KI Service Zentrum
by Hasso-Plattner-Institut

HPI

Gefördert durch:

Bundesministerium für Forschung, Technologie und Raumfahrt

Time Series Forecasting

2.2 Preparing the Target

Mario Tormo Romero

Design IT.
Create Knowledge.

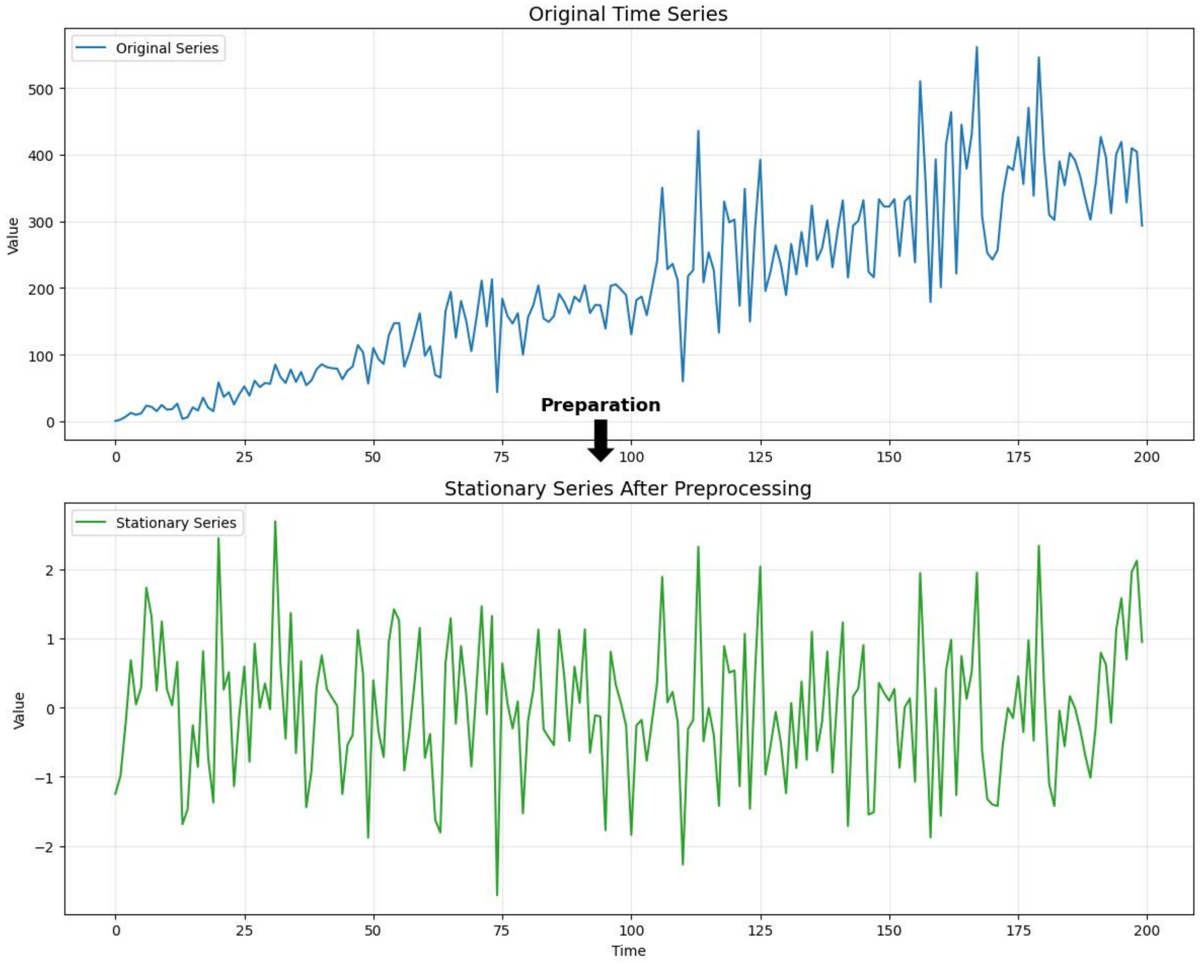
www.hpi.de

What we‘ll cover in this video

- Why preparing the target is crucial for forecasting accuracy
- Checking and fixing non-stationarity
  - ADF test
  - Differencing transforms
  - Variance stabilization (log, Box-Cox)
- Detecting and handling trends
  - Deterministic vs stochastic
  - Mann-Kendall trend test
  - Detrending methods
- Detecting and handling seasonality
  - STL decomposition

Why prepare the target?

