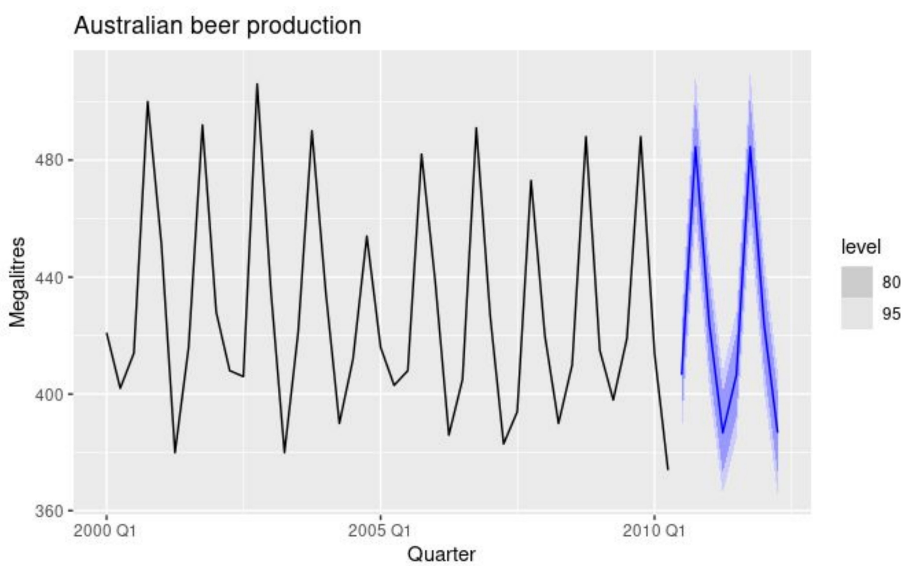
Time Series Forecasting

Time Series Forecasting

Time series refers to a sequence of data points collected, recorded, or otherwise measured over time intervals. Each data point in a time series is associated with a specific timestamp, allowing for the chronological organization of the data. Time series data is commonly used in various fields to analyze trends, patterns, and behaviors that evolve over time.
Time series forecasting - process of using historical time series data to make predictions about future values. The goal is to analyze patterns and trends within the data to build models that can then be used to forecast future points in the time series.

- Anything that is observed sequentially over time is a time series
- Forecasting aims to estimate how the sequence of observations will continue into the future
- Forecasting can be accompanied with indication of uncertainty



Types of Forecasting

Explanatory model: predictor variables capture reasons of the change[1]. The error term accounts for missing variables;

Electricity Demand = f(current temperature, strength of economy, population,time of day, day of week, error).

Typical models include classifiers and regression.

Timeseries model: prediction of the future is based on past values of a variable, but not on external variables that may affect the system.

Electricity Demand $ED_{t+1} = f(ED_t, ED_{t-1}, ED_{t-2}, ED_{t-3}, \ldots, \text{error})$

Mixed Models:

