MONASH University

MONASH BUSINESS SCHOOL

ETC3550: Applied forecasting for business and economics

Ch6. Time series decomposition

OTexts.org/fpp2/

Outline

Time series patterns

Recall

Trend pattern exists when there is a long-term increase or decrease in the data.

Cyclic pattern exists when data exhibit rises and falls that are *not of fixed period* (duration usually of at least 2 years).

Seasonal pattern exists when a series is influenced by seasonal factors (e.g., the quarter of the year, the month, or day of the week).

Time series decomposition

$$Y_t = f(S_t, T_t, R_t)$$

where

Y_t = data at period t

T_t = trend-cycle component at period t

S_t = seasonal component at period t

R_t = remainder component at period t

Time series decomposition

$$Y_t = f(S_t, T_t, R_t)$$

where
$Y_t =$ data at period t
$T_t =$ trend-cycle component at period t
$S_t =$ seasonal component at period t
$R_t =$ remainder component at period t

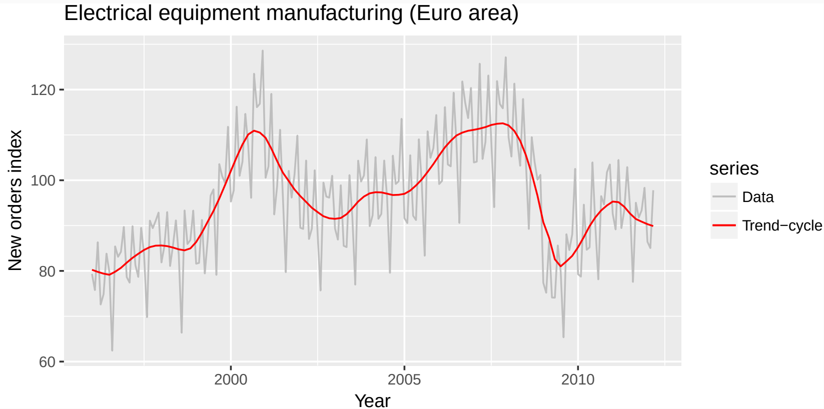
Additive decomposition: $Y_t = S_t + T_t + R_t$.

Multiplicative decomposition: $Y_t = S_t \times T_t \times R_t$.

Time series decomposition

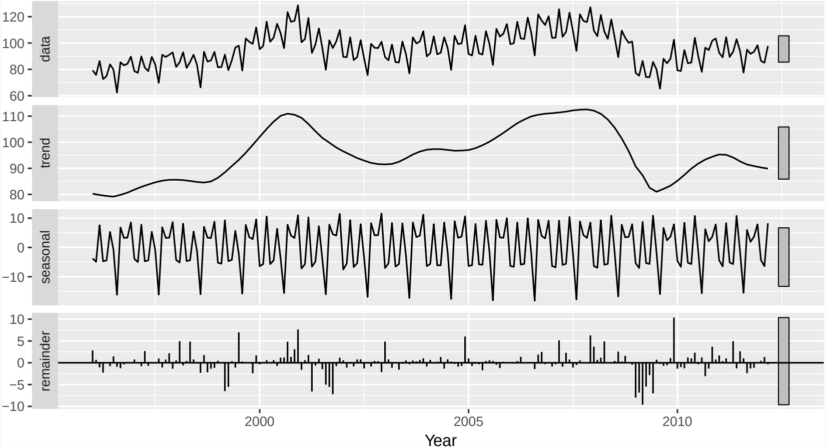
- Additive model appropriate if magnitude of seasonal fluctuations does not vary with level.
- If seasonal are proportional to level of series, then multiplicative model appropriate.
- Multiplicative decomposition more prevalent with economic series
- Alternative: use a Box-Cox transformation, and then use additive decomposition.
- Logs turn multiplicative relationship into an additive relationship:

Euro electrical equipment



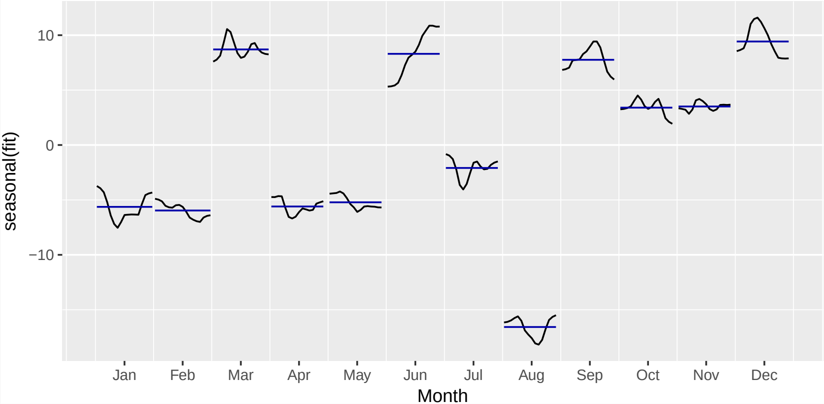
Euro electrical equipment

```
fit <- stl(elecequip, s.window=7)
autoplot(fit)
```