- Raw time series often break model assumptions
- Common issues:
    - Non-stationarity (changing mean over time)
    - Trends and seasonality
    - Heteroscedasticity (changing variance)
- Poor preparation → biased forecasts, lower accuracy
- **Goal**: Make the target variable stable, clean, and predictable



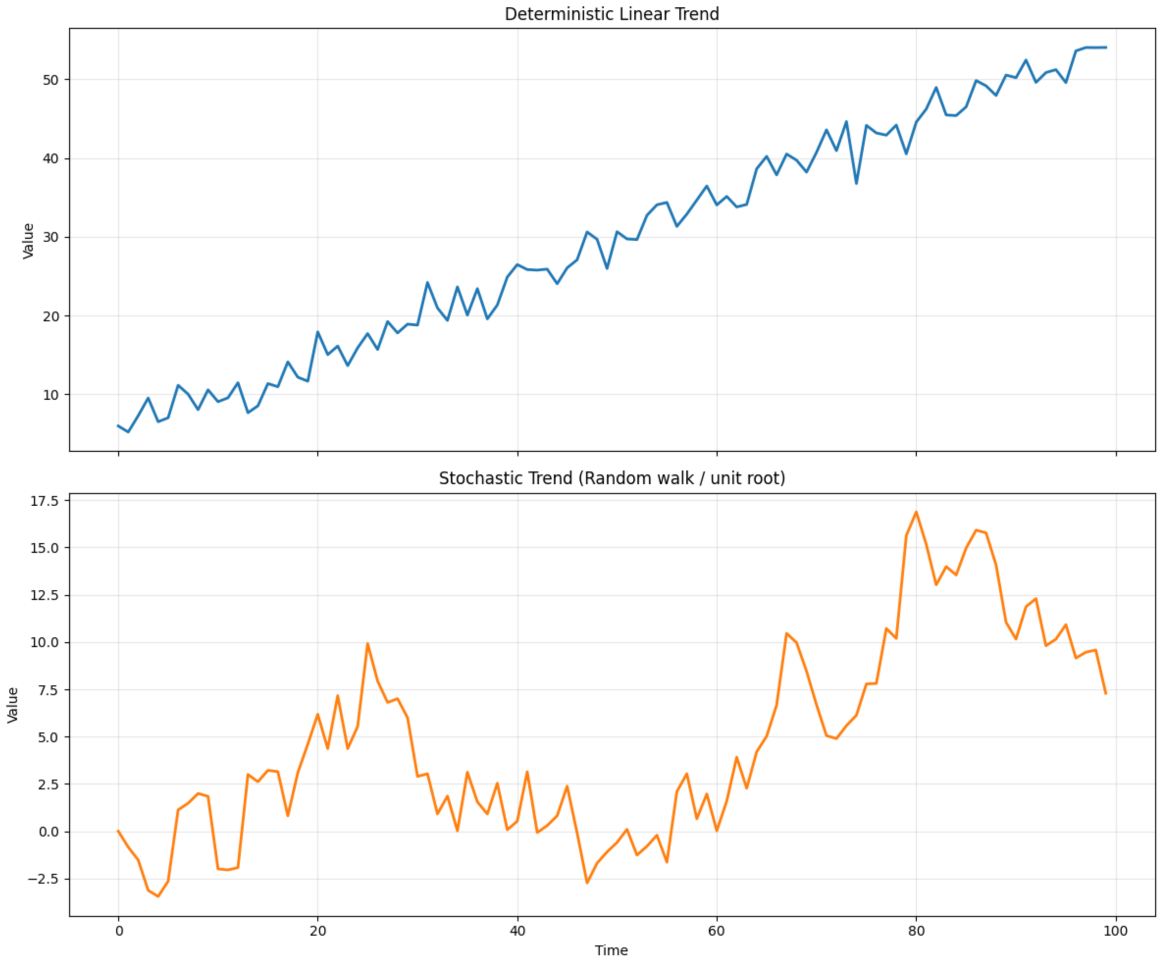
Stationarity

Types of trends

- Trends are long-term movements in the series

- Two main types:
  - Deterministic trend: predictable over time (e.g., straight line or known curve)
  - Stochastic trend: evolves randomly, often following a random walk

