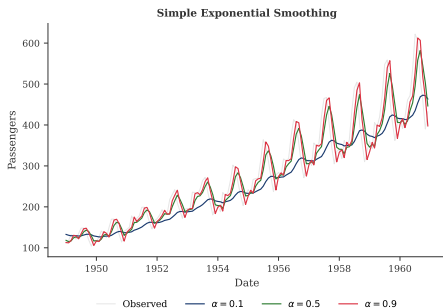


Smaller $\alpha$ produces smoother forecasts; larger $\alpha$ follows data more closely.

## Holt's Linear Trend Method

Extends SES to capture **linear trend** using two equations:

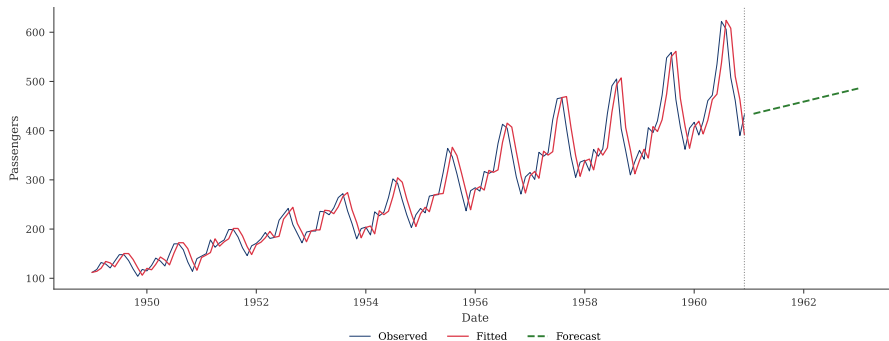**Level:** $\quad \ell_t = \alpha X_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$

**Trend:** $\quad b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$

**Forecast:** $\quad \hat{X}_{t+h|t} = \ell_t + h \cdot b_t$

**Parameters:**

- $\alpha \in (0,1)$: Level smoothing parameter
- $\beta^* \in (0,1)$: Trend smoothing parameter
- $\ell_t$: Estimated level at time $t$
- $b_t$: Estimated trend (slope) at time $t$

# Holt's Method: Visualization



Holt's method captures both level and trend, projecting them into the forecast horizon.

## Holt-Winters Seasonal Method

Extends Holt's method to include **seasonality** with three equations:

**Level:** $\quad \ell_t = \alpha(X_t - S_{t-s}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$

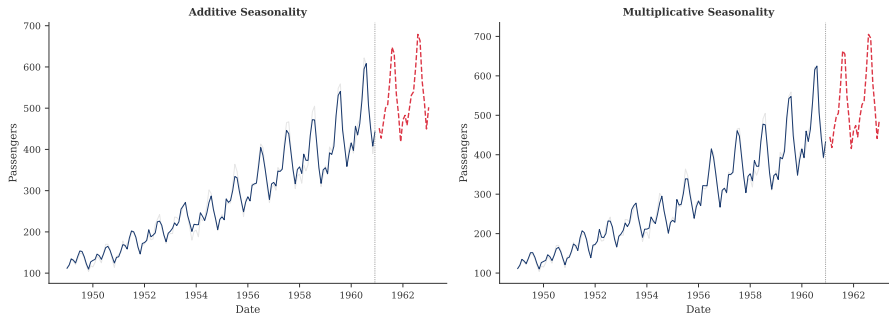**Trend:** $\quad b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$

**Seasonal:** $\quad S_t = \gamma(X_t - \ell_t) + (1-\gamma)S_{t-s}$

**Forecast:** $\quad \hat{X}_{t+h|t} = \ell_t + h \cdot b_t + S_{t+h-s(k+1)}$

**Parameters:**

- $\alpha$: Level smoothing
- $\beta^*$: Trend smoothing
- $\gamma$: Seasonal smoothing
- $s$: Seasonal period (e.g., 12 for monthly)

# Holt-Winters: Capturing Seasonality



Holt-Winters decomposes the series and produces seasonal forecasts.

ETS Framework: Error-Trend-Seasonal

## Definition 4 (ETS Models)

The **ETS framework** generalizes exponential smoothing with explicit error structure:

$$ETS(E, T, S)$$

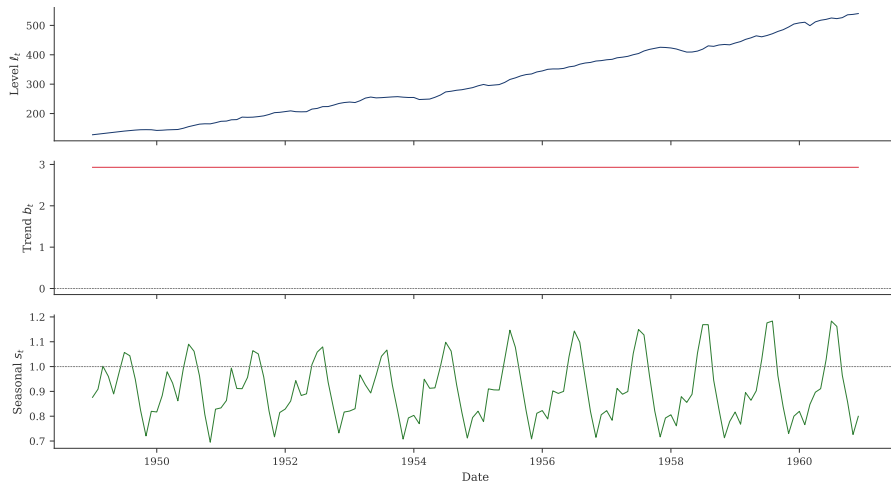where each component can be:

| 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(M,A,M) = Multiplicative errors, additive trend, multiplicative seasonal

# ETS Model Selection



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

# Damped Trend Methods

Introduces **damping parameter** $\phi \in (0, 1)$ to prevent over-projection:

**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 $\to \ell_t + \frac{\phi}{1-\phi} b_t$ (constant)
- Prevents unrealistic long-term extrapolation
- Often best for longer forecast horizons

## Forecast Accuracy Metrics

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

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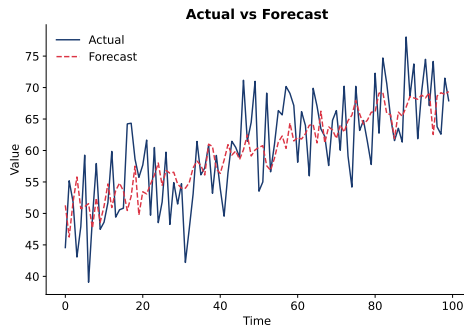- $\text{RMSE} = \sqrt{\text{MSE}}$