Euro electrical equipment

```
ggsubseriesplot(seasonal(fit))
```



Seasonal adjustment

- Useful by-product of decomposition: an easy way to calculate seasonally adjusted data.
- Additive decomposition: seasonally adjusted data given by

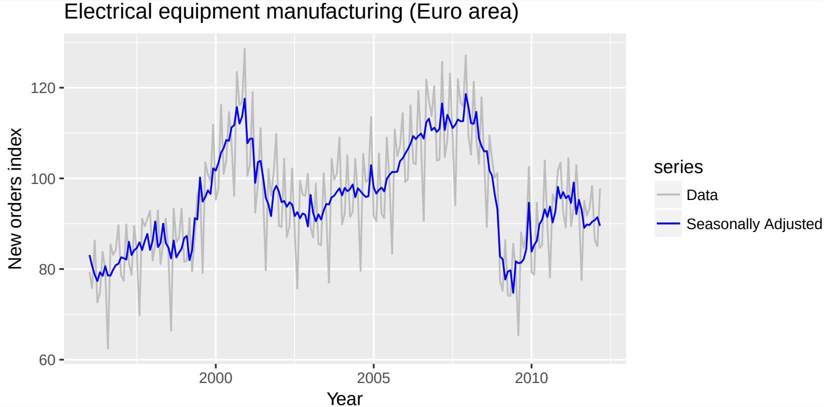
$$Y_t - S_t = T_t + R_t$$

- Multiplicative decomposition: seasonally adjusted data given by

$$Y_t / S_t = T_t \times R_t$$

Euro electrical equipment

```
autoplot(elecequip, series="Data") +
  autolayer(seasadj(fit), series="Seasonally Adjusted")
```



Seasonal adjustment

- We use estimates of S based on past values to seasonally adjust a current value.
- Seasonally adjusted series reflect **remainders** as well as **trend**. Therefore they are not “smooth”" and “downturns”" or “upturns” can be misleading.
- It is better to use the trend-cycle component to look for turning points.

The ABS stuff-up

NEWS ABC

LOCATION:
Clayton, Vic Change

Just In | Australia | World | Business | Sport | Analysis & Opinion | Fact Check | Programs

BREAKING NEWS Police arrest man in connection with stabbing death of 17-year-old Masa Vukotic in M

Print | Email | Facebook | Twitter | More

Treasurer Joe Hockey calls for answers over Australian Bureau of Statistics jobs data

By Michael Vincent and Simon Frazer

Updated 9 Oct 2014, 12:17pm

Federal Treasurer Joe Hockey says he wants answers to the problems the Australian Bureau of Statistics (ABS) has had with unemployment figures.

PHOTO: Joe Hockey says he is unhappy with the volatility of ABS unemployment figures. (AAP: Alan Porritt)

Mr Hockey, who is in the US to discuss Australia's G20 agenda, said last month's unemployment figures were "extraordinary".

The rate was 6.1 per cent after jumping to a 12-year high of 6.4 per cent the previous month.

The ABS has now taken the rare step of abandoning seasonal adjustment for its latest employment data.

The ABS stuff-up

NEWS

LOCATION:
Clayton, Vic Change

Just In | Australia | World | Business | Sport | Analysis & Opinion | Fact Check | Programs | M

BREAKING NEWS Police arrest man in connection with stabbing death of 17-year-old Masa Vukotic in Mel

Print | Email | Facebook | Twitter | More

ABS abandons seasonal adjustment for latest jobs data

By business reporter Michael Janda

Updated 8 Oct 2014, 4:19pm

The Australian Bureau of Statistics is taking the rare step of abandoning seasonal adjustment for its latest employment data.

The ABS uses seasonal adjustment, based on historical experience, to account for the normal variation between hiring and firing patterns between different months.

However, after a winter where the seasonally adjusted unemployment rate swung wildly from 6.1 to 6.4 and back to 6.1 per cent, the bureau released a statement saying it will not adjust the original figure for September for seasonal factors.

