Time-series Forecasting - Introduction

Dr. Josif Grabocka

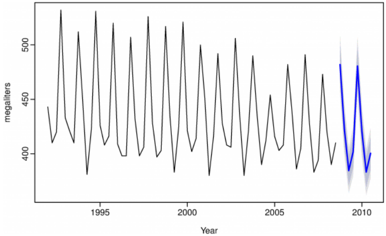
ISMLL, University of Hildesheim

Business Analytics

Time-series Forecasting

A time-series is a sequence of measurements **ordered** in time, such as:

- Daily stock prices
- Monthly rainfall
- Energy consumption/production
- Annual company profits
- Sales of products

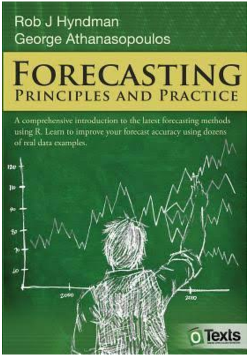


Australian Beer Production. Hyndman et al. 2014

Resources

The lectures on time-series forecasting are based on the book:

- Forecasting: principles and practice, Hyndman et al., 2014
- Freely vailable **online** `https://www.otexts.org/fpp`



Forecasting problem

- The forecasting problem demands estimating the values of a particular variable, given past measurements.
- The data can be of two types:
  - Time-series (**our focus**):
    - Given T measurements of a variable $y_1, y_2, \ldots, y_T$,
    - Accurately estimate $\hat{y}_{T+1|T}, \hat{y}_{T+2|T}, \ldots, \hat{y}_{T+h|T}$
  - Cross-sectional/Supervised learning (*our mandatory ML course*):
    - Given N^{Train} predictors $X \in \mathbb{R}^{N^{\text{Train}} \times M}$ and target $Y \in \mathbb{R}^{N^{\text{Train}}}$,
    - Accurately estimate the unknown targets $\hat{Y} \in \mathbb{R}^{N^{\text{Test}}}$ of a new test set of predictors $X \in \mathbb{R}^{N^{\text{Test}} \times M}$ and target

Steps of Forecasting

1. **Problem definition:** How forecasts are used? Who needs them? Talking to domain experts.

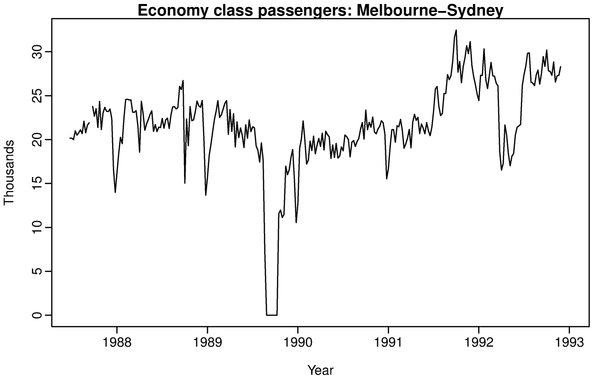
2. **Gathering information:** Mostly gather all the relevant recorded data and also the accumulated expertise

3. **Preliminary (explanatory) analysis**: **Graph the data.** Are there consistent patterns, trends, seasonality? Any business cycles? Unexplained outliers? How strong is the relationship between variables?

4. **Choosing and fitting models:** Understand data properties and choose a statistical model accordingly. Compare against other potential models.

5. **Using and evaluating a forecast model:** Compute accuracy measures to assess the quality of the predictions.

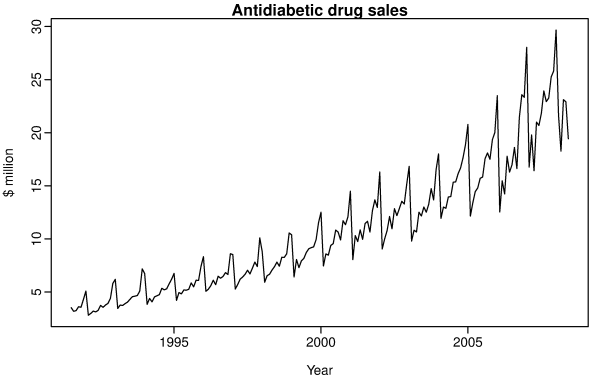
Forecasting Tools - Plot the Series



Weekly economy passengers on Ansett Airlines. Hyndman et al. 2014

'89 - industrial dispute, '92 - reduced load as economy seats moved to business, '91 - a boom, dips at start of year due to holidays.

