A Practical Introduction to Data Science

Part 6

Time Series Analysis

Gergely Zsombor Haász

haasz.zsombi@gmail.com

Course Agenda

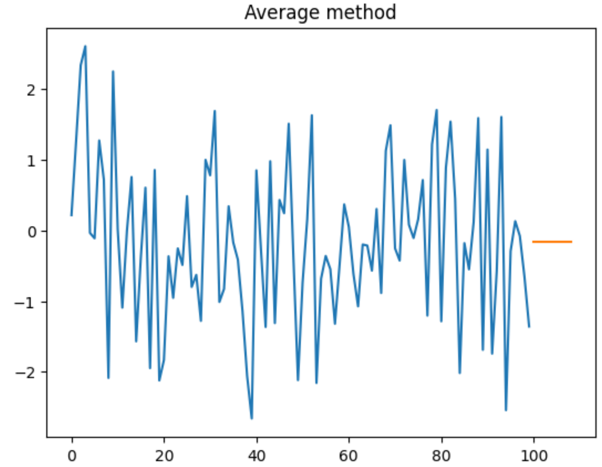
Time Series Analysis

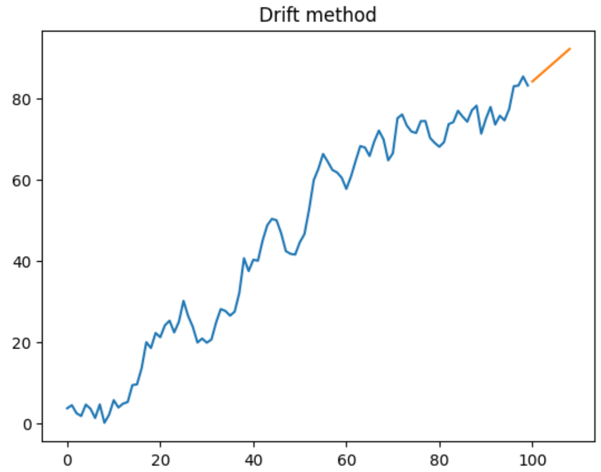
- Simple Forecasting Methods
- Time Series Decomposition
- Statistical Modelling
- Machine Learning

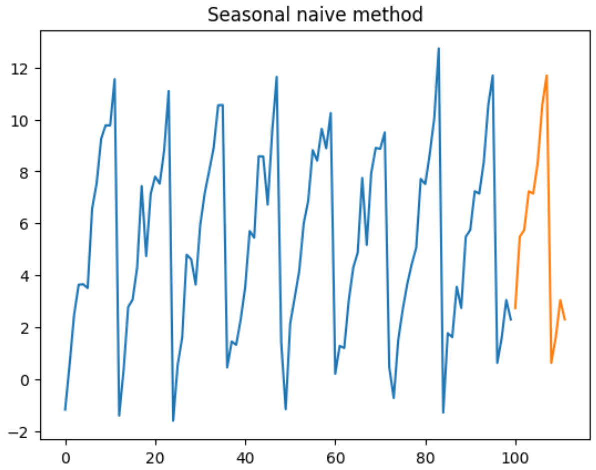
Simple Forecasting Methods

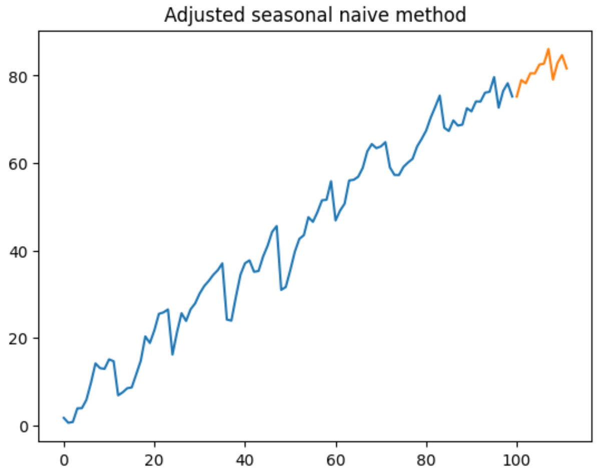
Simple Forecasting Methods

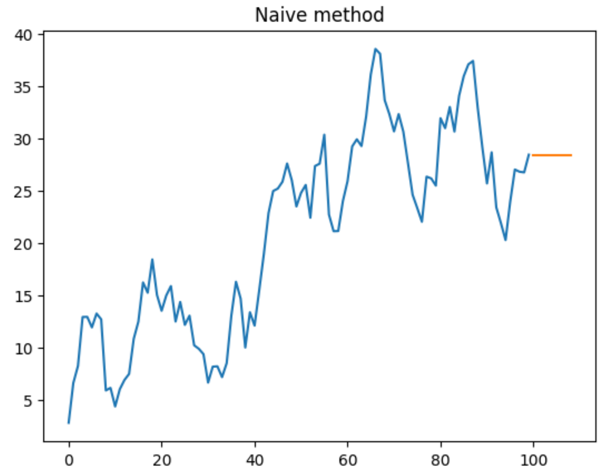
Average method	$\hat{y}_{T+h} = \bar{y}$
Naive method	$\hat{y}_{T+h} = y_T$
Seasonal naive method	$\hat{y}_{T+h} = y_{T+h-m}$
Drift method	$\hat{y}_{T+h} = y_T + h * \left(\frac{y_T - y_1}{T-1}\right)$