**Scale-Independent:**
- $\text{MAPE} = \frac{100}{n} \sum \left| \frac{e_t}{X_t} \right|$
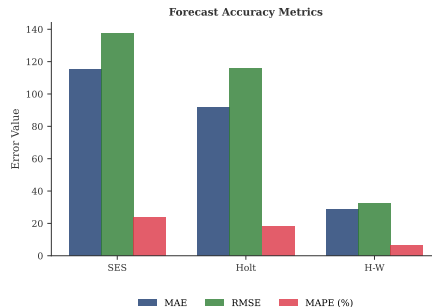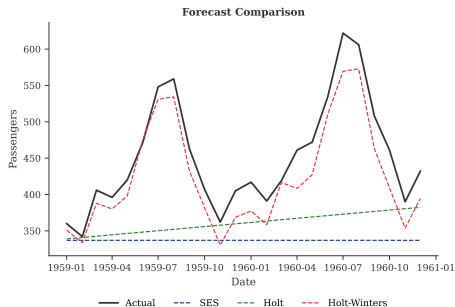- sMAPE (symmetric)

### Which to use?
- Same series: RMSE, MAE
- Compare across series: MAPE, sMAPE

# Forecast Evaluation: Visual Example
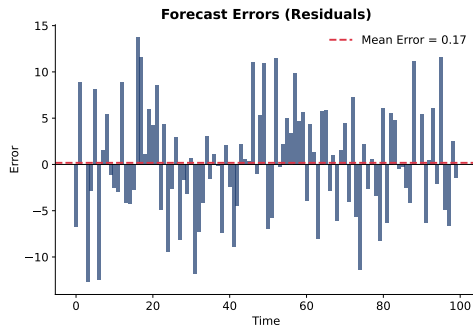


- **Top**: Actual values vs. forecasted values – visual assessment of fit
- **Bottom**: Residuals should be centered around zero with no pattern
- Good forecasts have small, random residuals with constant variance

# Comparing Forecast Methods



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

## Residual Diagnostics

**Good forecasts** should have residuals that are:

1. **Zero mean**: $\mathbb{E}[e_t] = 0$ (unbiased)
2. **Uncorrelated**: $\text{Cov}(e_t, e_{t-k}) = 0$ for $k \neq 0$
3. **Constant variance**: $\text{Var}(e_t) = \sigma^2$ (homoscedastic)
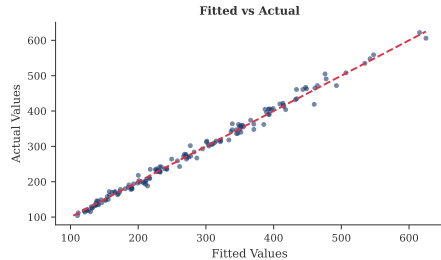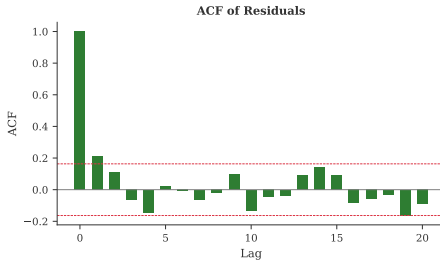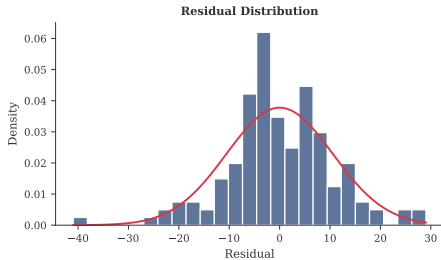4. **Normally distributed**: $e_t \sim N(0, \sigma^2)$ (for prediction intervals)
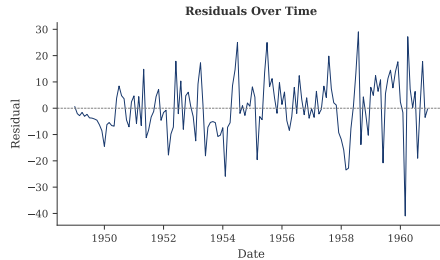
**Diagnostic tests:**

- **Ljung-Box test**: Tests for autocorrelation in residuals

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

- **Jarque-Bera test**: Tests for normality
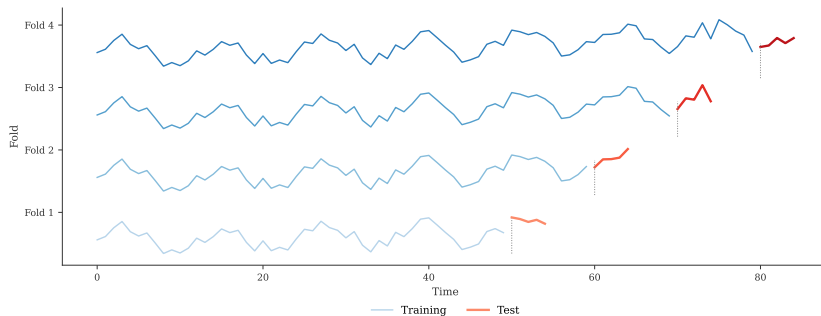
# Residual Diagnostics: Visualization

# Time Series Cross-Validation

**Standard CV** doesn't work for time series (temporal dependence).

**Rolling Origin CV:** Expanding windows

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

# Train / Validation / Test Split

**Three-way split** for model development:
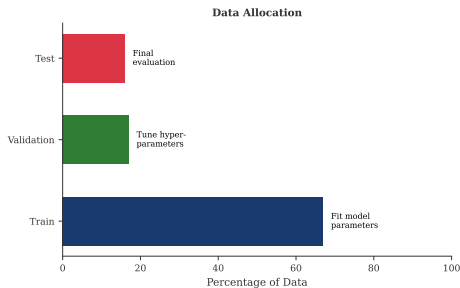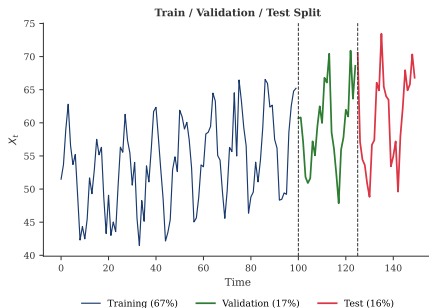
### Training Set
- Fit model parameters
- Largest portion (60–80%)
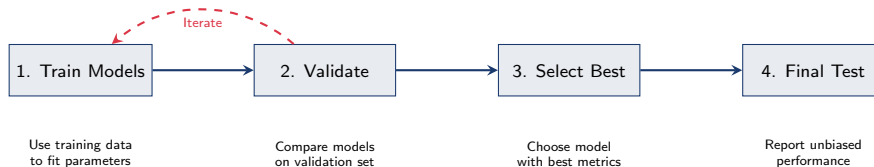- Used for estimation

