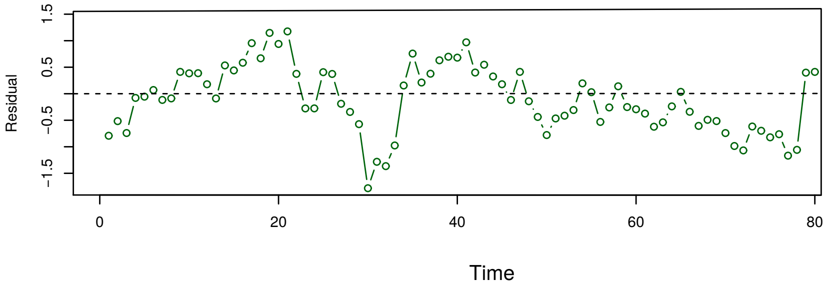
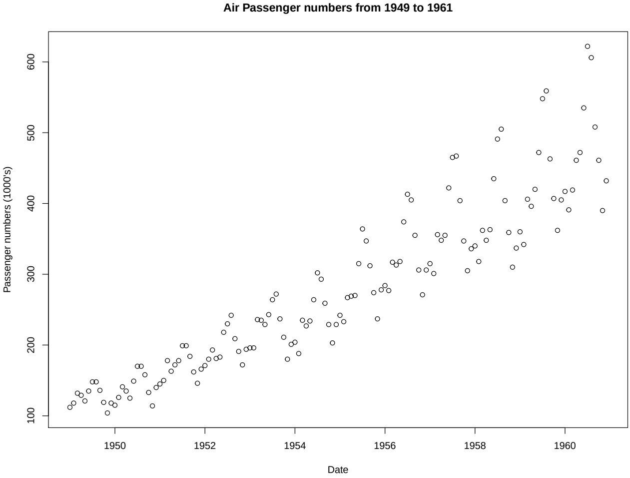
Time Series Analysis 1
DS351

Why can't we use linear regression



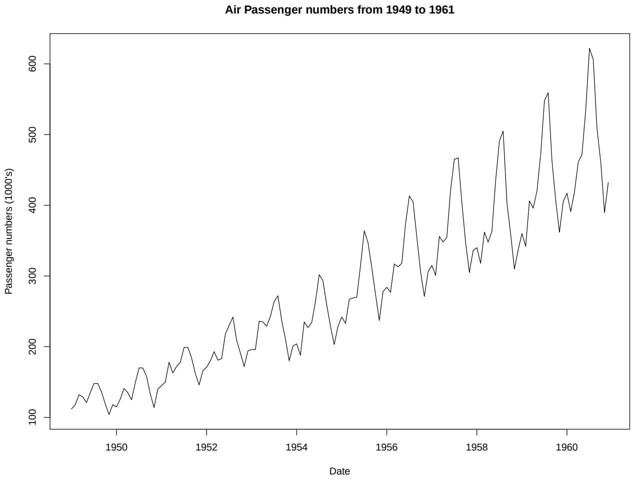
Error terms are correlated.

Why can't we use linear regression



Variance of the errors increases with time.

Why can't we use linear regression



Seasonality, which implies non-linearity!

Analyzing Time Series

Notations

Time series is often denoted by

$$\ldots \quad Y_{t-2} \quad Y_{t-1} \quad Y_t \quad Y_{t+1} \quad Y_{t+2} \quad \ldots$$

time index

Notations

Time series is often denoted by

$$\ldots \quad Y_{t-2} \quad Y_{t-1} \quad Y_t \quad Y_{t+1} \quad Y_{t+2} \quad \ldots$$

time index

Lag is an amount of time passed.