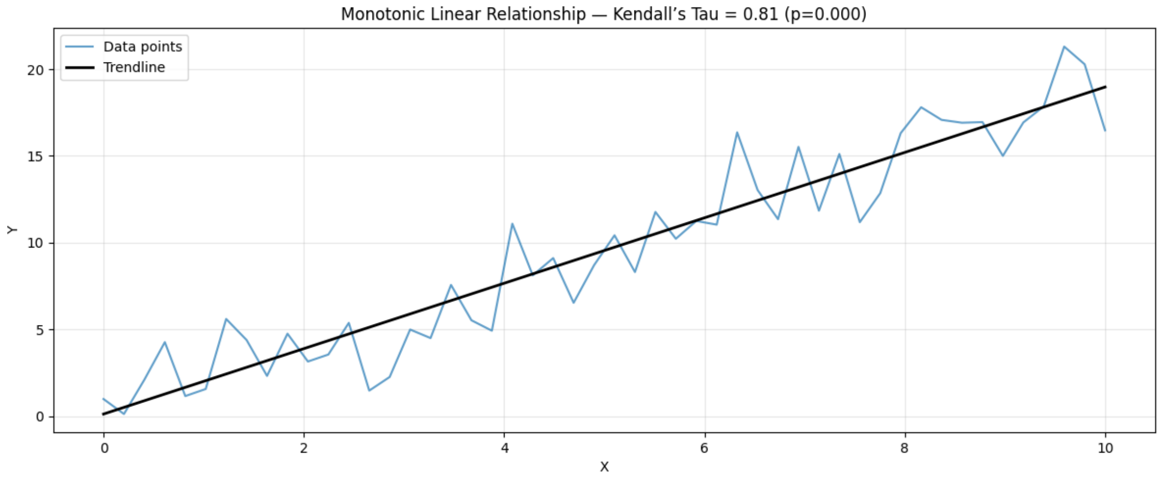
- Different trends require different treatments (detrending methods)