### Validation Set
- Tune hyperparameters
- Compare models
- Select best approach

### Test Set
- Final evaluation only
- Never used for tuning
- Unbiased performance

# Model Development Workflow



| 1. Train Models | → | 2. Validate | → | 3. Select Best | → | 4. Final Test |
| --- | --- | --- | --- | --- | --- | --- |
| Use training data to fit parameters | | Compare models on validation set | | Choose model with best metrics | | Report unbiased performance |

**Critical Rule**

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

# Real Data: Forecast Comparison
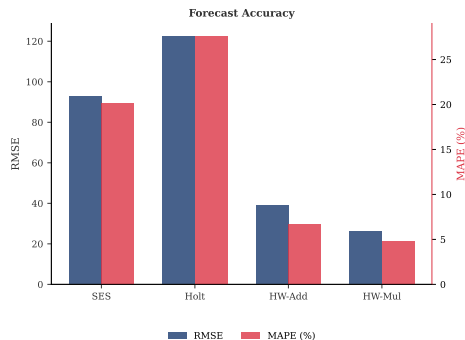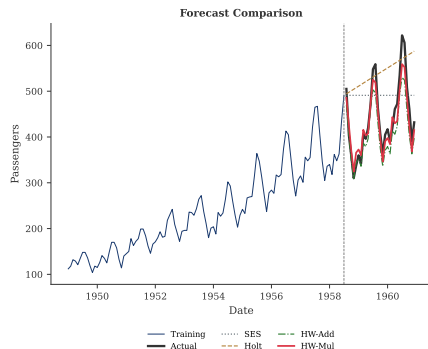


Airline passengers data: Holt-Winters Multiplicative performs best for seasonal data.

# Forecast Performance Across Datasets
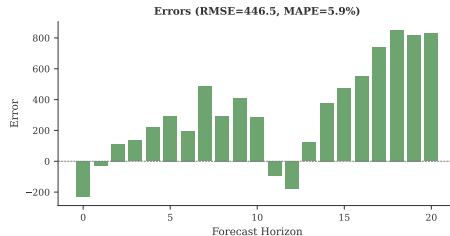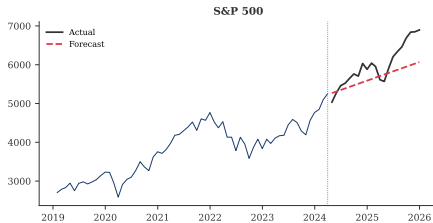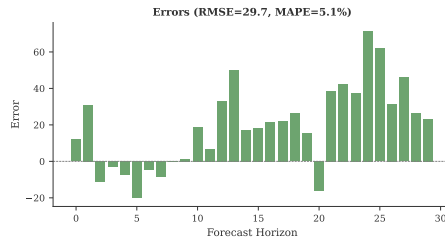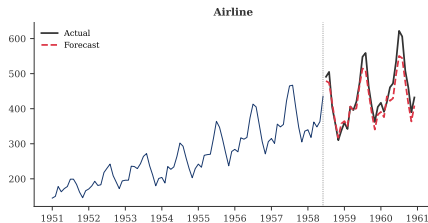


Different series require different models. Seasonal data needs seasonal methods.

## 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
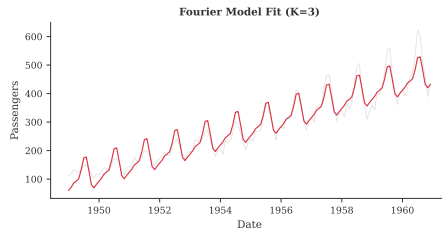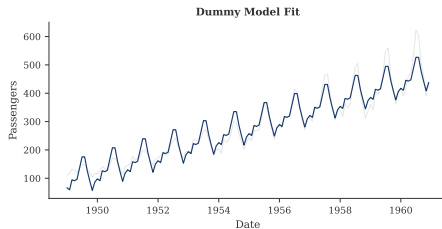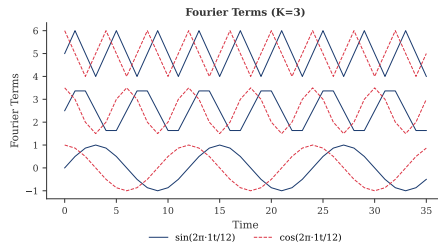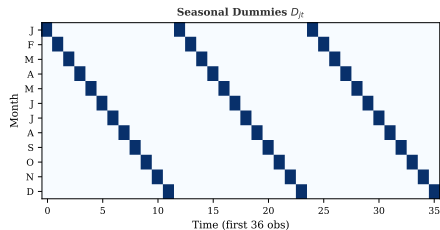- Any seasonal pattern

**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

## 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** when seasonal pattern is irregular or you need interpretable coefficients
- Use **Fourier** for smooth patterns, high-frequency seasonality (daily, hourly), or multiple seasonal periods
- **Fourier terms** are used in TBATS models 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.
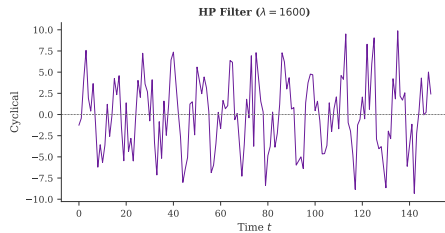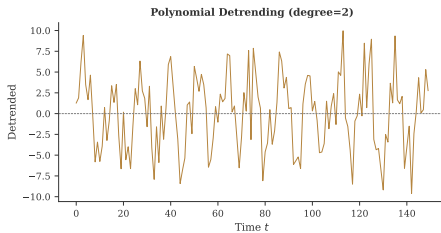
## Trend Removal Methods

**Six common detrending approaches:**

1. **Differencing**: $\Delta X_t = X_t - X_{t-1}$
2. **Linear regression**: Fit $\hat{T}_t = \hat{\beta}_0 + \hat{\beta}_1 t$
3. **Polynomial**: Fit 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
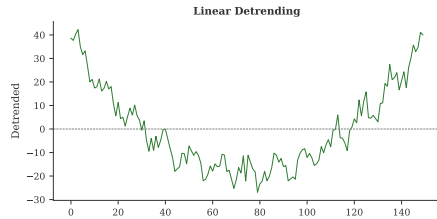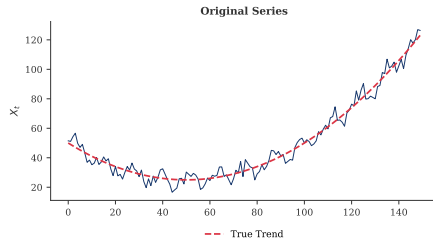
# Trend Estimation: Multiple Approaches
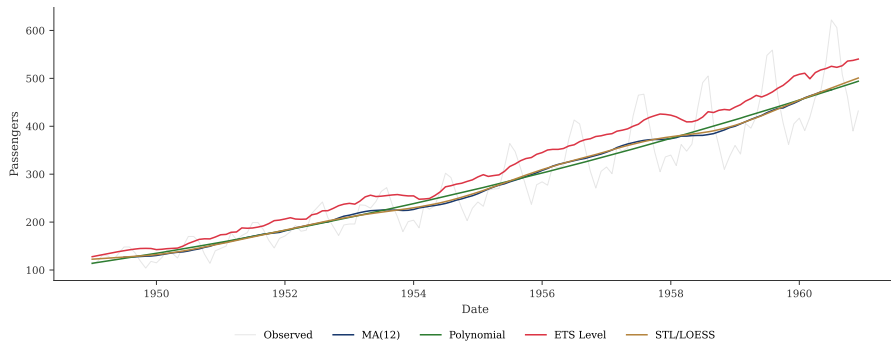


Different methods capture trend at varying levels of smoothness.