Plot the Series (2)

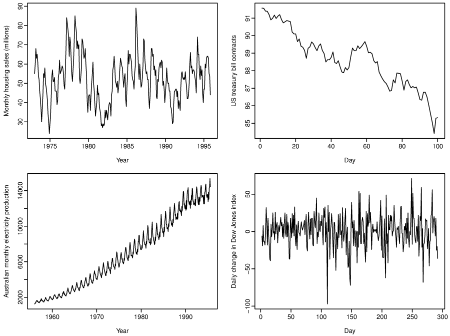


Monthly sales of antidiabetic drugs in Australia. Hyndman et al. 2014

Notice a trend and seasonality. Patients stockpile drugs due to gov. subsidization, leading to reduced sales at the end of each year.

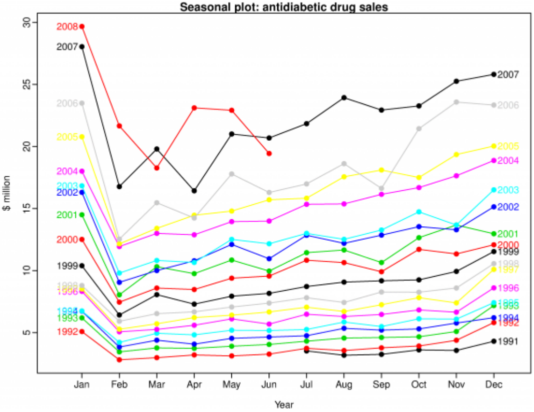
Time-series Patterns

- Trend: A long-term increase or decrease in the data
- Seasonality: Seasonal factors: Year, month, or day of week
- Cycle: A **non-periodic** fluctuation, also known as "business cycles"



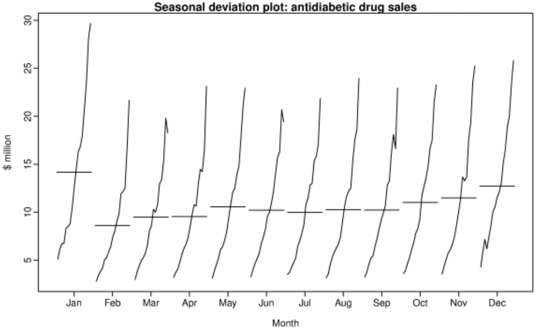
Hyndman et al. 2014

Inspection: Seasonal Plot



Atidiabetic drugs: Overlapping per-season series. Hyndman et al. 2014

Inspection: Seasonality Subseries Plot



Atidiabetic drugs: Plot per-season sub-series. Hyndman et al. 2014

Series Statistics

The mean of a series $y_1, y_2, \ldots, y_T$ is defined as:

$$\bar{y} = \frac{1}{T} \sum_{t=1}^{T} y_t$$

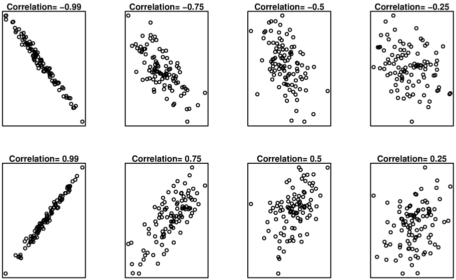
While its standard deviation:

$$\sigma = \sqrt{\frac{1}{T-1} \sum_{t=1}^{T} (y_t - \bar{y})^2}$$

Correlation Coefficient

The correlation between vectors $a_1, a_2, \ldots, a_N$ and $b_1, b_2, \ldots, b_N$ is:

$$r = \frac{\sum_{i=1}^{N} (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^{N} (a_i - \bar{a})^2} \sqrt{\sum_{i=1}^{N} (b_i - \bar{b})^2}}$$