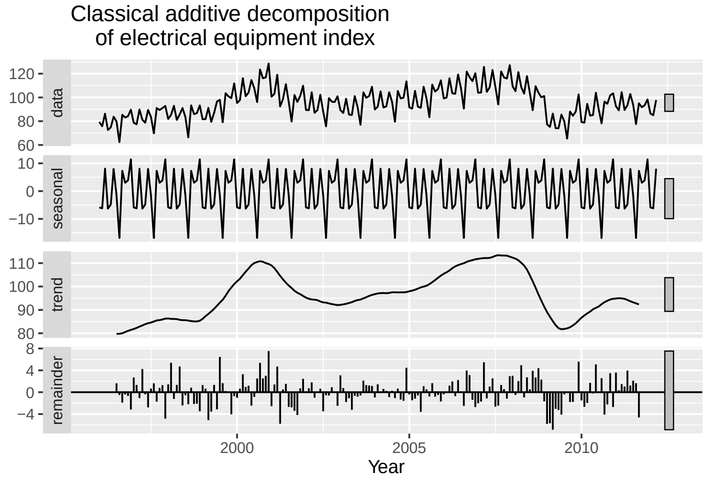
Example: lag 5 of Y_t is y_{t-5}.

Time series decomposition

Time series decomposition

Goal:

- Extract trend seasonality
- Visualize and improve understanding of time series



Classical decomposition

Two types of decomposition:

1. Additive decomposition

$$y_t = S_t + T_t + R_t,$$

where

- S_t is the seasonal component.
- T_t is the trend component.
- R_t is the remainder component.

Classical decomposition

Two types of decomposition:

2. Multiplicative decomposition

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

where

- S_t is the seasonal component.
- T_t is the trend component.
- R_t is the remainder component.

Additive decomposition

$$y_t = S_t + T_t + R_t,$$

Step 1: Estimate the **Trend** $\hat{T}_t$.

Moving average is a method to estimate the trend.

Pick m, usually the seasonal period.

$$\widehat{T}_t = \begin{cases} m\text{-MA} & \text{if } m \text{ is an odd number.} \\ 2 \times m\text{-MA} & \text{if } m \text{ is an even number.} \end{cases}$$

Moving averages

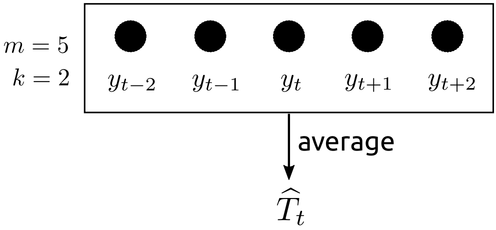
Moving average is a method to estimate the trend.

Time series: $y_t : y_1, y_2, \ldots, y_T$

Moving average of order m of y_t is

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

where $m = 2k + 1$.



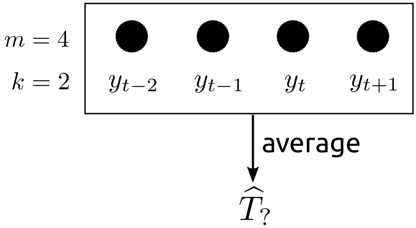
Example: electricity sold to customers in South Australia

Year	Sales (GWh)	5-MA
1989	2354.34	
1990	2379.71	
1991	2318.52	2381.53
1992	2468.99	2424.56
1993	2386.09	2463.76
1994	2569.47	2552.60
1995	2575.72	2627.70
1996	2762.72	2750.62
1997	2844.50	2858.35
⋮	⋮	⋮
1997	2844.50	2858.35
2006	3527.48	3485.43
2007	3637.89	
2008	3655.00	