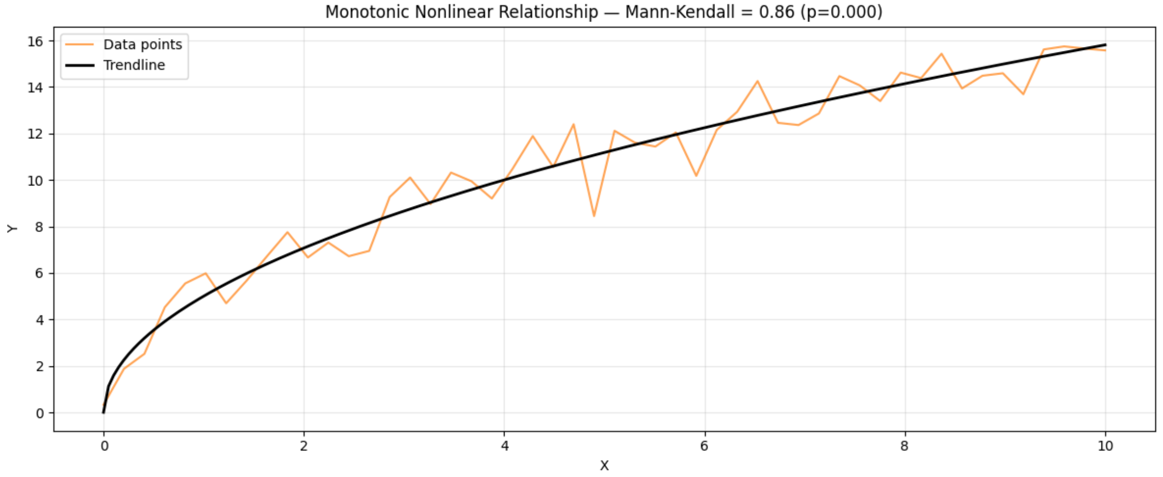


Stationarity

Trend detection tests

- **Kendall's Tau**: Measures strength of monotonic relationship between time and the variable
  - Value close to +1 → strong upward trend
  - Value close to −1 → strong downward trend
- **Mann–Kendall (MK) Test**: Statistical test for trend detection
  - **H_0**: No trend present
  - **H_1**: Trend exists (upward or downward)
  - Works well for non-linear trends and missing values

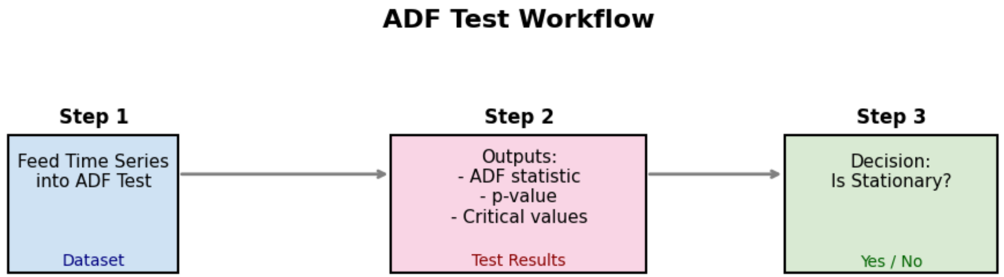


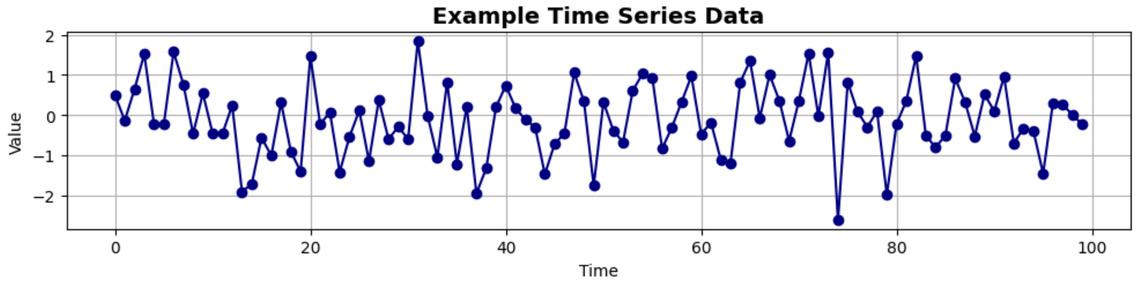


Stationarity

Checking for Unit Roots – ADF Test

- Purpose: Detect non-stationarity in a time series
- Augmented Dickey–Fuller (ADF) Test:
  - **H_0**: Unit root exists → series is non-stationary
  - **H_1**: No unit root → series is stationary
- Interpretation:
  - p-value < 0.05 → reject H_0 → stationary
  - p-value ≥ 0.05 → fail to reject H_0 → non-stationary
- Often followed by differencing or transformations if needed