It will also reset the seasonal adjustment for July and August to one, meaning that these months will

Sorry, this video has expired

VIDEO: Westpac chief economist Bill Evans discusses the ABS jobs data changes (ABC News)

RELATED STORY: Doubt the record breaking jobs figures? So does the ABS

RELATED STORY: Jobs increase record sees unemployment slashed

RELATED STORY: Unemployment surges to 12-year high at 6.4 pc

MAP: Australia

The ABS stuff-up

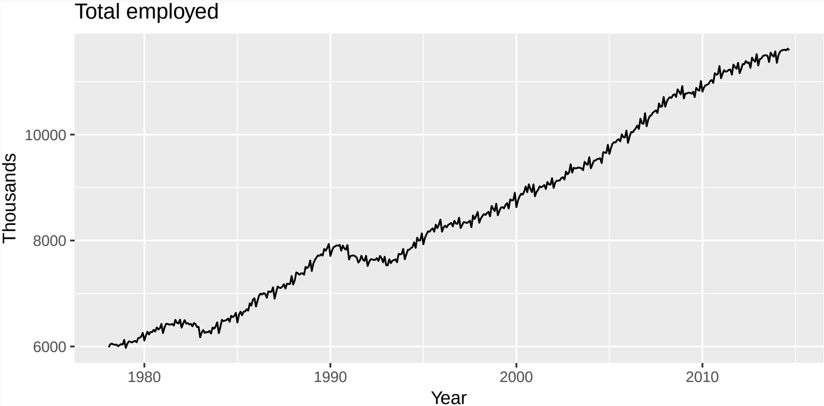
ABS jobs and unemployment figures – key questions answered by an expert

A professor of statistics at Monash University explains exactly what is seasonal adjustment, why it matters and what went wrong in the July and August figures

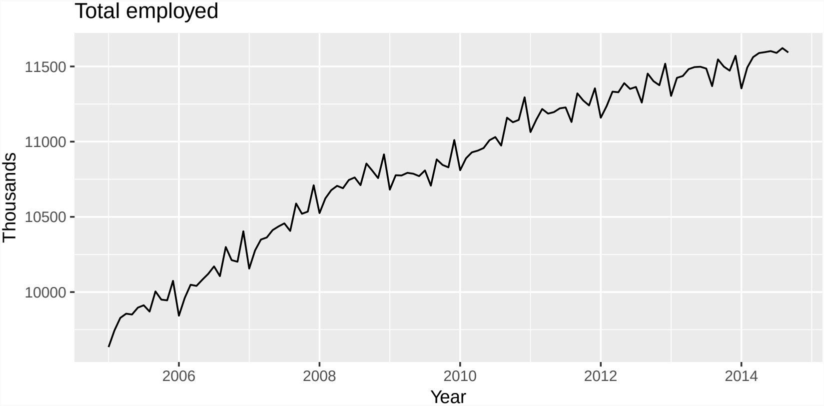
School leavers come on to the jobs market at the same time, causing a seasonal fluctuation. Photograph: Brian Snyder/Reuters

The Australian Bureau of Statistics has retracted its seasonally adjusted employment data for July and August, which recorded huge swings in the jobless rate. The ABS is also planning to review the methods it uses for seasonal adjustment to ensure its figures are as accurate as possible. Rob Hyndman, a professor of statistics at Monash University and member of the bureau's

The ABS stuff-up

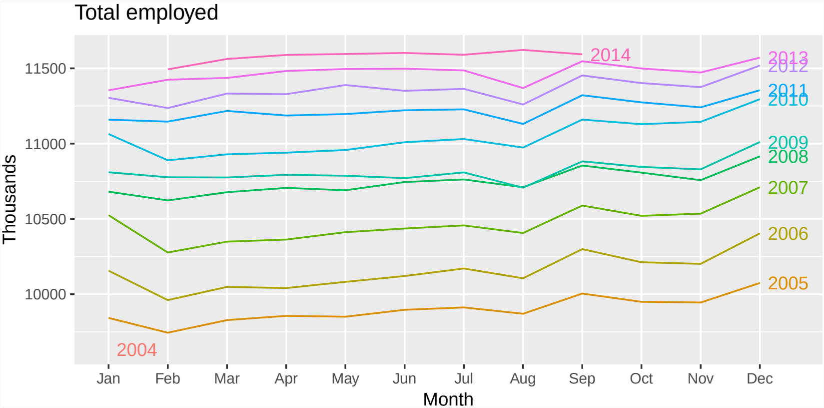


The ABS stuff-up

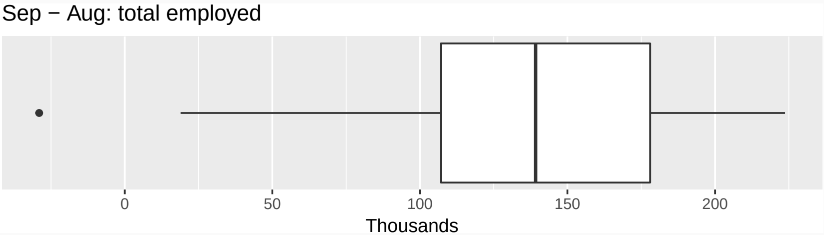


The ABS stuff-up

```
ggseasonplot(window(x,start=c(2005,1)), year.labels=TRU
  ggtitle("Total employed") + ylab("Thousands")
```