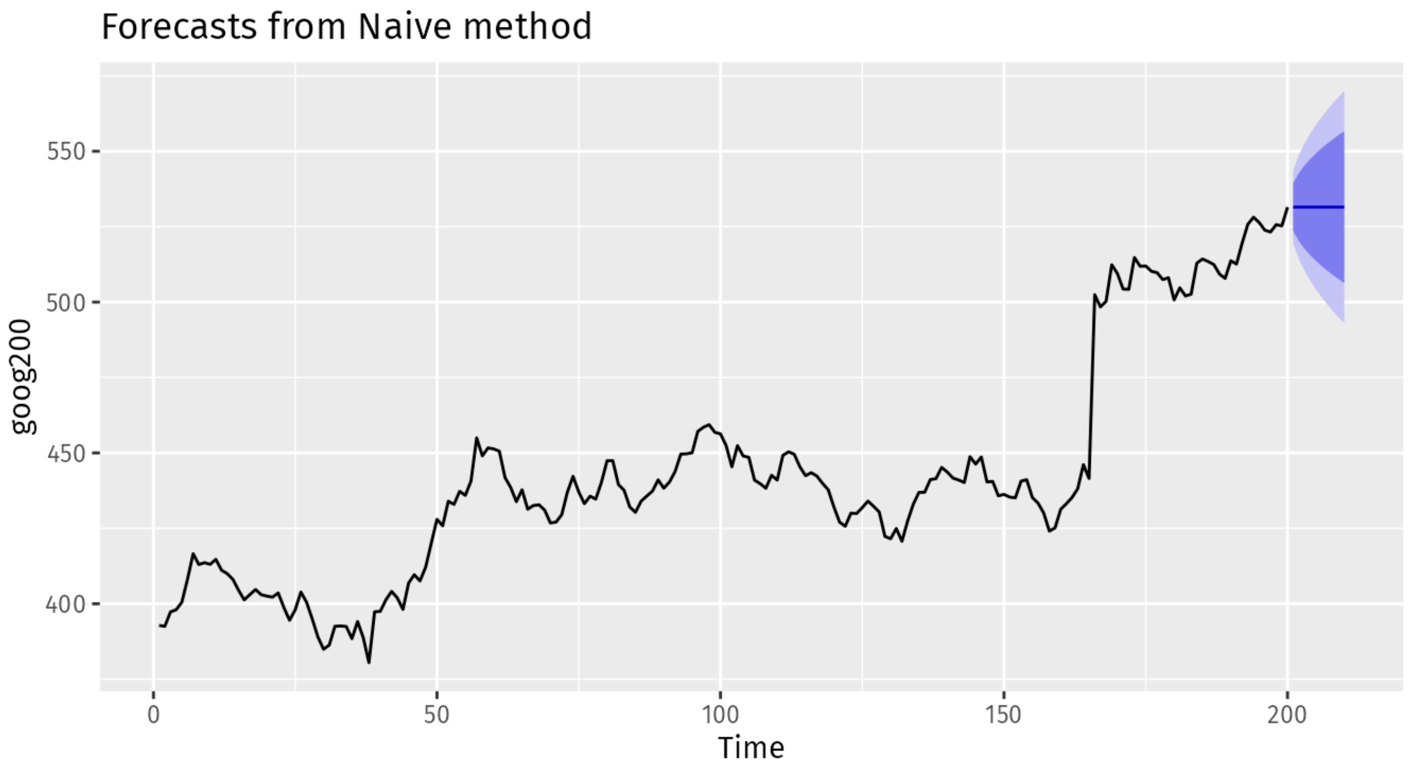




Prediction intervals

- A range of values that the random variable could take with relatively high probability
- Assuming that the forecast errors are normally distributed, a 95% prediction interval for the h-step forecast is

$$\hat{y}_{T+h} \pm 1.96 * \hat{\sigma}_h$$



Source: [Forecasting: Principles and Practice (2nd ed)]

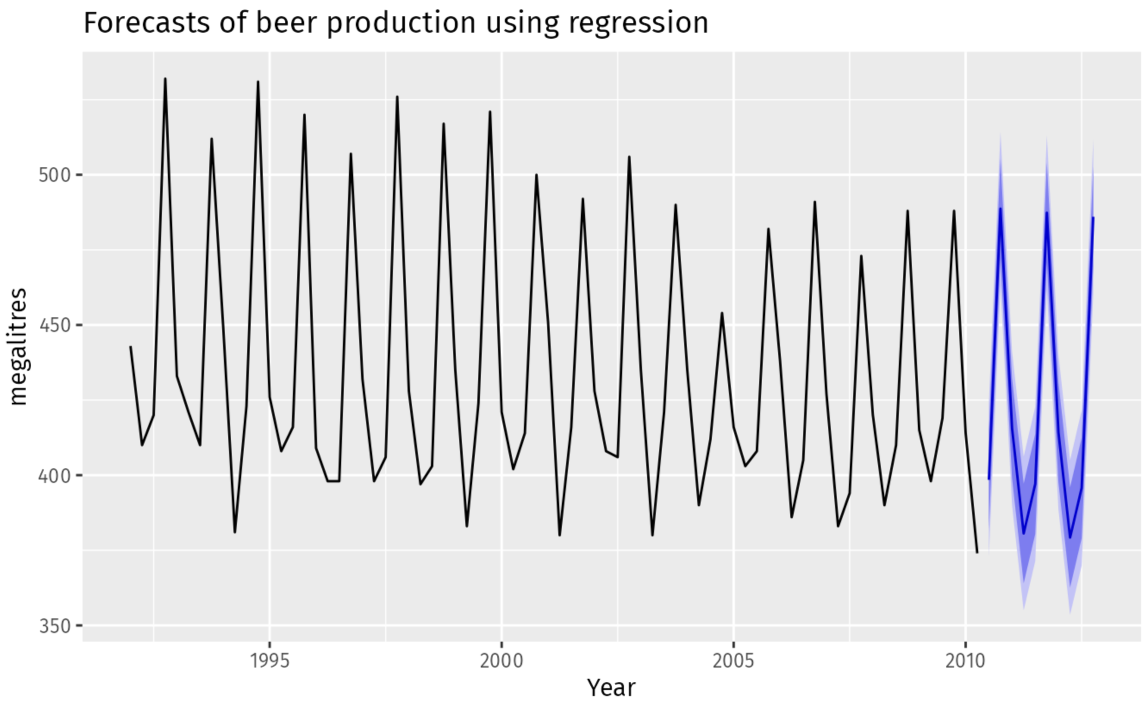
Simple Linear Models

- Trend:

$$\hat{y}_t = \beta_0 + \beta_1 * t$$

- Trend and seasonality dummies:

$$\hat{y}_t = \beta_0 + \beta_1 * t + \beta_{Q1} * I_{Q1} + \beta_{Q2} * I_{Q2} + \beta_{Q3} * I_{Q3}$$



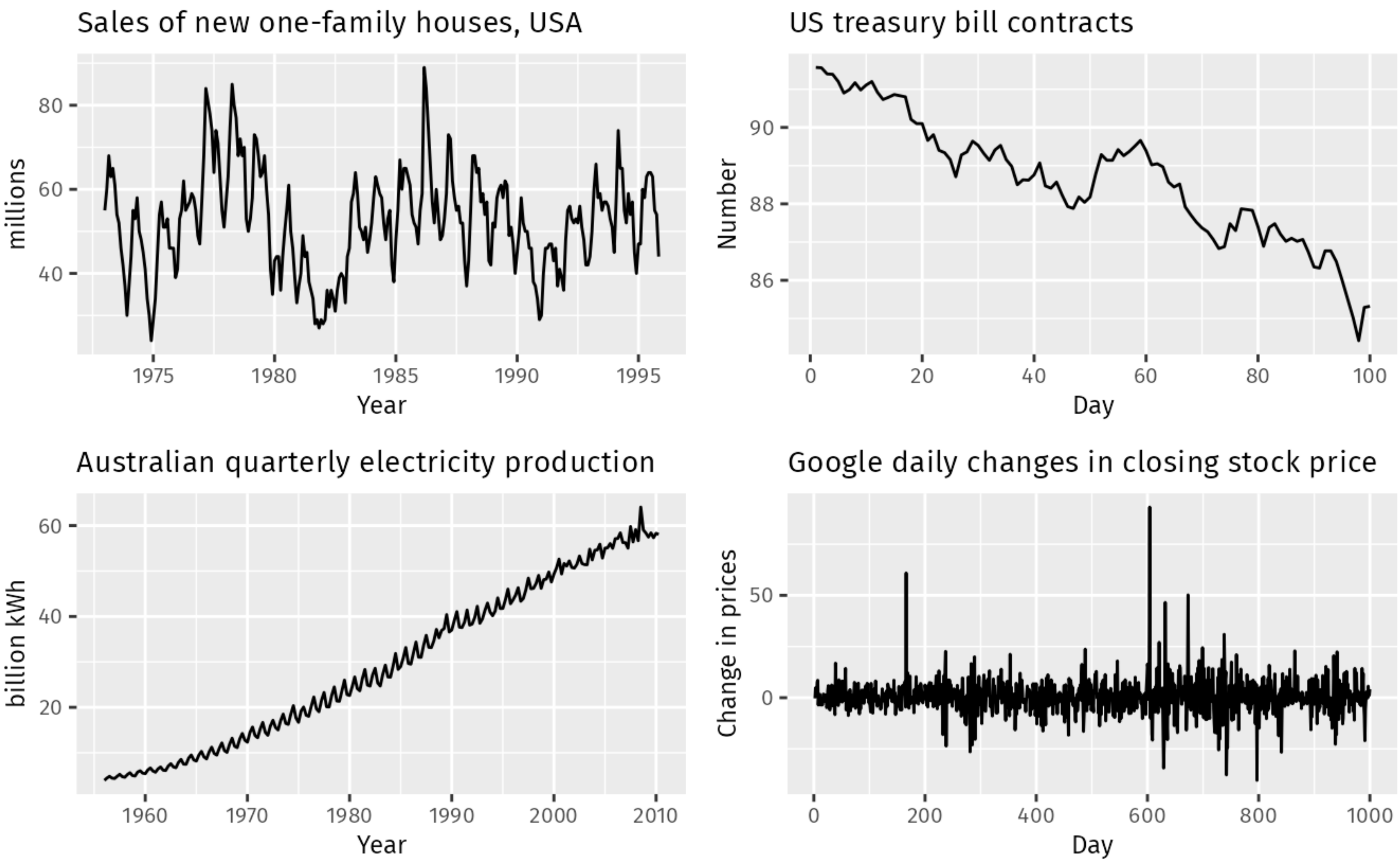
Source: Forecasting: Principles and Practice (2nd ed)