## 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



**Original Series**

**Seasonal Component $S_t$**

**Seasonally Adjusted:** $X_t/S_t$

**Seasonal Differencing:** $\Delta_{12}X_t$
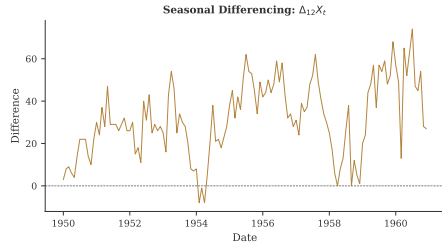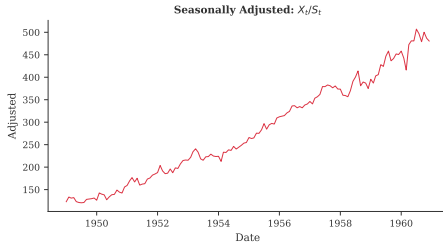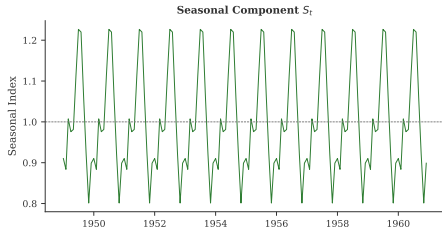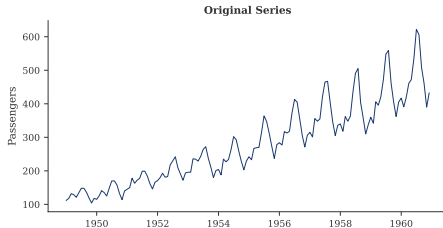
## Deterministic vs Stochastic Trend

**Deterministic Trend:**

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

- Trend is a function of time
- Detrend by regression
- $\varepsilon_t$ is stationary

**Stochastic Trend:**

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

- Random walk component
- Detrend by differencing
- $\Delta X_t$ is stationary

### Wrong Method = Problems

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

# Example: Deterministic Trend
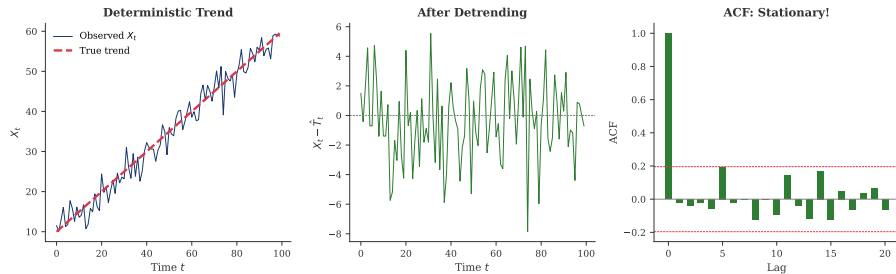


**Key:** Use regression to remove trend → residuals are stationary (ACF decays quickly).

# Example: Stochastic Trend (Random Walk)
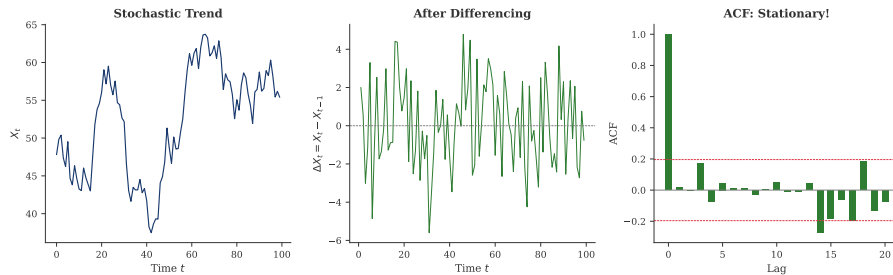


**Key:** Use differencing to remove trend → differences are stationary (white noise).

## Side-by-Side Comparison
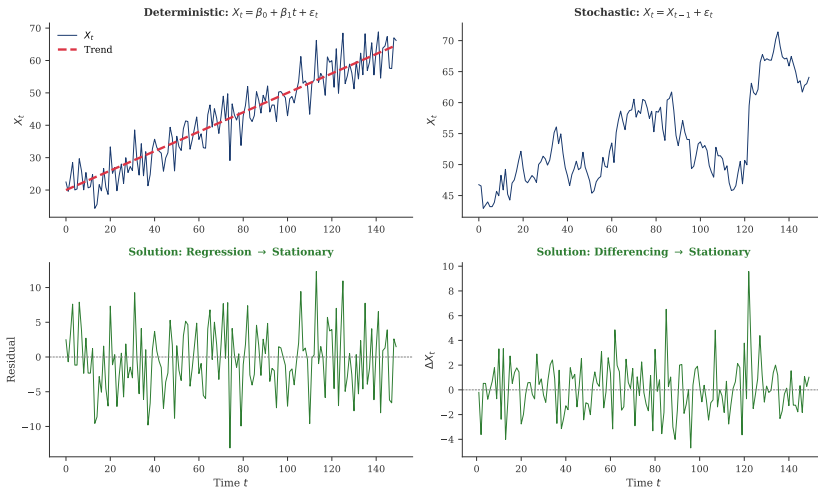


**Remember:** Deterministic → regression. Stochastic → differencing.

# Stochastic Process: Definition

## Definition 5 (Stochastic Process)

A **stochastic process** is a collection of random variables indexed by time:

$$\{X_t(\omega) : t \in \mathcal{T}, \omega \in \Omega\}$$

where $\Omega$ is the sample space of possible outcomes.

**Two perspectives:**

- **Fixed $\omega$:** A *realization* or *sample path* $\{X_t(\omega)\}_{t \in \mathcal{T}}$
- **Fixed $t$:** A *random variable $X_t$* with distribution $F_t(x)$

**Key insight:** A time series we observe is **one realization** of the underlying stochastic process. We use this single realization to infer properties of the process.