Example: moving average of different orders

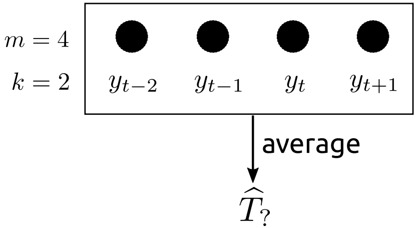
Moving average of even orders

For example, $m = 4$

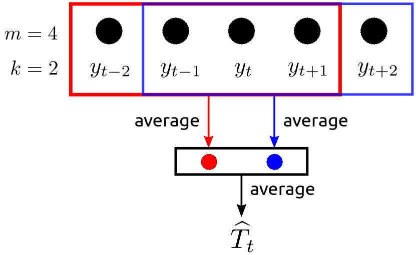


Moving average of even orders

For example, $m = 4$



Idea: use 2-MA **after** 4-MA



Australian quarterly beer production

Year	Quarter	Observation	4-MA	2x4-MA
1992	Q1	443		
1992	Q2	410	451.25	
1992	Q3	420	448.75	450
1992	Q4	532	451.5	450.12
1993	Q1	433	449	450.25
1993	Q2	421	444	446.5
1993	Q3	410	448	446
1993	Q4	512	438	443
1994	Q1	449	441.25	439.62
⋮	⋮	⋮	⋮	⋮
1996	Q3	398	433.75	430.88
1996	Q4	507	433.75	433.75

2×m-MA

The 2×4-MA of y_t is

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

2×m-MA

The 2×4-MA of y_t is

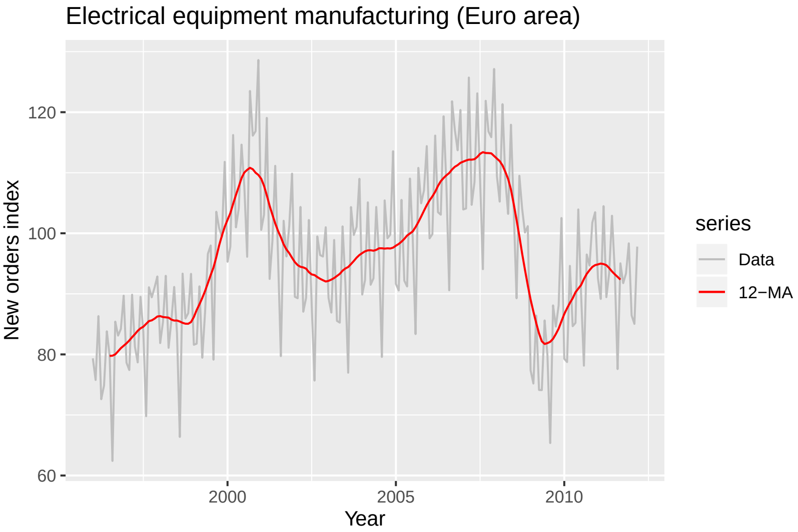
$$\begin{aligned}\widehat{T}_t &= \frac{1}{2}\left[\frac{1}{4}(y_{t-2}+y_{t-1}+y_t+y_{t+1})+\frac{1}{4}(y_{t-1}+y_t+y_{t+1}+y_{t+2})\right]\\ &= \frac{1}{8}y_{t-2}+\frac{1}{4}y_{t-1}+\frac{1}{4}y_t+\frac{1}{4}y_{t+1}+\frac{1}{8}y_{t+2}.\end{aligned}$$

In general, The 2×m-MA of y_t is

$$\widehat{T}_t = \frac{1}{2m}y_{t-k}+\ldots+\frac{1}{m}y_{t-1}+\frac{1}{m}y_t+\frac{1}{m}y_{t+1}+\ldots+\frac{1}{2m}y_{t+k},$$

where $m = 2k$.

Example: monthly data



2×4-MA for quarterly beer production, 7-MA for daily traffic data etc

Step 2: Calculate the detrended series

$$y_t - \widehat{T}_t$$

Additive decomposition

Step 3: Compute the mean of $y_t - \widehat{T}_t$ for each seasonal unit.
For example, for monthly data, we compute

$$
\begin{aligned}
S_1 &= \text{the mean of all values in January} \\
S_2 &= \text{the mean of all values in February} \\
&\quad \text{and so on...}
\end{aligned}
$$

Additive decomposition

Step 3: Compute the mean of $y_t - \widehat{T}_t$ for each seasonal unit.
For example, for monthly data, we compute

$$
\begin{aligned}
S_1 &= \text{the mean of all values in January} \\
S_2 &= \text{the mean of all values in February} \\
&\quad \text{and so on...}
\end{aligned}
$$

Then, these seasonal values are adjusted to have zero mean.