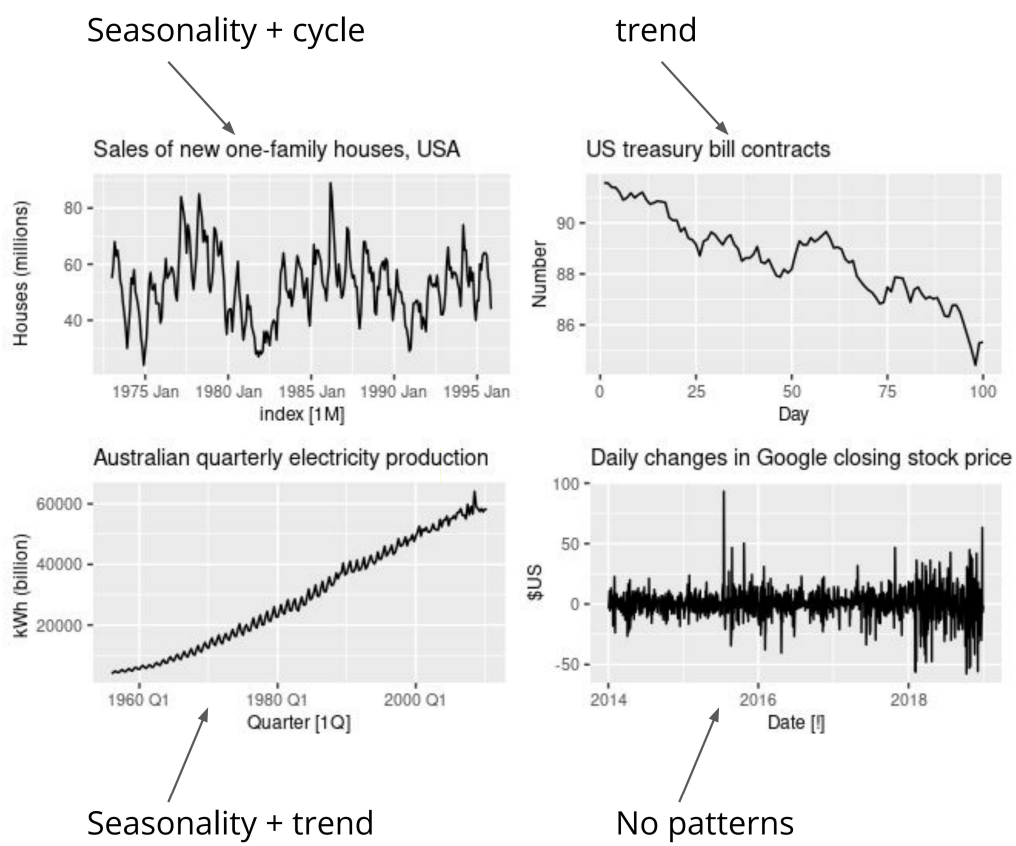
$ED_{t+1} = f(ED_t, \text{current temperature, time of day, day of week, error})$.

[1] https://www.sciencedirect.com/science/article/pii/S136481521730542X

Time Series Components

A time series may consist of these components:

- **Trend:** long-term increase or decrease in the data.
- **Seasonality:** when a time series is affected by seasonal factors such as the time of the year or the day of the week. Seasonality is always of a fixed and known period.
- **Cycles:** similar to seasonality, yet lacks a fixed frequency. These fluctuations are usually due to economic conditions, and are often related to the "business cycle." The duration of these fluctuations is usually at least 2 years.