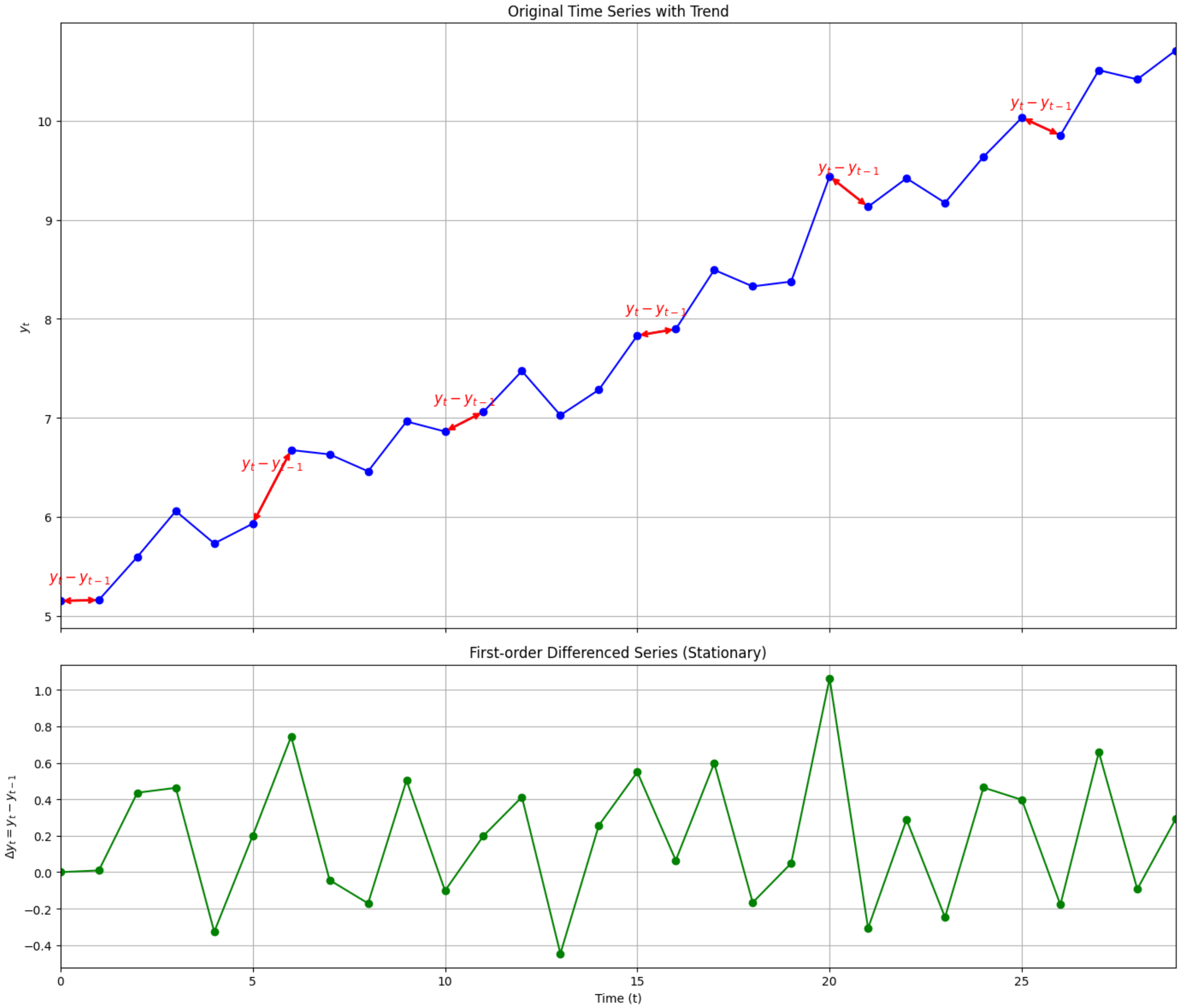
ADF Test Result

ADF Statistic	p-value	Critical Value (5%)	Decision
-3.25	0.015	-2.90	Stationary

Stationarity

Differencing Transform

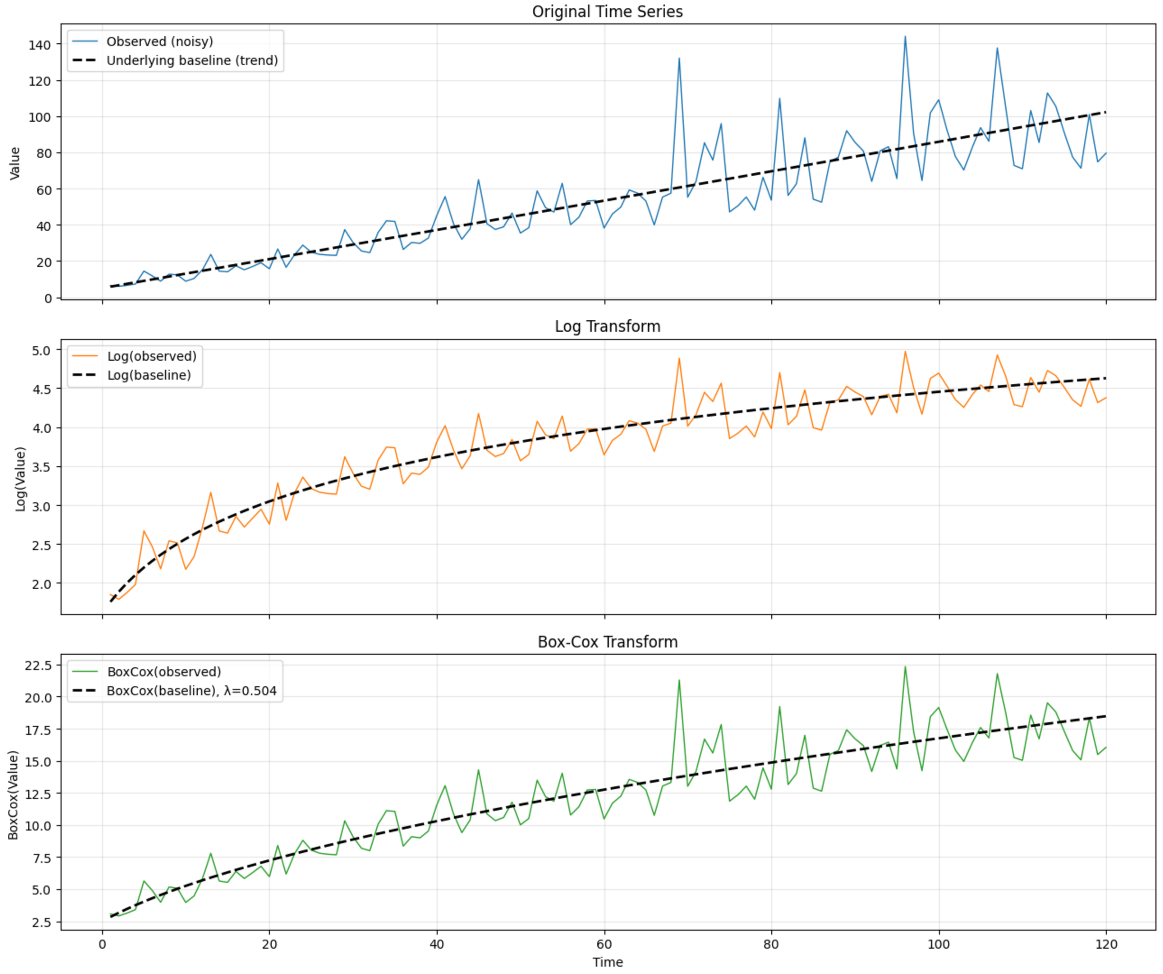
- Purpose: Remove trends and stabilize the mean
- **Additive differencing**: $y'_t = y_t - y_{t-1}$
- **Multiplicative differencing**: $y'_t = \frac{y_t}{y_{t-1}} - 1$
- Apply differencing once or multiple times until stationarity is reached
- Often the first step after ADF test signals non-stationarity