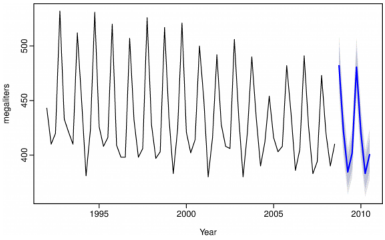
Autocorrelation

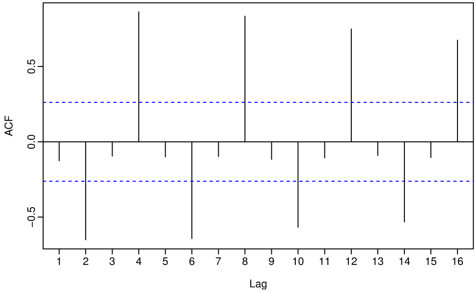
- The correlation coefficient between a series and **a lagged** version of itself.
- r_k measures the correlation between $y_{k+1}, y_{k+2}, \ldots, y_T$ and $y_1, y_2, \ldots, Y_{T-k}$:

$$r_k = \frac{\sum_{t=k+1}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sqrt{\sum_{t=k+1}^{T} (y_t - \bar{y})^2} \sqrt{\sum_{t=1}^{T-k} (y_t - \bar{y})^2}}$$

$$r_k \approx \frac{\sum_{t=k+1}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{T} (y_i - \bar{y})^2}$$

Autocorrelation Function - ACF - Plot $r_1, r_2, \ldots$

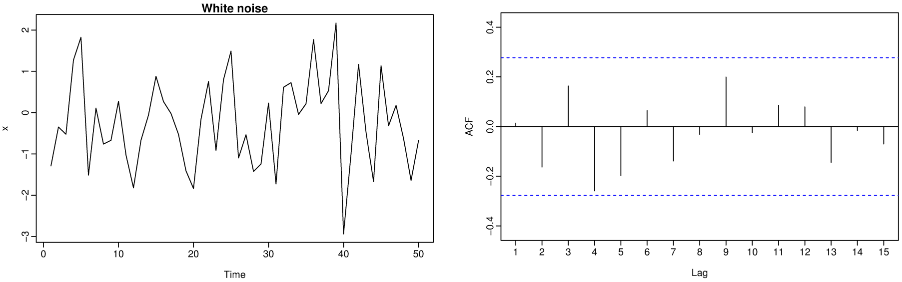




ACF of Beer Production. Hyndman et al. 2014

- Trend when **high** values of ordered coefficients $r_1 > r_2 > r_3 \ldots$
- Seasonal of period m where have high values of $r_t, r_{t-m}, r_{t-2m} \ldots$
- Noisy series have **small** coefficients (next slide ...)

ACF - White Noise



Hyndman et al. 2014

- The dashed blue lines show the $\pm \frac{2}{\sqrt{T}}$
- Coefficients of non-noisy series exceed those bounds

Some Simple Forecasting Methods (1)

- Predict the **average** value:

$$\hat{y}_{T+h|T} = \bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

- Naive: Predict the **last** observed value:

$$\hat{y}_{T+h|T} = y_T$$

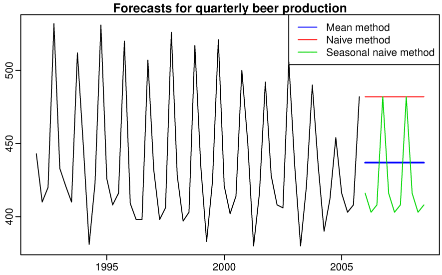
- Seasonal Naive: Predict the **last** observed **periodic** value:

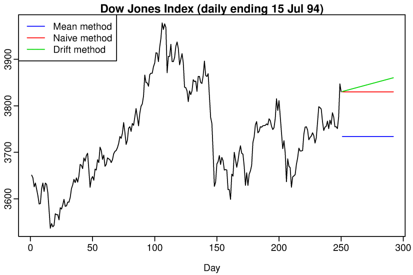
$$\hat{y}_{T+h|T} = y_{T+h-km}, \text{ for } k = \lfloor \frac{h-1}{m} \rfloor + 1$$

Few Simple Forecasting Methods (2)

- Drift method: Use the average change/drift:

$$\hat{y}_{T+h|T} = y_T + \frac{h}{T-1}\sum_{t=2}^{T}(y_t - y_{t-1}) = y_T + h\frac{y_T - y_1}{T-1}$$