# Realizations vs Ensemble



**Multiple Realizations** — **Distribution at $t = 50$**
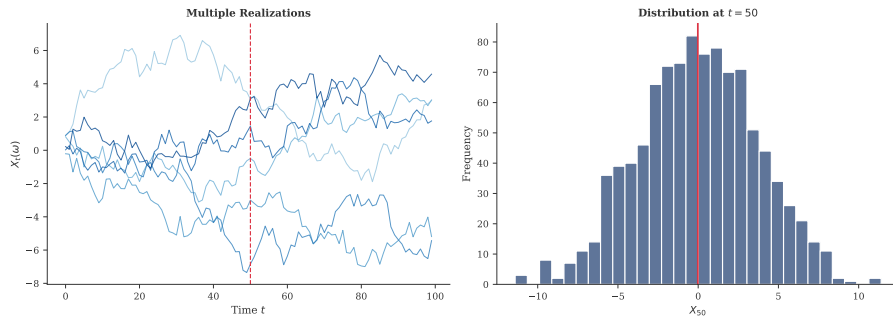
**Left:** Multiple realizations from the same process. **Right:** Distribution at fixed time $t = 50$ (ensemble average).

## Moments of a Stochastic Process

**First two moments characterize weak properties:**

**Mean Function:** $\mu_t = \mathbb{E}[X_t]$

**Autocovariance Function (ACVF):**

$$\gamma(t, s) = \text{Cov}(X_t, X_s) = \mathbb{E}[(X_t - \mu_t)(X_s - \mu_s)]$$

**Autocorrelation Function (ACF):**

$$\rho(t, s) = \frac{\gamma(t, s)}{\sqrt{\text{Var}(X_t) \cdot \text{Var}(X_s)}}$$

**Properties:** $\rho(t, s) \in [-1, 1]$ and $\rho(t, t) = 1$

# Why Stationarity Matters

**Stationarity** is a fundamental assumption for time series analysis:

**Without Stationarity:**
- Mean, variance change over time
- Past may not predict future
- Standard methods fail
- Spurious correlations

**With Stationarity:**
- Statistical properties constant
- Can estimate from one realization
- Valid inference possible
- Models are meaningful

### Key Principle

Most time series models (ARMA, ARIMA, etc.) require stationarity. Non-stationary series must be transformed (e.g., differencing) before modeling.

# Stationary vs Non-Stationary: Visual Comparison
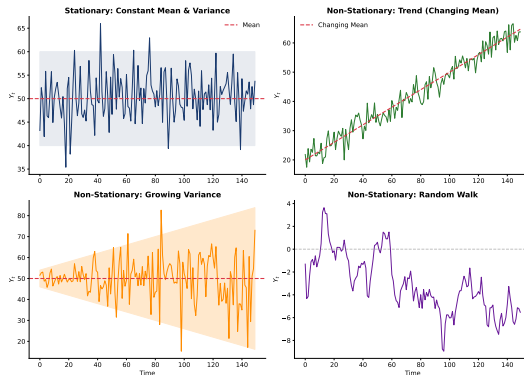


- **Stationary**: Constant mean and variance – fluctuates around a fixed level
- **Non-stationary**: Mean and/or variance change over time
- Visual inspection is the first step; formal tests (ADF, KPSS) confirm

# Strict Stationarity

## Definition 6 (Strict (Strong) Stationarity)

A process $\{X_t\}$ is **strictly stationary** if for all $k$, all $t_1, \ldots, t_k$, and all $h$:

$$(X_{t_1}, X_{t_2}, \ldots, X_{t_k}) \stackrel{d}{=} (X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_k+h})$$

**Interpretation:** The joint distribution of any collection of observations is **invariant to time shifts**.

**Implications:**

- All marginal distributions $F_{X_t}(x)$ are identical
- $\mathbb{E}[X_t] = \mu$ (constant mean)
- $\text{Var}(X_t) = \sigma^2$ (constant variance)
- Joint distributions depend only on time *differences*

**Note:** Strict stationarity is a strong condition, often impractical to verify.

# Weak (Covariance) Stationarity

---

### Definition 7 (Weak Stationarity)

A process $\{X_t\}$ is **weakly stationary** (or covariance 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 $h$)

---

**Key property:** Autocovariance is a function of lag only:

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

**Autocorrelation function:**

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{\text{Cov}(X_t, X_{t+h})}{\text{Var}(X_t)}$$

Note: $\rho(0) = 1$ and $\rho(h) = \rho(-h)$ (symmetry)

## Properties of the Autocovariance Function

For a weakly stationary process, the ACVF $\gamma(h)$ satisfies:

1. **Symmetry:** $\gamma(h) = \gamma(-h)$

2. **Maximum at zero:** $|\gamma(h)| \leq \gamma(0)$

3. **Non-negative definiteness**

**Implication:** Not every function can be an autocovariance function.

# White Noise Process

## Definition 8 (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 $t \neq s$
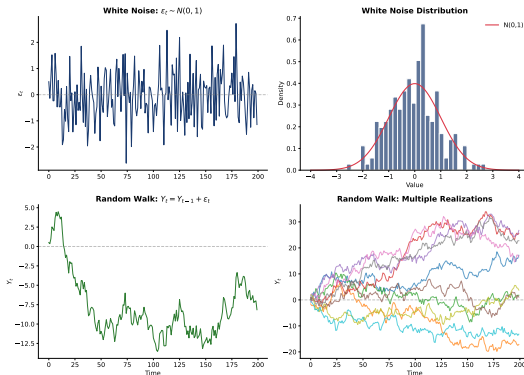
**ACF of White Noise:**

$$\rho(h) = \begin{cases} 1 & \text{if } h = 0 \\ 0 & \text{if } h \neq 0 \end{cases}$$

**Types:**

- **Weak white noise**: Uncorrelated (conditions above)
- **Strong white noise**: Independent and identically distributed (i.i.d.)
- **Gaussian white noise**: $\varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$

# White Noise vs Random Walk: Comparison



- **White noise**: Fluctuates around zero – stationary, constant variance
- **Random walk**: Cumulative sum of white noise – wanders away, non-stationary
- Random walk is the simplest non-stationary process (unit root)

# White Noise: Visualization
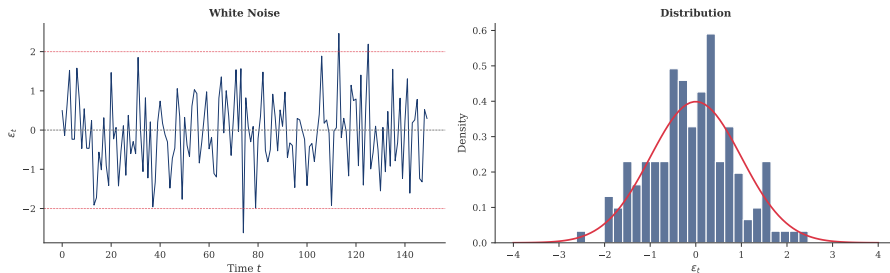


White noise is the building block of time series models. It represents unpredictable random shocks.