Stationarity

Stabilizing the Variance

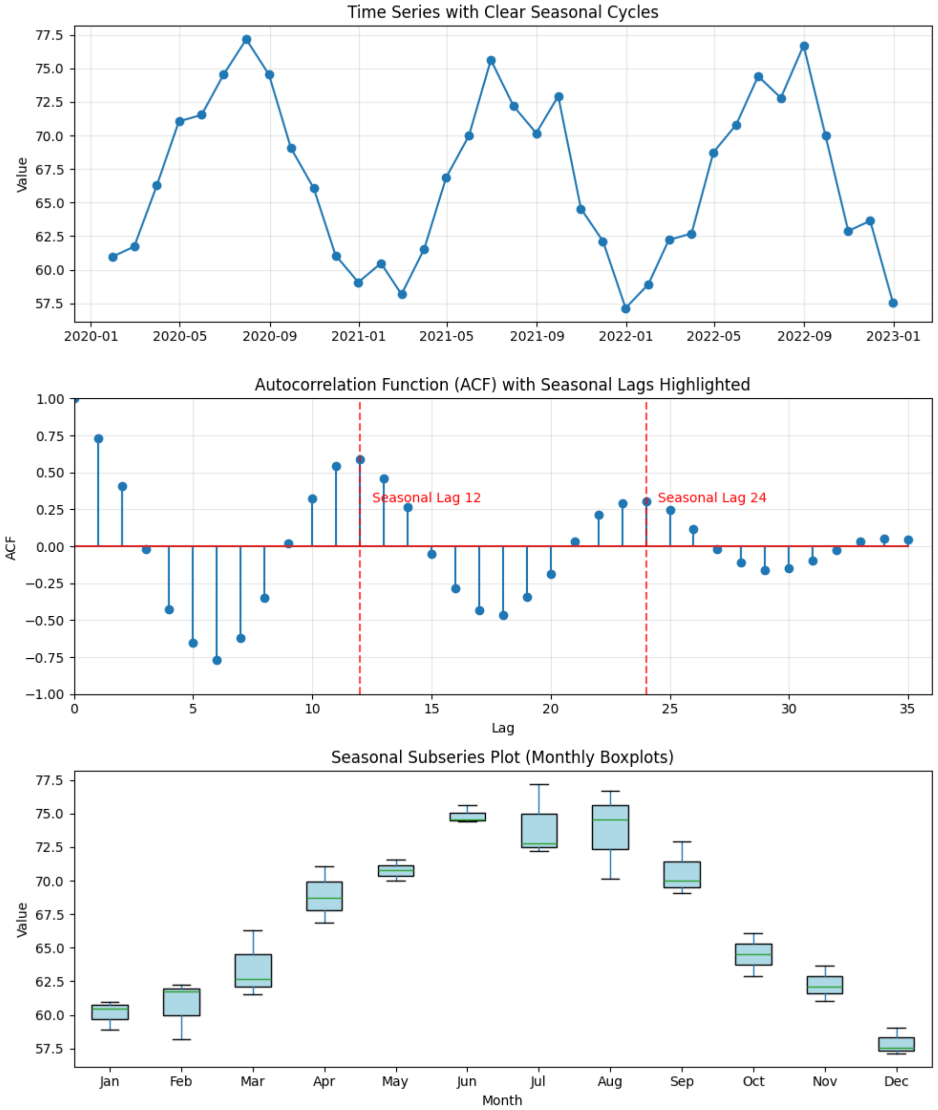
- Heteroscedasticity → changing variance over time
- Common test: **White Test** to detect heteroscedasticity
- Variance stabilizing transforms:
  - **Log transform** (for positive values)
  - **Box-Cox transformation** (flexible power transform)
- Goal: Make variance more constant for better forecasting