Time Series Decomposition

Decomposition

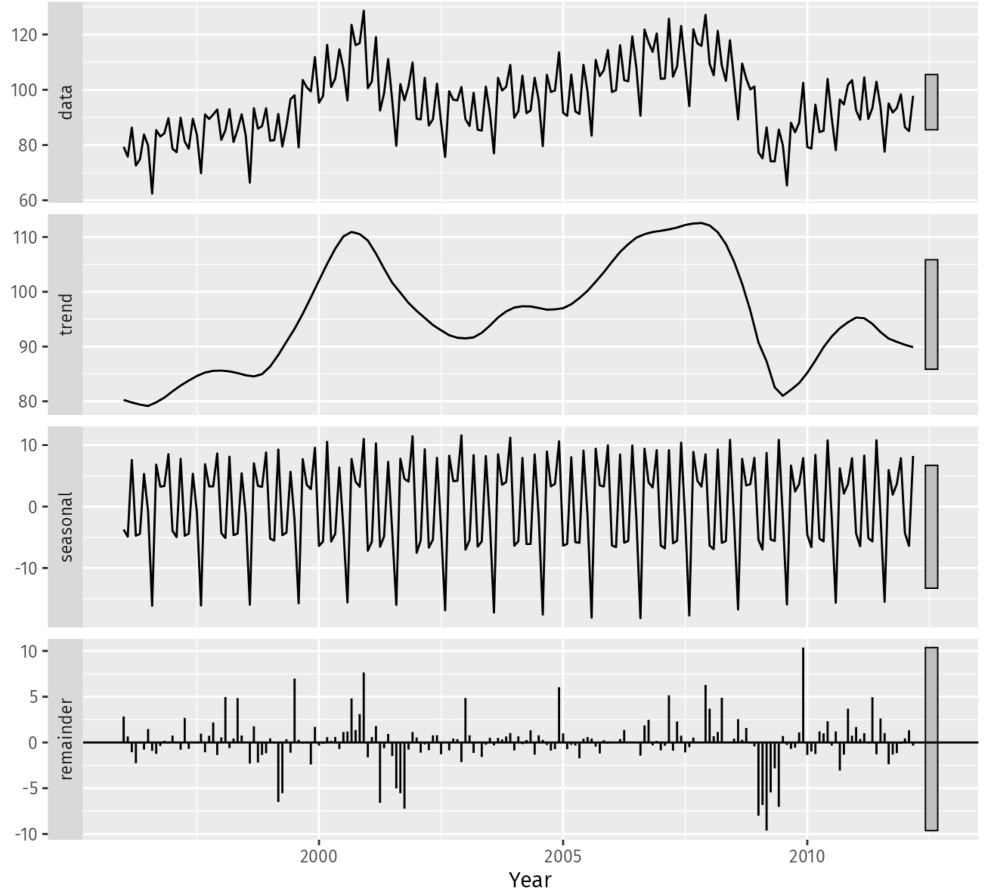
Components

- Trend
- Seasonality
- Cyclical patterns
- Random noise
- Other drivers
    - Autoregression
    - External variables
    - Special events



Source: [Forecasting: Principles and Practice (2nd ed)]

Decomposition



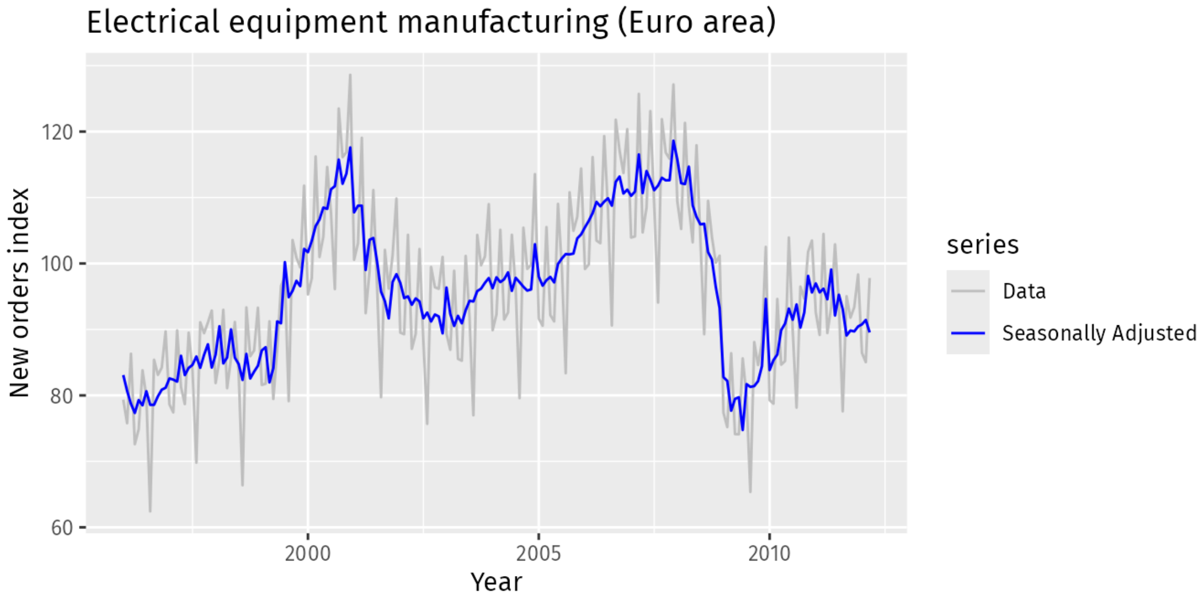
Additive model:

$$y_t = S_t + T_t + R_t$$

Multiplicative model:

$$y_t = S_t * T_t * R_t$$

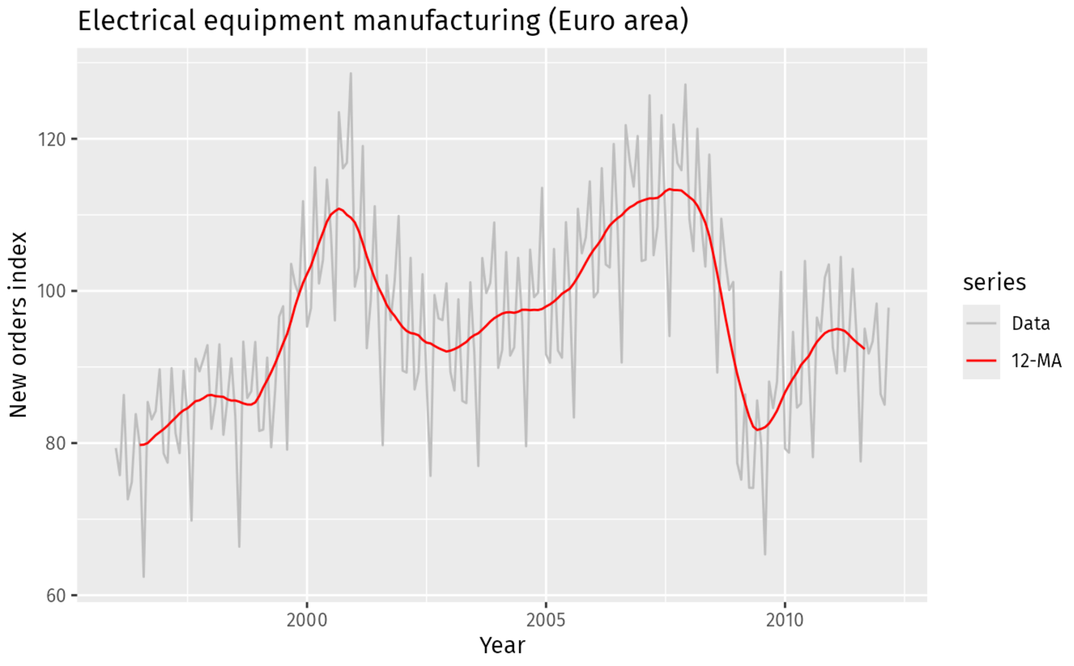
Seasonally adjusted data: the remaining data after removing the seasonal component. Helpful, if the variation due to seasonality is not of primary interest