# Random Walk Process

**Definition:** $X_t = X_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim WN(0, \sigma^2)$, $X_0 = 0$

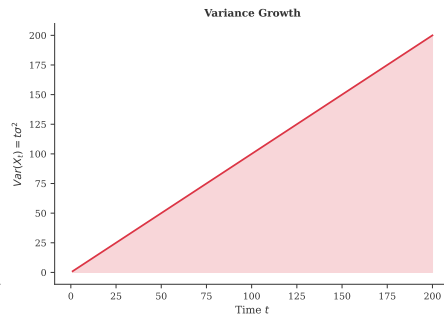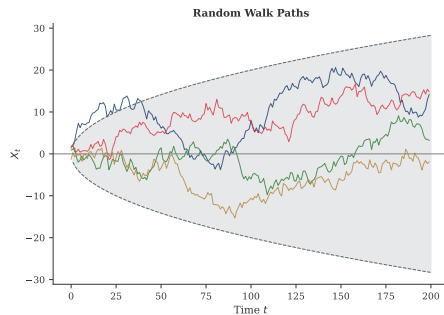**Explicit form:** $X_t = \sum_{i=1}^{t} \varepsilon_i$

**Properties:**

- $\mathbb{E}[X_t] = 0$ (constant mean)
- $\text{Var}(X_t) = t\sigma^2$ (variance grows with time!)
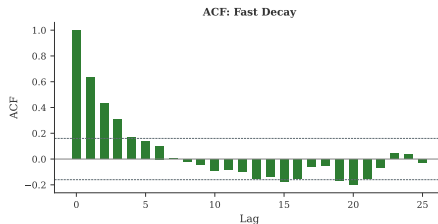- $\text{Cov}(X_t, X_s) = \min(t, s) \cdot \sigma^2$

## Non-Stationary!

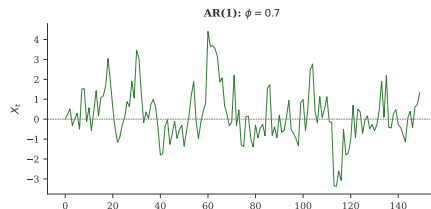Random walk is **not stationary** because variance depends on $t$.

# Random Walk: Visualization



**Random Walk Paths** / **Variance Growth**

**Left:** Multiple paths diverge over time. **Right:** Variance grows linearly: $\text{Var}(X_t) = t\sigma^2$.

# Stationary vs Non-Stationary: Comparison
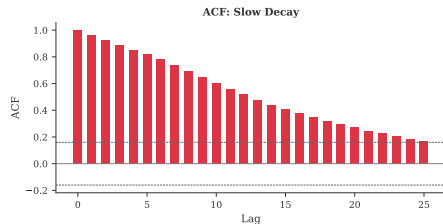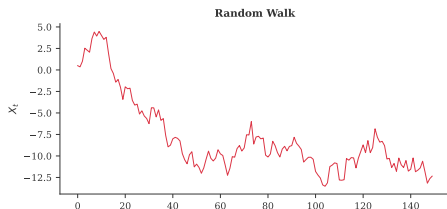


**Key diagnostic:** ACF of stationary process decays quickly; ACF of random walk decays very slowly.

## Sample Autocorrelation Function

**Sample ACF at lag $h$:**

$$\hat{\rho}(h) = \frac{\sum_{t=1}^{T-h}(x_t - \bar{x})(x_{t+h} - \bar{x})}{\sum_{t=1}^{T}(x_t - \bar{x})^2}$$
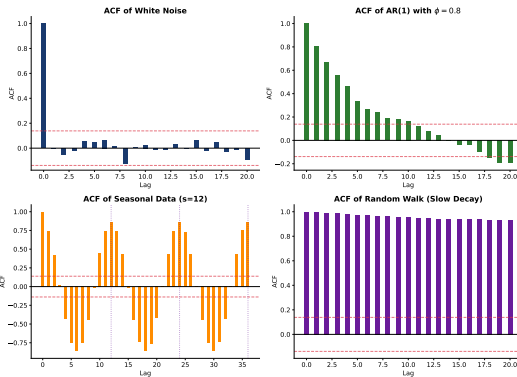
**Properties:**

- $\hat{\rho}(0) = 1$ always
- $|\hat{\rho}(h)| \leq 1$

**Significance test:** Under white noise, $\hat{\rho}(h) \approx N(0, 1/T)$

**95% bounds:** $\pm 1.96/\sqrt{T}$

# ACF Patterns for Different Processes



- **White noise**: ACF drops to zero immediately (no dependence)
- **AR(1)**: ACF decays exponentially – indicates autoregressive structure
- **Seasonal**: ACF shows spikes at seasonal lags (e.g., 12, 24 for monthly)
- **Random walk**: ACF decays very slowly – sign of non-stationarity