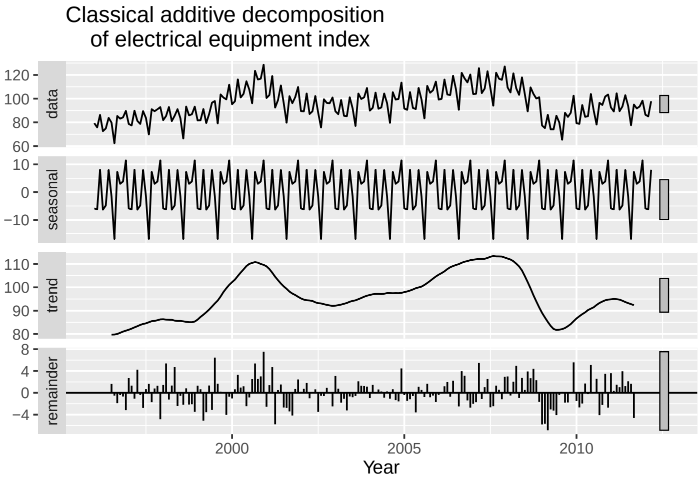
$$
\begin{aligned}
\widehat{S}_1 &= S_1 - \overline{S} \\
\widehat{S}_2 &= S_2 - \overline{S} \\
&\quad \text{and so on...}
\end{aligned}
$$

where $\overline{S} = \frac{1}{12}\sum_{i=1}^{12} S_i$

Additive decomposition

Step 4: The remainder component is

$$\widehat{R}_t = y_t - \widehat{T}_t - \widehat{S}_t.$$



Multiplicative decomposition

- *Step 1*: Pick m, usually the seasonal period.

$$\widehat{T}_t = \begin{cases} m\text{-MA} & \text{if } m \text{ is an odd number.} \\ 2 \times m\text{-MA} & \text{if } m \text{ is an even number.} \end{cases}$$

Multiplicative decomposition

- *Step 1*: Pick m, usually the seasonal period.

$$\widehat{T}_t = \begin{cases} m\text{-MA} & \text{if } m \text{ is an odd number.} \\ 2 \times m\text{-MA} & \text{if } m \text{ is an even number.} \end{cases}$$

- *Step 2*: Calculate the detrended series

$$\frac{y_t}{\widehat{T}_t}.$$

Multiplicative decomposition

- *Step 3*: Compute the mean of $y_t/\widehat{T}_t$ for each seasonal unit. For example, for monthly data, we compute

$$S_1 = \text{the mean of all values in January}$$
$$S_2 = \text{the mean of all values in February}$$
$$\text{and so on...}$$

Multiplicative decomposition

- *Step 3*: Compute the mean of $y_t/\widehat{T}_t$ for each seasonal unit. For example, for monthly data, we compute

$$
\begin{aligned}
S_1 &= \text{the mean of all values in January} \\
S_2 &= \text{the mean of all values in February} \\
&\quad \text{and so on...}
\end{aligned}
$$

Then, these seasonal values are adjusted to have sum of 1 .