Source: [Forecasting: Principles and Practice (2nd ed)]

Decomposition

$$\hat{T}_t = \frac{1}{m}\sum_{j=-k}^{k} y_{t+j}$$



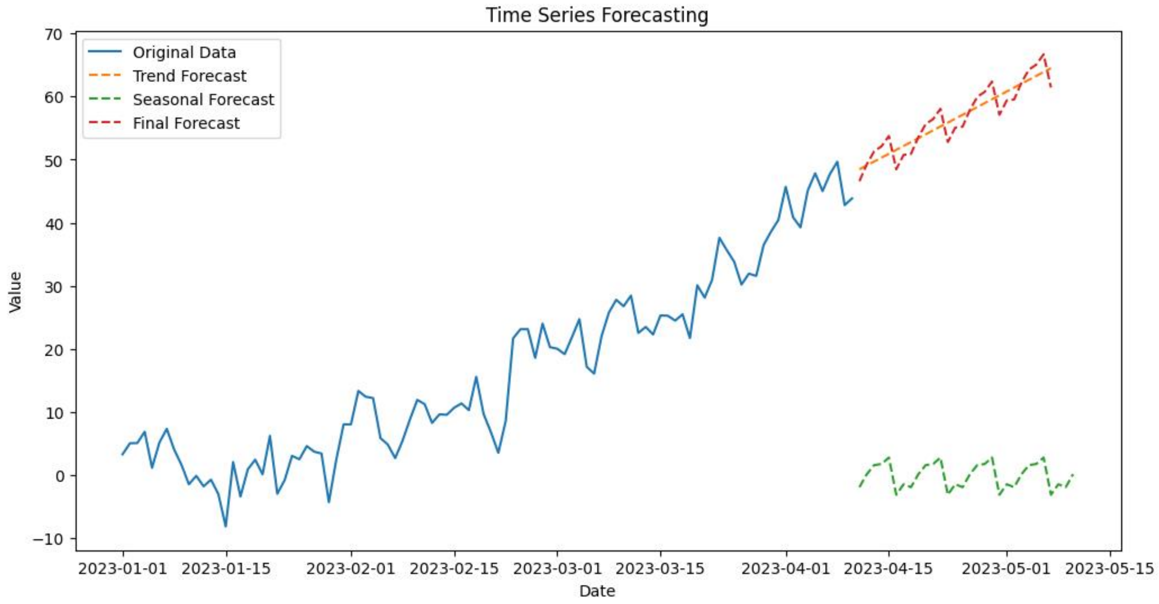
Moving Averages: estimate the trend-cycle component with the moving average of the time series

Decomposition Steps:

1. Compute the trend-cycle component using MA
2. Calculate the detrended series
3. To estimate the seasonal component for each season, simply average the detrended values for that season. For example, with monthly data, the seasonal component for March is the average of all the detrended March values in the data
4. The remainder component is calculated by subtracting the estimated seasonal and trend-cycle components

Source: [Forecasting: Principles and Practice (2nd ed)]

Forecasting with decomposition



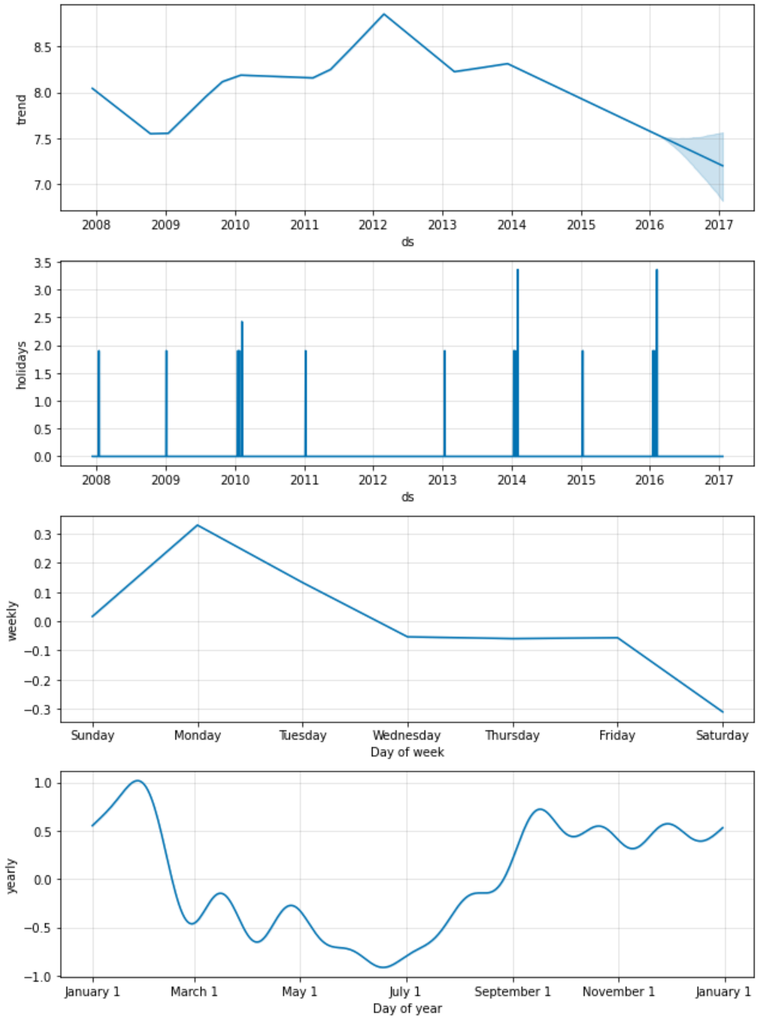
Forecasting Steps:

1. Decompose the time series
2. Forecast the components separately
    - The seasonal component can usually be considered constant
    - Use dummy variables for holiday effects
    - Use any forecasting method for the trend-cycle component and the remaining component (regression, ARIMA etc.)
3. Combine the component forecasts

When to use decomposition?

- Understand the underlying components
- Complex seasonal pattern that needs to be isolated
- Preprocessing step before applying other forecasting methods

Prophet



Source: Seasonality, Holiday Effects, And Regressors | Prophet

Prophet: Easy to use framework in R and Python

Additive model:

- Piecewise-linear trend (trend with changepoints)
- Seasonal effects
- Holiday effects

Features:

- Automatic changepoint detection
- Automatic seasonality modelling with Fourier terms
- Additional regressors
- Easy cross-validation and hyperparameter tuning
- Uncertainty Intervals

NeuralProphet: Fusing traditional time series algorithms using standard deep learning methods

Statistical Modelling

Statistical Time Series Models

- Exponential smoothing (Exponential Moving Average, EMA)
  - $s_0 = x_0$
  - $s_t = \alpha * x_t + (1 - \alpha) * s_{t-1}$
  - Weights decrease exponentially
  - Longer memory compared to simple MA
- ARIMA model family
- Python:
  - statsmodels
  - pmdarima

Stationarity

Stationarity is a common assumption in many time series techniques.