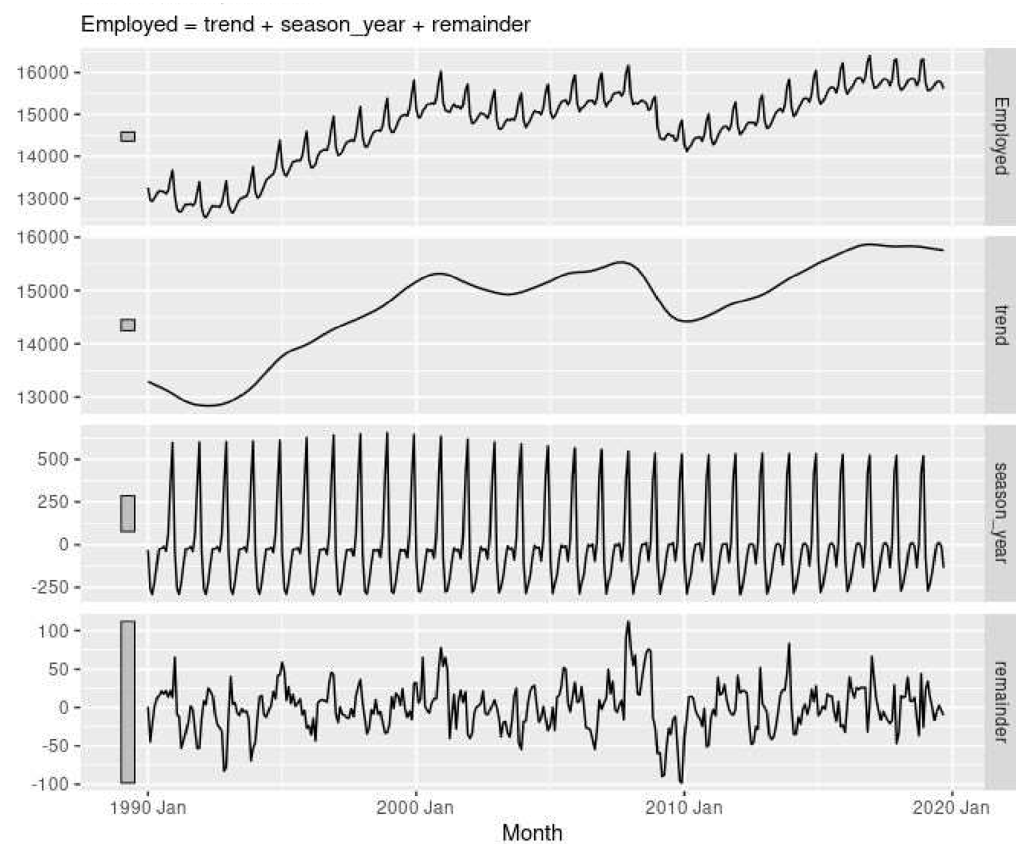
Time Series Decomposition

- We usually combine the trend and cycle into a single **trend-cycle** component (often just called the **trend** for simplicity)
- The seasonal components may slowly change over time
- The remainder component is what is left over when the seasonal and trend-cycle components have been subtracted from the data.
- additive decomposition

$$y_t = S_t + T_t + R_t$$

- multiplicative decomposition

$$y_t = S_t \times T_t \times R_t.$$



Adjustments Before Decomposition

- Calendar adjustments
  - Sum of monthly sales => daily sales

    Not all months have the same number of days/work days
- Population adjustments
  - => per-capita
- Inflation adjustment
  - If z_t denotes the price index and y_t denotes the original house price in year t, then

$$x_t = y_t / z_t * z_{2000}$$

Moving Averages to Estimate the Trend-Cycle

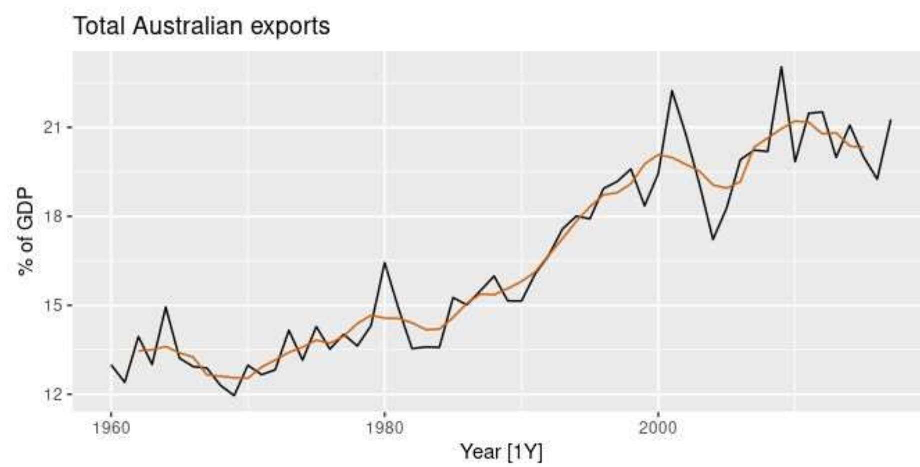
$$\hat{T}_t = \frac{1}{m} \sum_{j=-k}^{k} y_{t+j},$$

- m=2k+1, k is number of neighbours
- Every value is replaced by the average of itself and its k nearest neighbourhood
- Results in a smooth trend-cycle component
- Ex: K=2, m=5, the result is called 5-MA
- Averaging is applied using a sliding window