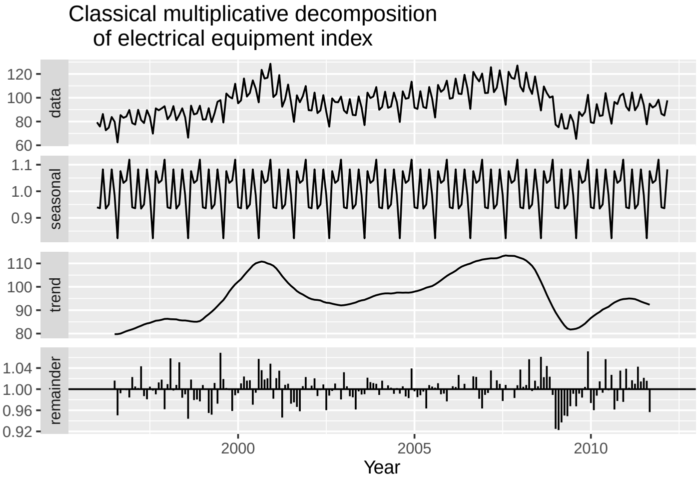
$$
\begin{aligned}
\widehat{S}_1 &= S_1/\overline{S} \\
\widehat{S}_2 &= S_2/\overline{S} \\
&\quad \text{and so on...}
\end{aligned}
$$

where $\overline{S} = \sum_{i=1}^{12} S_i$.

Multiplicative decomposition

- *Step 4*: The remainder component is

$$\widehat{R}_t = \frac{y_t}{\widehat{T}_t \widehat{S}_t}.$$



Strength of trend (Wang, Smith & Hyndman, 2006)

Back to additive decomposition:

$$y_t = T_t + S_t + R_t.$$

Observation: for a time series with strong trend,

$$\frac{\text{Var}(R_t)}{\text{Var}(T_t + R_t)} \quad \text{should be small.}$$

Strength of trend (Wang, Smith & Hyndman, 2006)

Back to additive decomposition:

$$y_t = T_t + S_t + R_t.$$

Observation: for a time series with strong trend,

$$\frac{\mathrm{Var}(R_t)}{\mathrm{Var}(T_t + R_t)} \text{ should be small.}$$

for a time series with strong seasonality,

$$\frac{\mathrm{Var}(R_t)}{\mathrm{Var}(S_t + R_t)} \text{ should be small.}$$

Strength of trend (Wang, Smith & Hyndman, 2006)

So we define the strength of trend as

$$F_T = \max\left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(T_t + R_t)}\right)$$

and the strength of seasonality as

$$F_S = \max\left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(S_t + R_t)}\right).$$

Higher value = Stronger effect

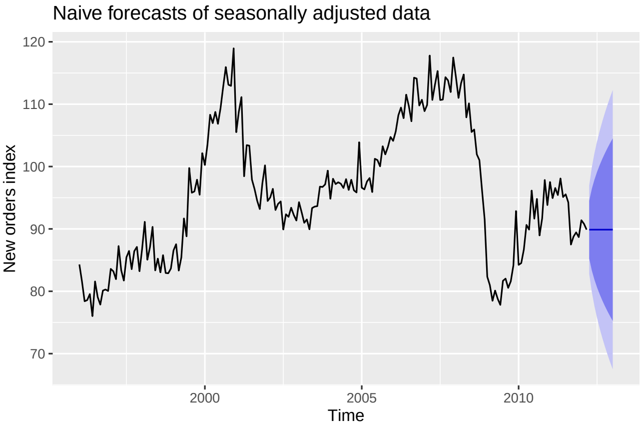
This is useful when we have a collection of time series and we want to find the one with the most trend or seasonality.

Forecasting with decomposition

We can make forecast from the decomposition

$$y_t = \widehat{S}_t + (\widehat{T}_t + \widehat{R}_t),$$

where we can use time series model to forecast the seasonally adjusted component $\widehat{A}_t = \widehat{T}_t + \widehat{R}_t$ and then add back the seasonal component S_t.



Forecasting with decomposition

We can make forecast from the decomposition

$$y_t = \widehat{S}_t + (\widehat{T}_t + \widehat{R}_t),$$

where we can use time series model to forecast the seasonally adjusted component $\widehat{A}_t = \widehat{T}_t + \widehat{R}_t$ and then add back the seasonal component S_t.