Definition:

- Statistical properties, such as mean and variance, do not change over time
- The joint probability distribution of the process remains the same when shifted in time.
- No trend, no seasonality, no heteroskedasticity – the process should look the same at any time

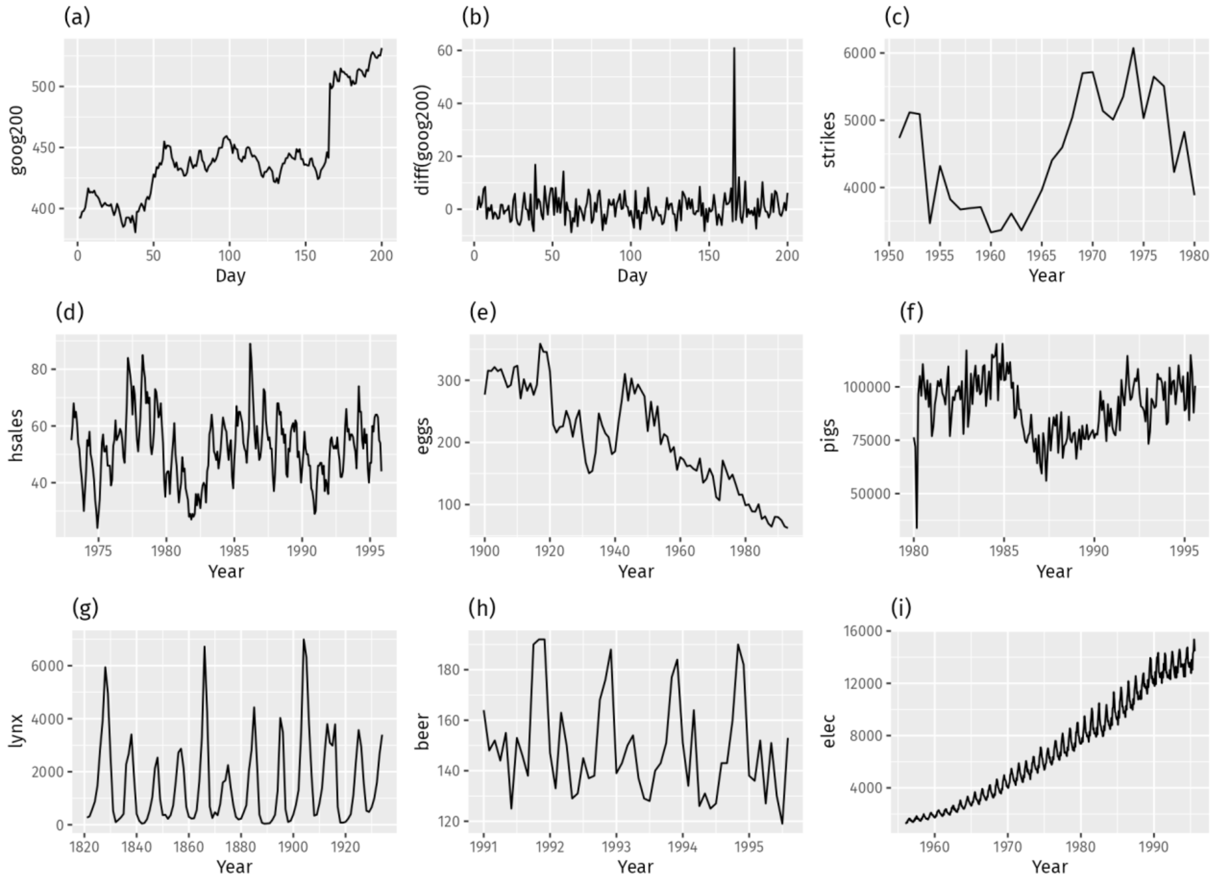
Transformations to achieve stationarity:

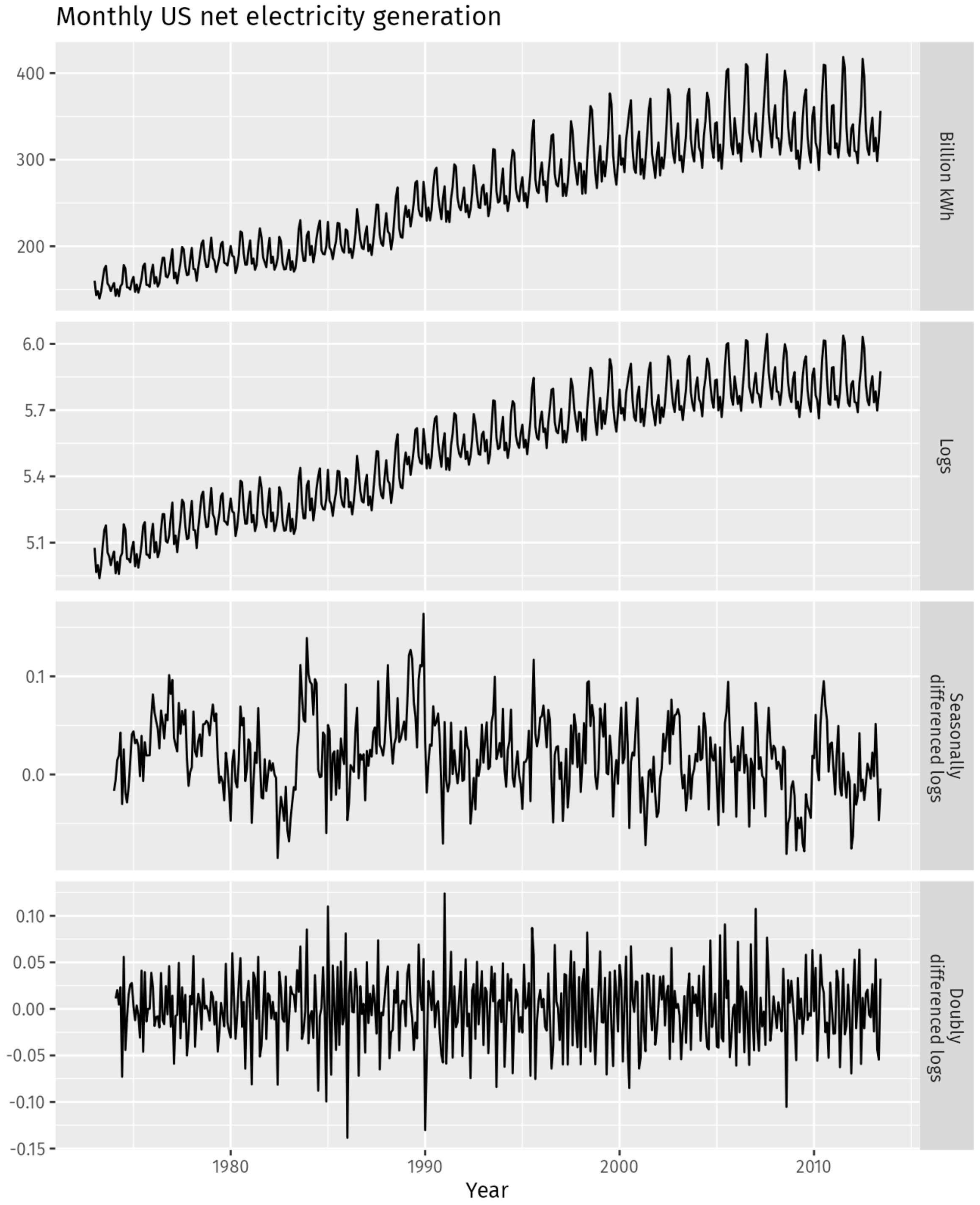
- Remove trend and seasonality before modelling
- Differencing (for the mean) $\mathrm{y}'_\mathrm{t} = \mathrm{y}_\mathrm{t} - \mathrm{y}_{\mathrm{t}-1}$
- Seasonal differencing $\mathrm{y}'_\mathrm{t} = \mathrm{y}_\mathrm{t} - \mathrm{y}_{\mathrm{t}-\mathrm{m}}$
- Log transformation (for variance)