Seasonality

Detecting seasonality

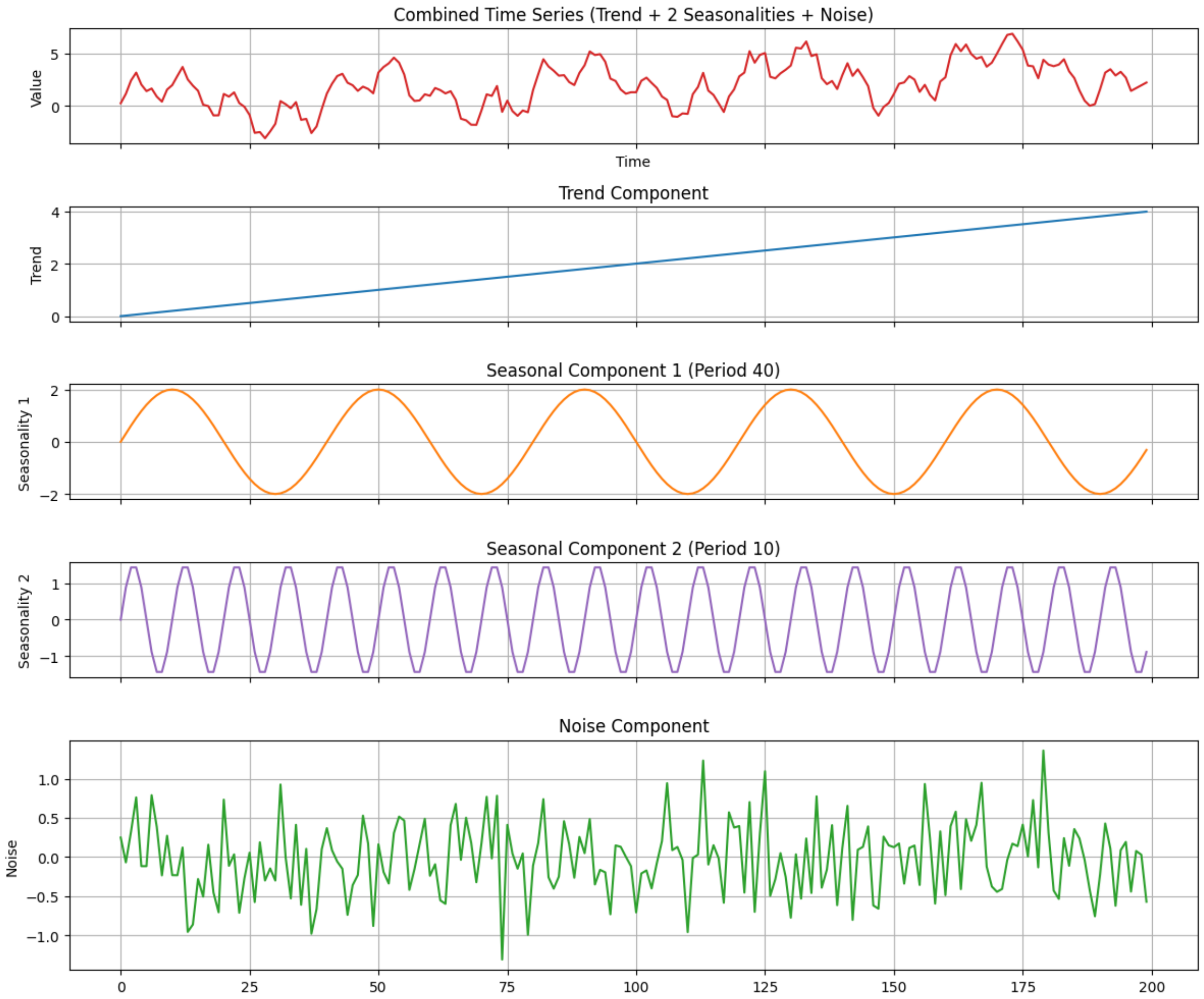
- Common detection methods:
  - Visual inspection of plots
  - Autocorrelation Function (ACF) to spot repeating patterns
  - Seasonal subseries plots
- Statistical tests to detect seasonality:
  - **Kruskal-Wallis test**: checks for differences between seasonal groups
  - **Friedman test**: for related samples across seasons
  - **Edwards' test**: detects cyclical patterns, especially in event data



STL Decomposition

Breaking Down the Series

- STL decomposes a time series into three components:
  - Trend
  - Seasonal
  - Residual (noise)
- Uses Loess smoothing, making it very flexible and robust
- Works well with complex seasonal patterns and missing data
- Allows combining with forecasting methods like Naive or ARIMA on components



What we‘ve learnt

- Stationarity matters — trends and seasonality can mislead models if not addressed
- Differencing helps remove trends (additive or multiplicative)
- Log and Box-Cox transforms help stabilize variance (hello again, TBATS & BATS!)
- Trends can be detected with statistical tests like Kendall’s Tau & M-K test
- Seasonality can be spotted visually, with ACF, subseries plots, and tests like Kruskal-Wallis, Friedman, and Edwards’
- STL decomposition is a powerful tool to break down, clean, and prepare series for modeling