# Partial Autocorrelation Function (PACF)

**PACF** $\phi_{hh}$: Correlation between $X_t$ and $X_{t+h}$ after removing the linear effect of $X_{t+1}, \ldots, X_{t+h-1}$.
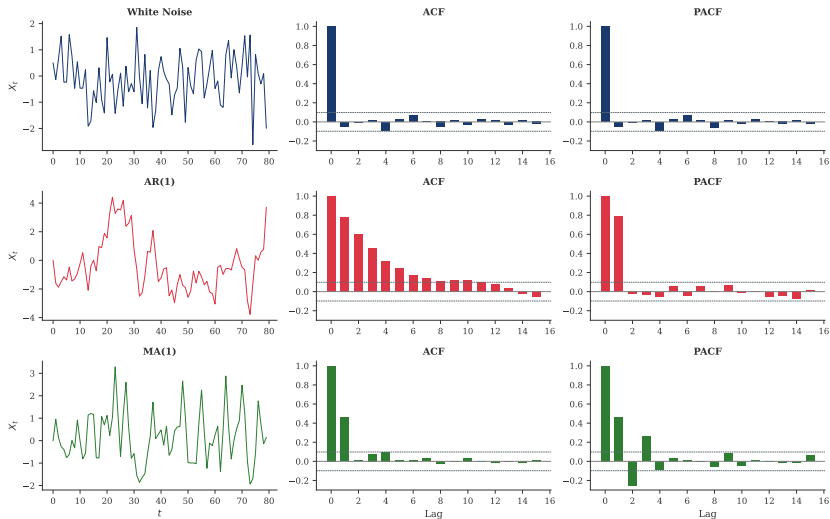
**Interpretation:**
- $\phi_{11} = \rho(1)$ (same as ACF at lag 1)
- $\phi_{22} =$ correlation of $X_t, X_{t+2}$ controlling for $X_{t+1}$
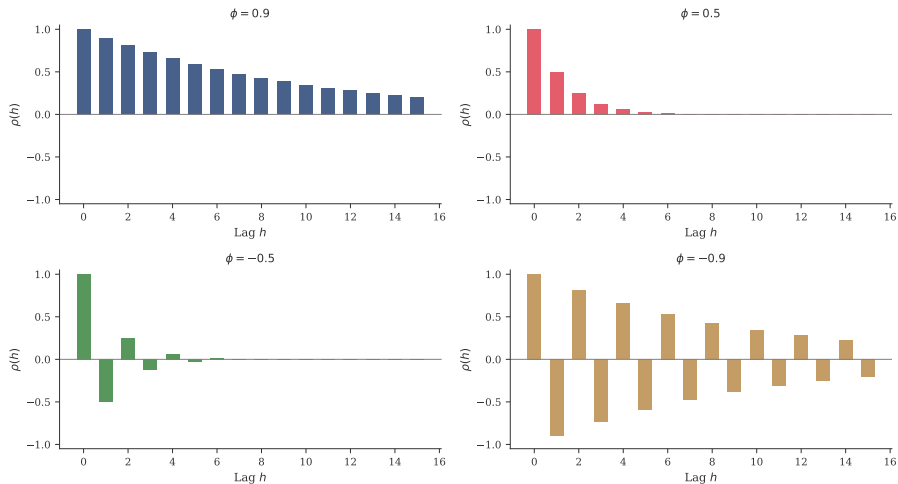- Measures *direct* dependence at lag $h$

**Key application:** Identify AR order
- For AR($p$): PACF **cuts off** after lag $p$
- For MA($q$): ACF **cuts off** after lag $q$

# ACF and PACF Patterns

# Theoretical ACF for AR(1)



For AR(1): $X_t = \phi X_{t-1} + \varepsilon_t$, the theoretical ACF is $\rho(h) = \phi^h$.

## The Lag Operator

### Definition 9 (Lag Operator)

The **lag operator** (or backshift operator) $L$ is defined by:

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

**Properties:**

- $L^k X_t = X_{t-k}$ (lag by $k$ periods)
- $L^0 = I$ (identity)
- $(1 - \phi L)X_t = X_t - \phi X_{t-1}$

**Examples:**

- AR(1): $(1 - \phi L)X_t = \varepsilon_t$
- MA(1): $X_t = (1 + \theta L)\varepsilon_t$
- AR($p$): $(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)X_t = \varepsilon_t$

## Differencing

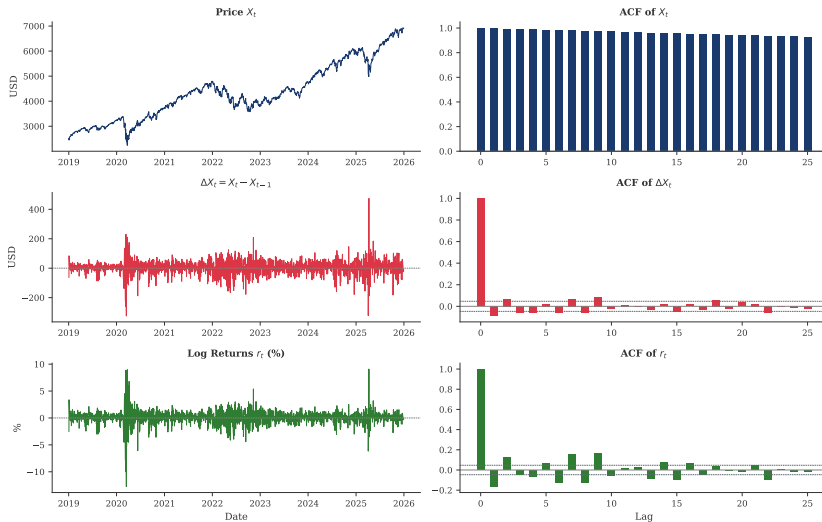**First difference:** $\Delta X_t = X_t - X_{t-1} = (1 - L)X_t$

**Why difference?**

- Removes trend and unit root
- Random walk: $\Delta X_t = \varepsilon_t$ (white noise)

**Integrated process:** $X_t \sim I(d)$ if $\Delta^d X_t$ is stationary

- $I(0)$: Stationary (no differencing needed)
- $I(1)$: One difference needed
- $I(2)$: Two differences needed

# Effect of Differencing: S&P 500

## Augmented Dickey-Fuller (ADF) Test

**Model:** $\Delta X_t = \alpha + \gamma X_{t-1} + \sum_{i=1}^{p} \delta_i \Delta X_{t-i} + \varepsilon_t$

**Hypotheses:**

- $H_0$: $\gamma = 0$ (unit root)
- $H_1$: $\gamma < 0$ (stationary)

**Decision:**

- $\tau <$ critical value $\Rightarrow$ Reject $H_0$ $\Rightarrow$ Stationary
- $\tau \geq$ critical value $\Rightarrow$ Non-stationary

Critical values: Dickey-Fuller distribution (not normal)

**Test statistic:**

$$\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})}$$

# KPSS Test

**Model:** $X_t = \xi t + r_t + \varepsilon_t$ where $r_t = r_{t-1} + u_t$

**Hypotheses (opposite of ADF):**

- $H_0$: $\sigma_u^2 = 0$ (stationary)
- $H_1$: $\sigma_u^2 > 0$ (unit root)

**Test statistic:**

$$LM = \frac{\sum_{t=1}^{T} S_t^2}{T^2 \hat{\sigma}^2}$$

where $S_t = \sum_{i=1}^{t} \hat{e}_i$

**Decision:**

- $LM >$ critical value $\Rightarrow$ Reject $H_0$ $\Rightarrow$ Non-stationary
- $LM \leq$ critical value $\Rightarrow$ Stationary

**Note:** KPSS complements ADF—use both for robust conclusions.

## Using ADF and KPSS Together

**Confirmatory testing** for robust conclusions:

| ADF | KPSS | Conclusion |
|-----|------|------------|
| Reject $H_0$ | Fail to reject $H_0$ | Stationary |
| Fail to reject $H_0$ | Reject $H_0$ | Unit Root |
| Reject $H_0$ | Reject $H_0$ | Inconclusive |
| Fail to reject $H_0$ | Fail to reject $H_0$ | Inconclusive |

**Recommended workflow:**

1. Run ADF test (null = unit root)
2. Run KPSS test (null = stationary)
3. If results agree, proceed with confidence
4. If inconclusive, consider alternative tests (PP, DF-GLS)