Tests: the **Augmented Dickey-Fuller test** (ADF) tests the null hypothesis that a **unit root** is present, which means non-stationarity

White Noise: the variables are independent and identically distributed with a mean of zero

Stationarity





Source: [Forecasting: Principles and Practice (2nd ed)]

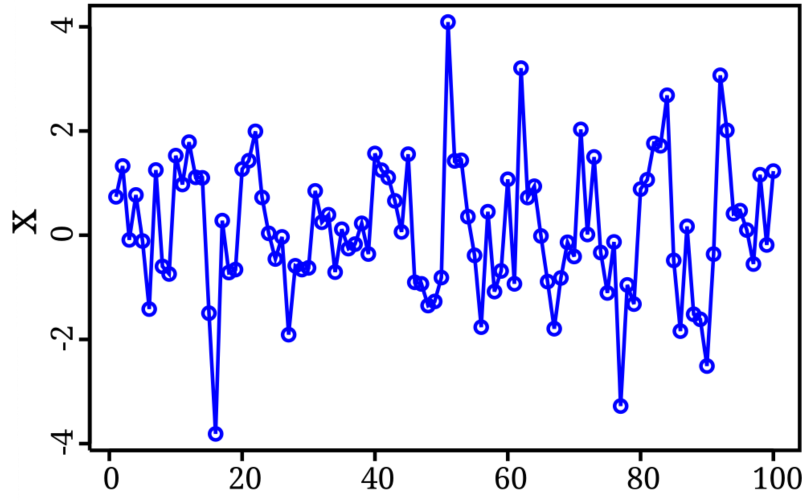
Autocorrelation

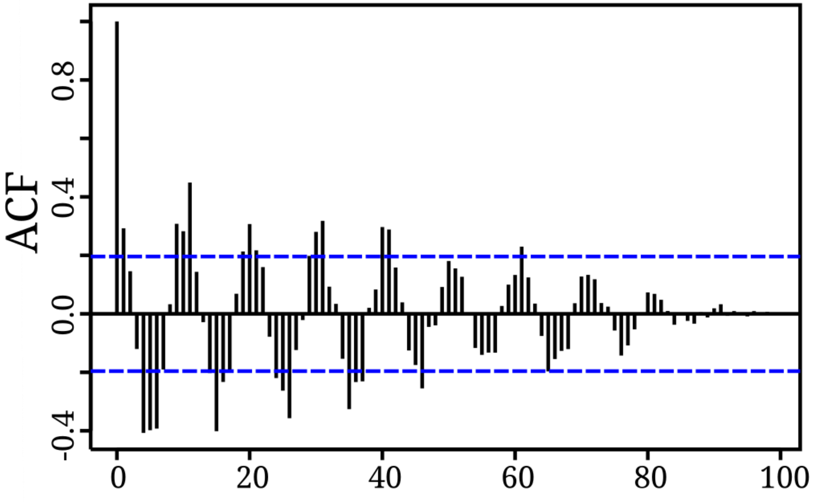
Autocorrelation Function (ACF)

- ACF measures the correlation between observations of a time series
- If the ACF shows significant correlations at certain lags, it indicates that past values influence future values.

Partial Autocorrelation Function (PACF)

- PACF measures the correlation between observations of a time series, but with the influence of intermediate lags removed.
- It helps identify the direct relationship between an observation and its lagged values, excluding the indirect effects.



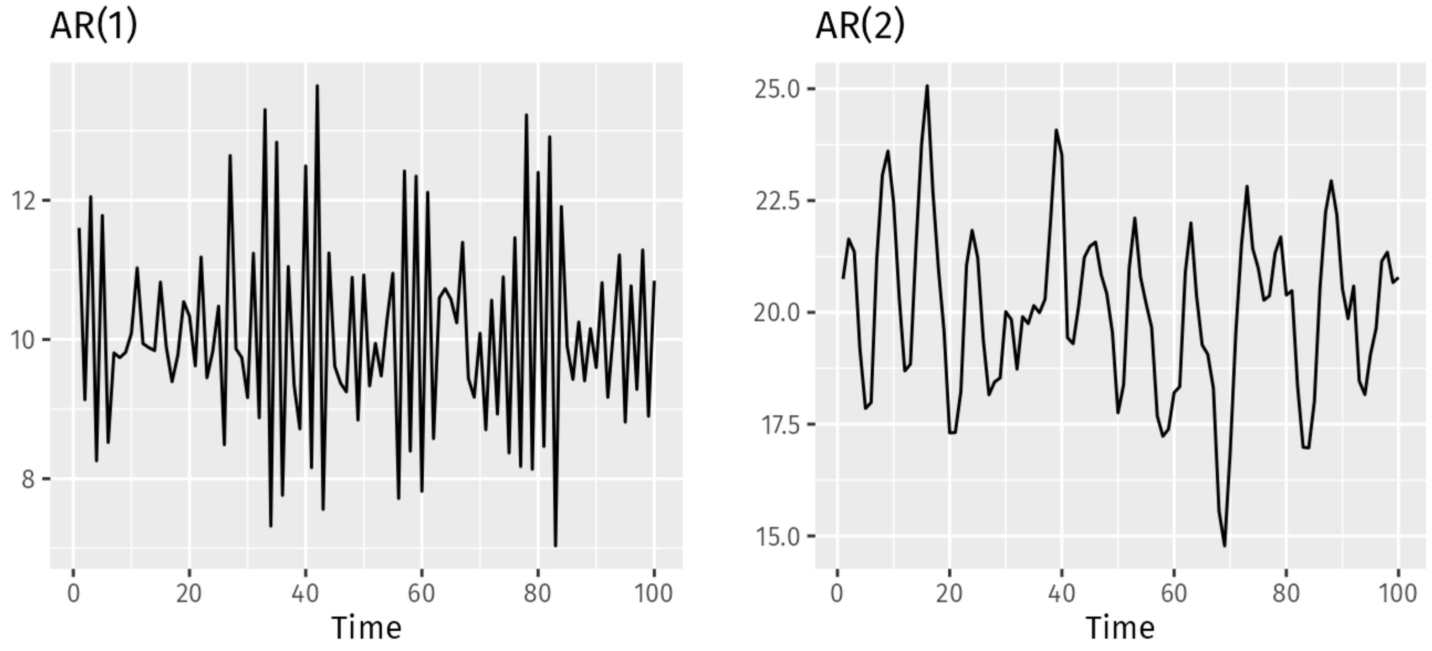


Source: Autocorrelation – Wikipedia

Autoregressive Models

AR(p): $$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \ldots + \varphi_p X_{t-p} + \varepsilon_t$$

- The PACF will show significant spikes up to lag (p) and then drop to zero
- The ACF typically tails off gradually.