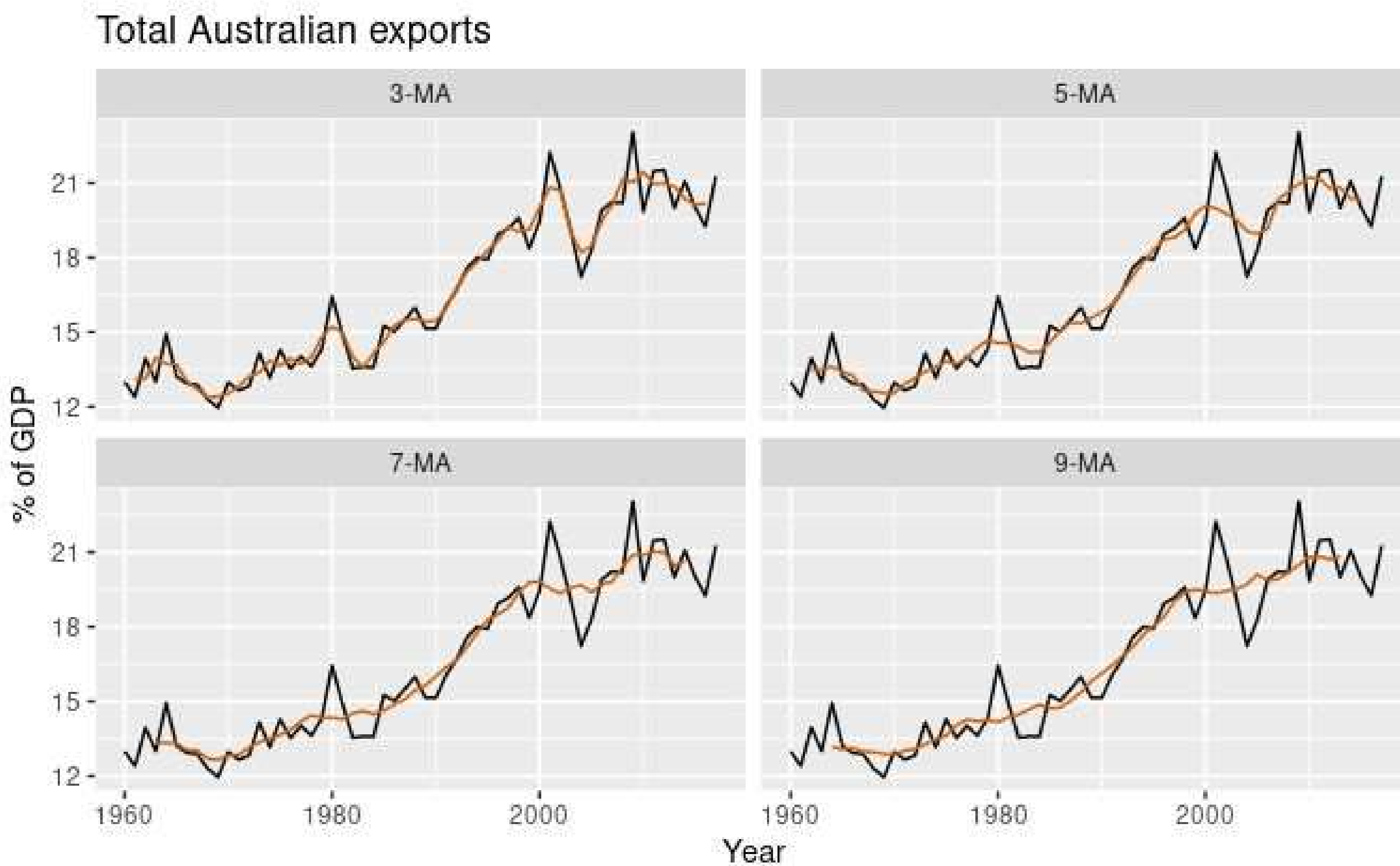
Moving Averages to Estimate the Trend-Cycle

Table 3.1: Annual Australian exports of goods and services: 1960–2017.