Hyndman et al. 2014

Evaluating Forecast Accuracy

- Scale-dependent errors:
  - Mean **Absolute** Error:

$$\text{MAE} = \frac{1}{T}\sum_{t=1}^{T} \left| y_t - \hat{y}_{t|t-1} \right|$$

  - Root Mean **Square** Error:

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T} (y_t - \hat{y}_{t|t-1})^2}$$

- Percentage errors are independent to scale
  - Mean Absolute **Percentage** Error:

$$\text{MAPE} = \frac{1}{T}\sum_{t=1}^{T} \left| \frac{y_t - \hat{y}_{t|t-1}}{y_t} \right|$$

Evaluating Forecast Accuracy (2)

- Percentage errors are undefined when $y_t = 0$ and produce **extreme** values when $y_t \approx 0$
- A solution is to use scaled errors as an alternative to percentage errors
- Hence, the Mean Absolute Scaled Error:

$$\text{MASE} = \frac{1}{T} \sum_{t=1}^{T} \left| \frac{y_t - \hat{y}_{t|t-1}}{\frac{1}{T-1} \sum_{t'=2}^{T} |y_{t'} - y_{t'-1}|} \right|$$

- Both nominator and denominator are on the same scale, however cannot be perceived percentually

Exercise

Given a series $y_1 = 2, y_2 = -3, y_3 = 3, y_4 = -2$, predict $\hat{y}_{5|4} =?$ using

Note $y_5 = 2$.

- Average method:
  $\hat{y}_{5|4} = \frac{1}{4}(2 - 3 + 3 - 2) = 0$, MAE=2
- Naive method:
  $\hat{y}_{5|4} = y_4 = -2$, MAE=3
- Seasonal Naive method, $m = 2$:
  $\hat{y}_{5|4} = y_3 = 3$, MAE=1
- Drift Method, $m = 2$:
  $\hat{y}_{5|4} = -2 + (-2 - 2)/3 = -3.33$, MAE=5.33

Why is the Seasonal Naive method performing better?

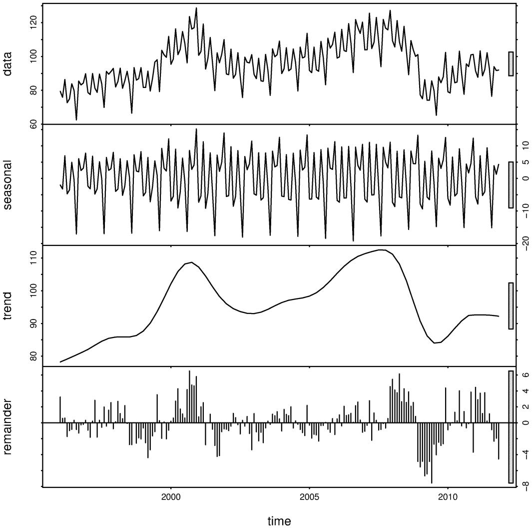
Time-series Decomposition

Time series can be thought as comprising of three components: a seasonal, trend-cycle and a remainder.

- **Additive** decomposition model:
  $y_t = S_t + T_t + E_t$
- **Multiplicative** decomposition model:
  $y_t = S_t \times T_t \times E_t$

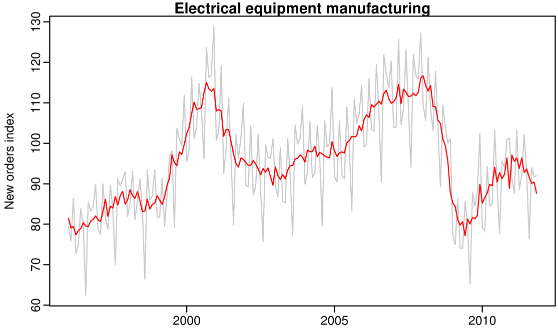
- Where S_t is the seasonality, T_t is the trend-cycle, while E_t is the remainder (error or irregular) at time t
- Please note that $y_t = S_t \times T_t \times E_t$ is equivalent to
  $\log y_t = \log S_t + \log T_t + \log E_t$

Time-series Decomposition (Hyndman et al. 2014)



Seasonally-Adjusted Series

- **Sometimes** remove the seasonal component, i.e.
  - Additive: $y_t^{\text{new}} := y_t^{old} - S_t$
  - Multiplicative: $y_t^{\text{new}} := y_t^{old} / S_t$
- Seasonality of unemployment caused by school leavers seeking work is not interesting, while the seasonal trend matters.