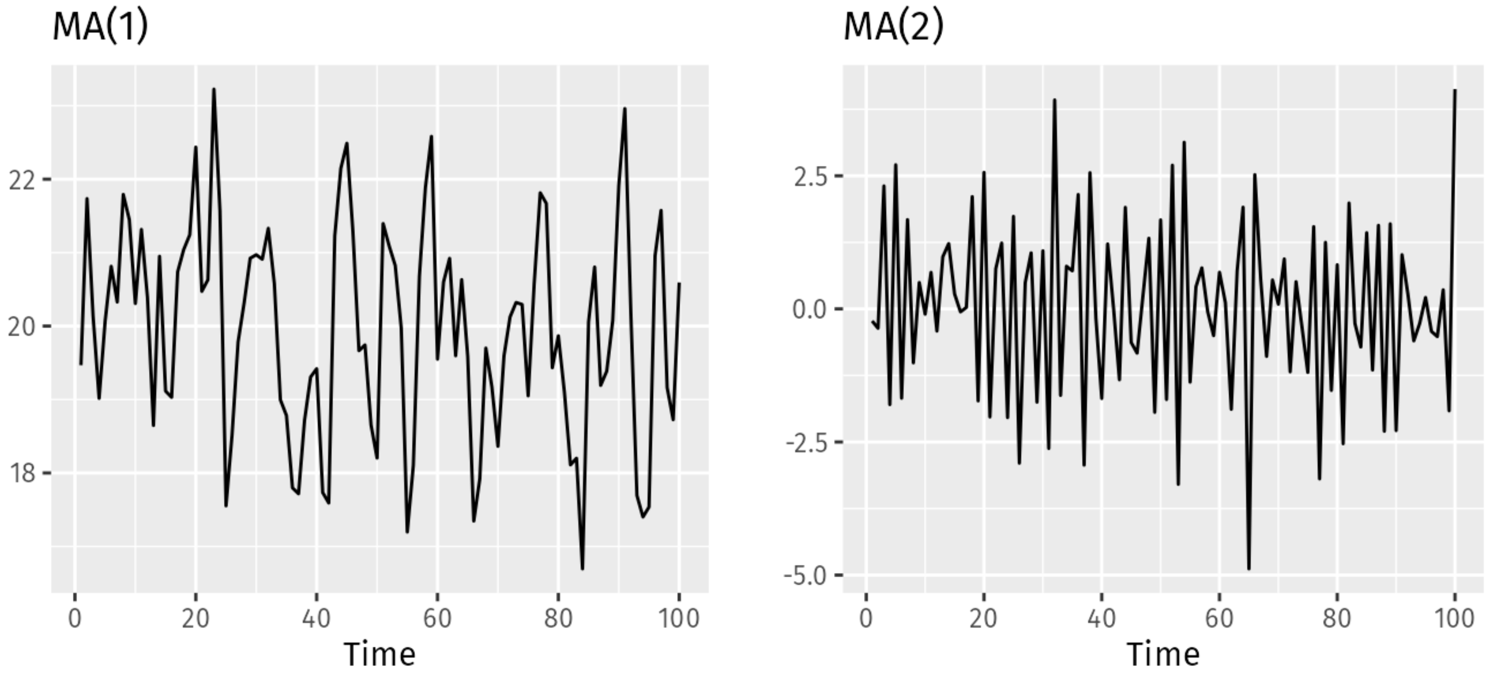


Source: [Forecasting: Principles and Practice (2nd ed)]

Moving Average Models

MA(q): $X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$

- The ACF will show significant spikes up to lag (q) and then drop to zero.
- The PACF typically tails off gradually.



Source: Forecasting: Principles and Practice (2nd ed)

ARIMA(p,d,q)

Autoregressive Integrated Moving Average

- p = order of the autoregressive part
- d = degree of first differencing involved
- q = order of the moving average part

When to use ARIMA?

- Univariate data
- No seasonality (or removed)
- Short-term forecasting
- Interpretable parameters
- Finance, economics

Steps:

1. Visualize the data
2. Check stationary (ADF test)
3. Differencing until stationarity
4. Plot ACF and PACF to determine the order of AR and MA terms (p and q values)
5. Fit the models with selected (p,d,q) orders. Compare different candidates with AIC
   - Akaike Information Criterion: $AIC = 2k - 2\ln(\hat{L})$
6. Check if the residuals are white noise
7. Forecast and evaluate performance

SARIMAX

- Seasonal ARIMA: $\text{SARIMA}(p, d, q)\ (P, D, Q)_m$
- Seasonal ARIMA with Exogenous Variables: SARIMAX

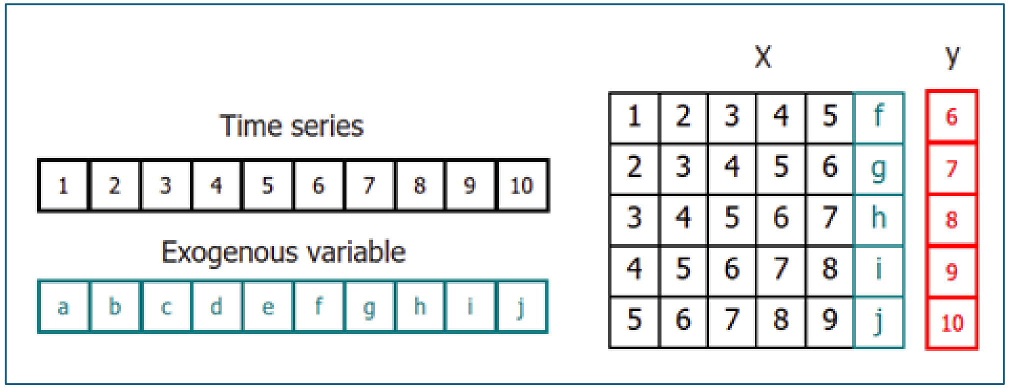
Machine Learning

Forecasting with Machine Learning

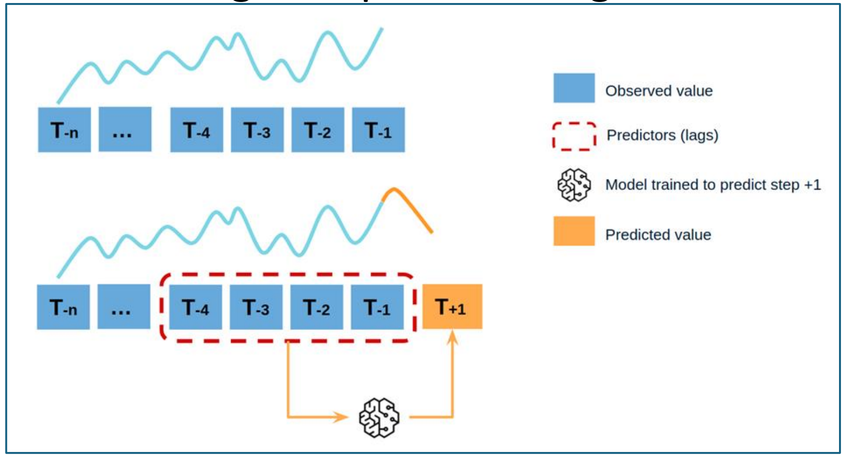
- Use ML algorithms like Random Forest or XGBoost
- Use Deep Learning Architectures like RNN, LSTM and Transformers
- Input features:
  - Lags of the output variable
  - External variables
- When use machine learning?
  - Multivariate data with complex relationships (e.g. interactions)
  - When statistical models are not performing well
  - Large amount of data
- Python packages:
  - Skforecast
  - TensorFlow
  - PyTorch Forecasting