Year	Exports	5-MA
1960	12.99	
1961	12.40	
1962	13.94	13.46
1963	13.01	13.50
1964	14.94	13.61
1965	13.22	13.40
1966	12.93	13.25
1967	12.88	12.66
...	...	...
2010	19.84	21.21
2011	21.47	21.17
2012	21.52	20.78
2013	19.99	20.81
2014	21.08	20.37
2015	20.01	20.32
2016	19.25	
2017	21.27	



Moving Averages to Estimate the Trend-Cycle



Moving Averages of Moving Averages

Quarter	Beer	4-MA	2x4-MA
1992 Q1	443.00		
1992 Q2	410.00	451.25	
1992 Q3	420.00	448.75	450.00
1992 Q4	532.00	451.50	450.12
1993 Q1	433.00	449.00	450.25
1993 Q2	421.00	444.00	446.50
...	...	...	...
2009 Q1	415.00	430.00	428.88
2009 Q2	398.00	430.00	430.00
2009 Q3	419.00	429.75	429.88
2009 Q4	488.00	423.75	426.75
2010 Q1	414.00		
2010 Q2	374.00		

$T_t = \frac{1}{2}\left[\frac{1}{4}(y_{t-2}+y_{t-1}+y_t+y_{t+1}) + \frac{1}{4}(y_{t-1}+y_t+y_{t+1}+y_{t+2})\right] = \frac{1}{8}y_{t-2} + \frac{1}{4}y_{t-1} + \frac{1}{4}y_t + \frac{1}{4}y_{t+1} + \frac{1}{8}y_{t+2}$

Common for quarterly data

Moving Averages of Moving Averages

- Odd-order moving averages are symmetric
- Even-order MA are not symmetric. To make symmetric, we compute MA of MA
- The choice depends on the seasonality of data, e.g.,
  - Quarterly data with annual seasonality => 2 x 4-MA
  - Monthly data with annual seasonality ?
  - Daily data with weekly seasonality ?
- The goal is to average out seasonal variations, so that the smoothed timeseries has only the trend-cycle component.
- Other choices for the order of the MA will usually result in trend-cycle estimates being contaminated by the seasonality in the data.

how to even use garne ki odd kati order ko use garne?

Classical Additive Decomposition

Step 1

Compute the trend-cycle component T_t using a 2×m-MA or m-MA, for even-odd seasonality respectively.

Step 2

Calculate the detrended series: $y_t - T_t$

Step 3

Seasonal component for each season= average of detrended values for that season in all data

Adjust the seasonal components to ensure that they add to zero

Concat all seasonal components to form S_t

Step 4

The remainder component is $R_t = y_t - T_t - S_t$

Classical Decomposition

- Multiplicative decomposition works in a similar way
- The estimate of the trend-cycle is unavailable for the first few and last few observations.

  How many ?
- Assumes that the seasonal components repeats unchanged
- Tends to over smooth rapid rises/falls
- No special handling of unusual values, e.g., covid effect on air transport
- STL is currently commonly used:

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. J. (1990). STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, *6*(1), 3–33. http://bit.ly/stl1990

Simple Forecasting

- Mean Method: all future data are equal to the average of the history

$$y_{T+h} = (y_1 + \cdots + y_T) / T$$

- Naive Method: all future data are equal to last observed value

$$y_{T+h} = y_T$$

- Seasonal Naive: each forecast is equal to the last observed value in same season
- Drift Method

$$y_{T+h} = y_T + h \left(\frac{y_T - y_1}{T - 1} \right)$$

Forecasting With Decomposition

- Seasonally adjusted component $A_t = T_t + R_t$
- To forecast a decomposed time series, we separately forecast the seasonal component, S_t, and the seasonally adjusted component A_t
- The seasonal component is simply repeated
- The seasonally adjusted component is forecast using any method, e.g., drift