Moving Average Smoothing

Given a neighborhood of size $2k$, the moving average smoothing is:

$$\hat{T}_t^{(1)} = \frac{1}{m}\sum_{j=-k}^{k} y_{t+j}, (m = 2k+1); \text{ or } \quad \hat{T}_t^{(1)} = \frac{1}{m}\sum_{j=-k}^{k-1} y_{t+j}, (m = 2k)$$

- Smoothing helps retrieve the **trend** component
- Chain smoothing:

$$\hat{T}_t^{(2)} = \frac{1}{m}\sum_{j=-q}^{q} T_{t+j}^{(1)}, (m = 2q+1); \text{ or } \quad \hat{T}_t^{(2)} = \frac{1}{m}\sum_{j=-q}^{q-1} T_{t+j}^{(1)}, (m = 2q)$$

- For instance a 4-MA followed by 2-MA is denoted as 2×4-MA:

$$\begin{aligned}\hat{T}_t^{(2)} &= \frac{1}{2}\left[\frac{1}{4}(y_{t-3} + y_{t-2} + y_{t-1} + y_t) + \frac{1}{4}(y_{t-2} + y_{t-1} + y_t + y_{t+1})\right] \\ &= \frac{1}{8}y_{t-3} + \frac{1}{4}y_{t-2} + \frac{1}{4}y_{t-1} + \frac{1}{4}y_t + \frac{1}{8}y_{t+1}.\end{aligned}$$

Classical Decomposition - Additive

Assuming a seasonal period m, the **Additive** decomposition is:

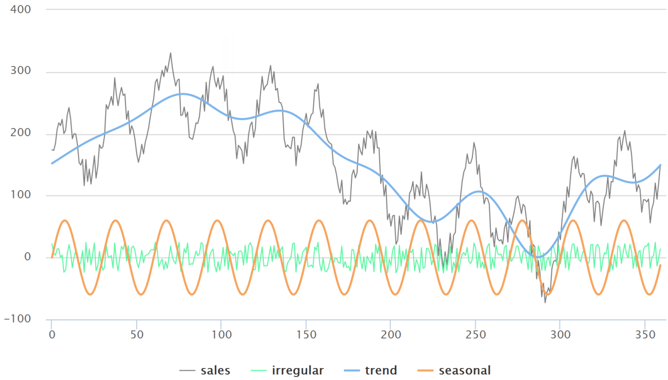
- **Step 1: Trend** is $\hat{T} = \begin{cases} m \bmod 2 = 0, & 2 \times m - \text{MA} \\ m \bmod 2 = 1, & m - \text{MA} \end{cases}$
- **Step 2:** Calculate de-trended series $\hat{D}_t = y_t - \hat{T}_t, \ t = 1, \ldots, T$
- **Step 3: Seasonality**, say, of each month, is the average among the de-trended values $\hat{D}$ of that month. Ensure that seasonal indices add to zero, yielding $\hat{S}_t, t = 1, \ldots, T$.
- **Step 4: Remainder** is $\hat{E}_t = y_t - \hat{T}_t - \hat{S}_t, \ t = 1, \ldots, T$

Classical Decomposition - Multiplicative

Assuming a seasonal period m, the **Multiplicative** decomposition is:

- **Step 1: Trend** is $\hat{T} = \begin{cases} m \bmod 2 = 0, & 2 \times m - \text{MA} \\ m \bmod 2 = 1, & m - \text{MA} \end{cases}$
- **Step 2:** Calculate de-trended series $\hat{D}_t = y_t / \hat{T}_t, \; t = 1, \ldots, T$
- **Step 3: Seasonality**, say, of each month, is the average among the de-trended values $\hat{D}$ of that month. Ensure that seasonal indices add to m, yielding $\hat{S}_t, t = 1, \ldots, T$.
- **Step 4: Remainder** is $\hat{E}_t = y_t / (\hat{T}_t \hat{S}_t), \; t = 1, \ldots, T$

Example - Classical Decomposition - Additive



Source: `http://www.alanzucconi.com/`