# ADF Test: Visualization with S&P 500
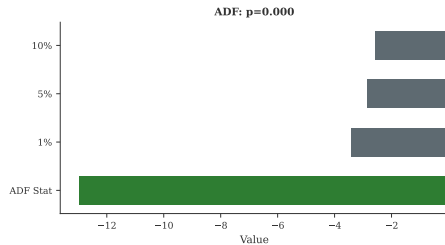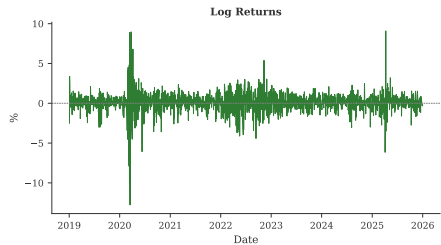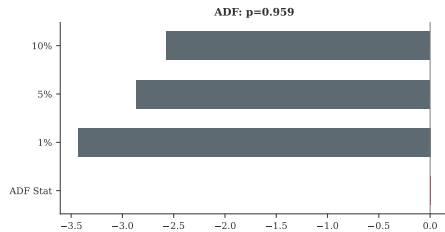
# S&P 500 Analysis: Overview
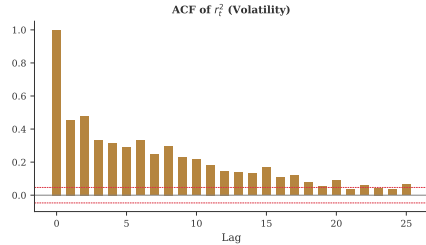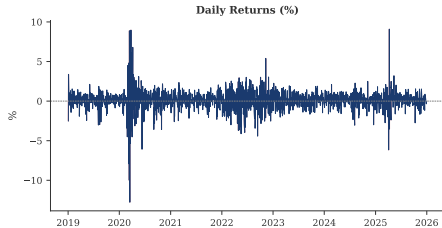
# Stylized Facts of Financial Returns



Distribution

$\mu$=0.057%
$\sigma$=1.260%
Skew=-0.67
Kurt=15.5

Q-Q Plot
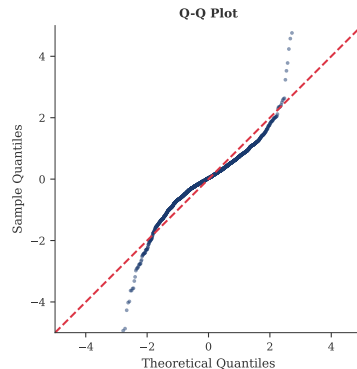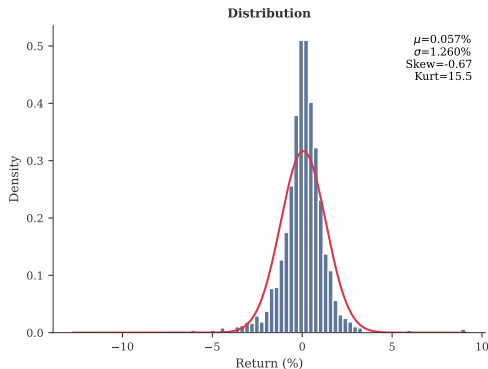
**Observed properties:**

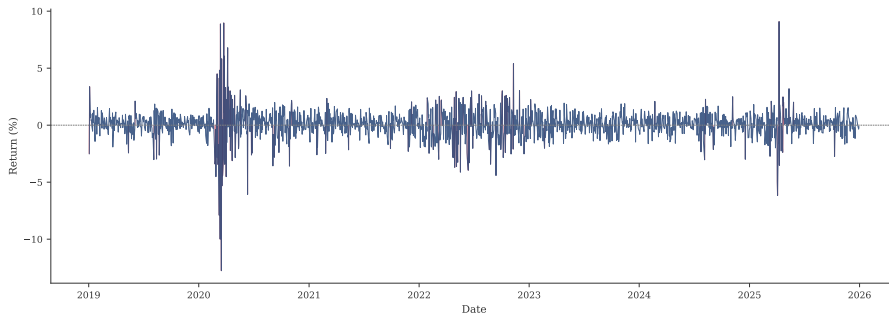- Negative skewness (left tail)
- Excess kurtosis ($\gg 3$)
- Heavy tails (fat tails)

**Implications:**

- Normal distribution inadequate
- Extreme events more likely
- Need Student-t or similar

# Volatility Clustering



---

**Stylized Fact**

Large returns (positive or negative) tend to be followed by large returns. This **volatility clustering** motivates ARCH/GARCH models (future chapters).

# Key Takeaways

1. **Time series** $=$ observations indexed by time with temporal dependence
2. **Decomposition**: Additive $X_t = T_t + S_t + \varepsilon_t$ or Multiplicative
3. **Exponential Smoothing**: SES (level), Holt (trend), Holt-Winters (seasonal)
4. **Forecast Evaluation**: MAE, RMSE, MAPE; use train/validation/test splits
5. **Seasonality Modeling**: Dummy variables (any pattern) or Fourier terms (smooth)
6. **Trend Handling**: Differencing (stochastic) or regression (deterministic)
7. **Stationarity**: Mean, variance, autocovariance constant over time
8. **ACF/PACF**: Essential for identifying dependence structure
9. **Unit root tests**: ADF ($H_0$: unit root) vs KPSS ($H_0$: stationary)

## Important Formulas I

### Decomposition

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

### Simple Exponential Smoothing (SES)

$\hat{X}_{t+1|t} = \alpha X_t + (1-\alpha)\hat{X}_{t|t-1}$    where $\alpha \in (0,1)$

### Holt's Linear Trend

$\ell_t = \alpha X_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$    $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$

### Holt-Winters Additive

$\ell_t = \alpha(X_t - S_{t-s}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$    $S_t = \gamma(X_t - \ell_t) + (1-\gamma)S_{t-s}$

# Important Formulas II

## Moving Average (Trend Estimation)

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

## Autocovariance and Autocorrelation

$\gamma(h) = \text{Cov}(X_t, X_{t+h})$ $\qquad \rho(h) = \frac{\gamma(h)}{\gamma(0)}$

## Random Walk

$X_t = X_{t-1} + \varepsilon_t \quad \Rightarrow \quad \text{Var}(X_t) = t\sigma^2$ (non-stationary)

## Differencing

$\Delta X_t = (1 - L)X_t = X_t - X_{t-1}$

## Next Chapter Preview

**Chapter 2: ARMA Models**

- Autoregressive (AR) models
- Moving Average (MA) models
- Combined ARMA models
- Model identification using ACF/PACF
- Parameter estimation
- Model diagnostics
- Forecasting

# Thank You!

Questions?

All charts generated using Python (yfinance, statsmodels, matplotlib)

Data source: Yahoo Finance (2019–2025)