Forecasting with Machine Learning

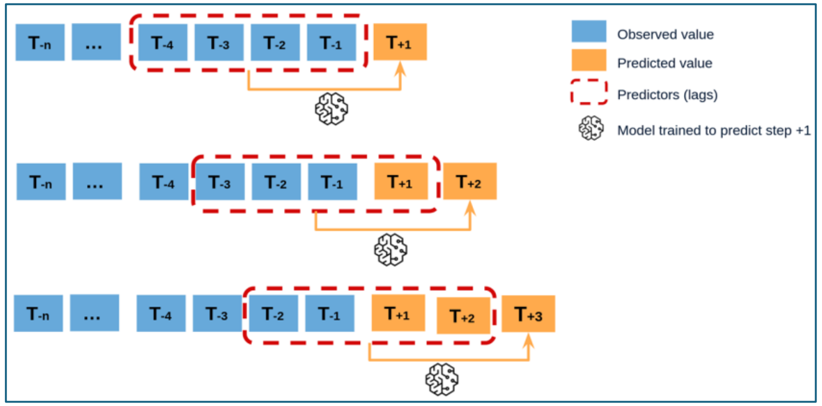
Reshape the time series data to X and y



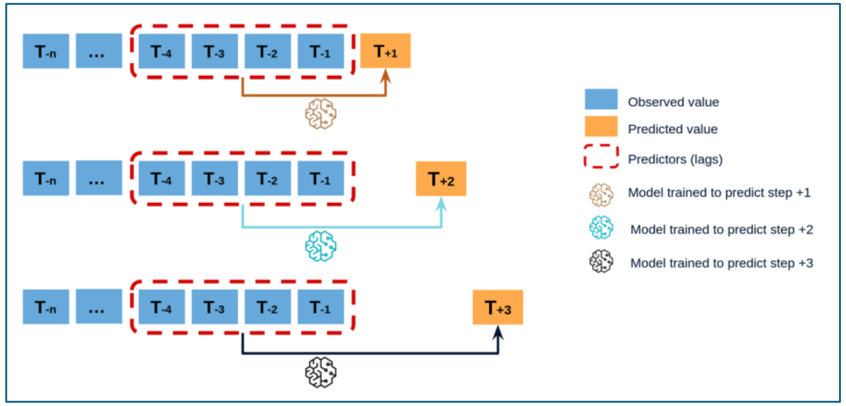
Single-step forecasting



Recursive multi-step forecasting



Direct multi-step forecasting

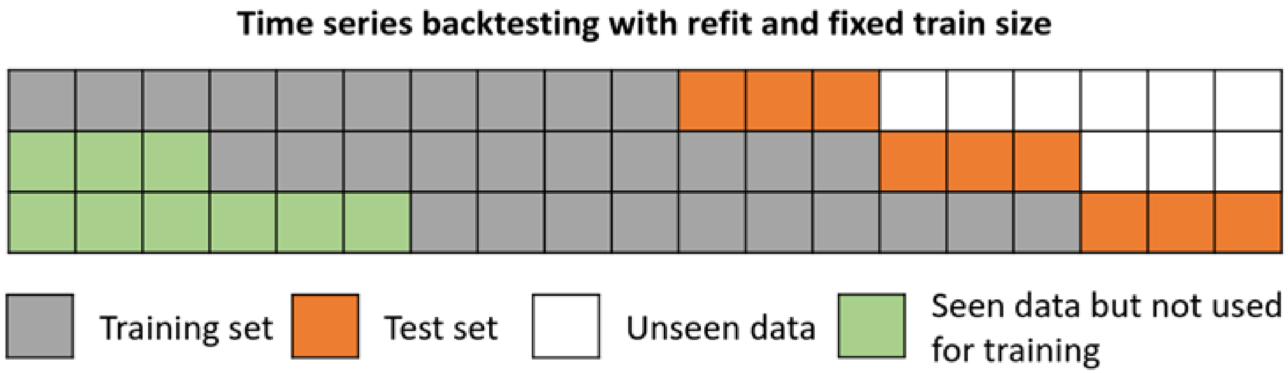


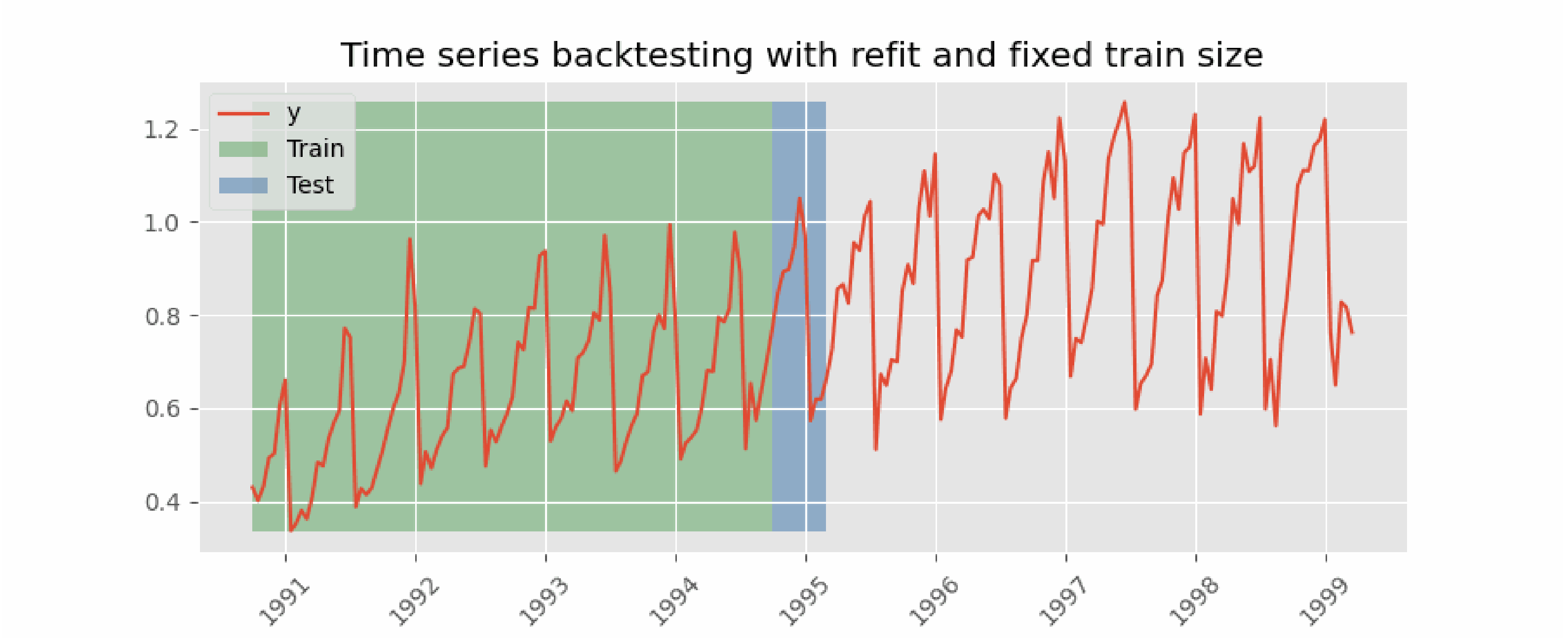
Source: [Intro to Forecasting - Skforecast Docs]

Preprocessing

- Deal with missing values and data errors
  - Consider business meaning (e.g. public holiday: sales = 0 + add dummy)
  - Forward fill
  - Interpolation
- Feature engineering (when using predictors)
  - Lags
  - Rolling statistics
  - External variables
- Seasonality
  - Add binary variables for the days of the week
  - Use 7 lags
  - Fourier series

Evaluation





Source: [Intro to Forecasting - Skforecast Docs]

Evaluation

Simple error term: $\epsilon_{T+h} = y_{T+h} - \hat{y}_{T+h}$

Scale-dependent errors:

$$MAE = mean(|\epsilon_t|)$$

$$RMSE = \sqrt{mean(\epsilon_t^2)}$$

- Not suitable for comparison

Percentage errors:

$$MAPE = mean\left(\left|\frac{\epsilon_t}{y_t}\right|\right)$$

- Unstable if y_t is close to zero
- Percentage is not always meaningful

Scaled errors:

$$MASE = \frac{mean(|\epsilon_t|)}{\frac{1}{T-1}\sum_{i=2}^{T}|y_t - y_{t-1}|} = \frac{MAE}{MAE_{Naive}^{Training}}$$

Time Series Analysis

- Simple Forecasting Methods
- Time Series Decomposition
- Statistical Modelling
- Machine Learning

Thank you for your attention!

Your feedback would be much appreciated:

Any Questions?

Gergely Zsombor Haász

haasz.zsombi@gmail.com