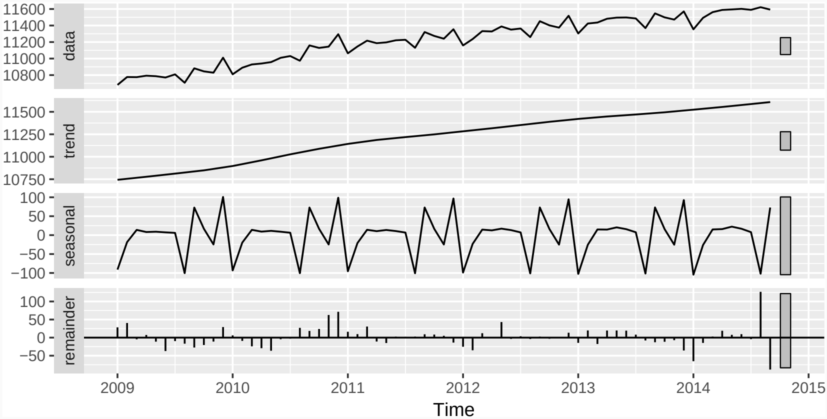


The ABS stuff-up

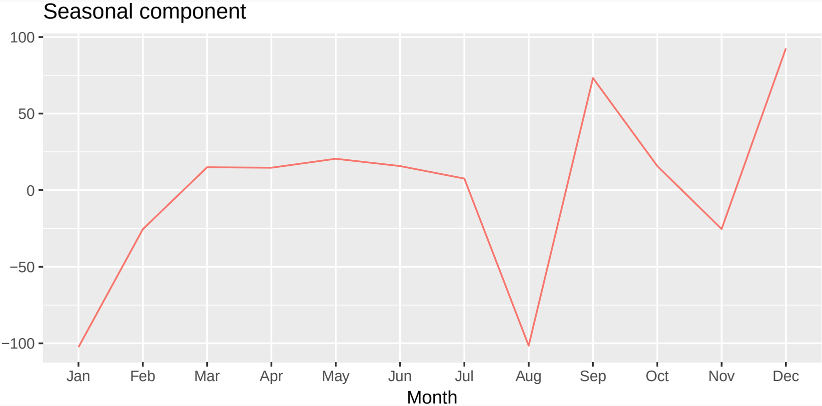


The ABS stuff-up

```
x %>% window(start=2009) %>%
  stl(s.window=11, robust=TRUE) -> fit
autoplot(fit)
```

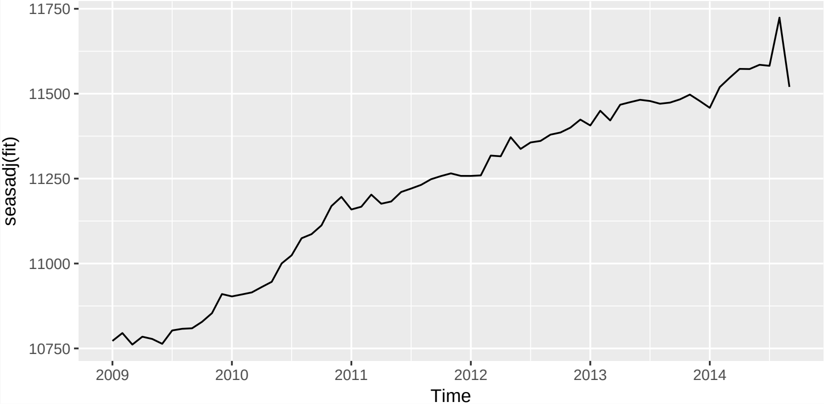


The ABS stuff-up



The ABS stuff-up

```r
autoplot(seasadj(fit))
```



The ABS stuff-up

- August 2014 employment numbers higher than expected.
- Supplementary survey usually conducted in August for employed people.
- Most likely, some employed people were claiming to be unemployed in August to avoid supplementary questions.
- Supplementary survey not run in 2014, so no motivation to lie about employment.
- In previous years, seasonal adjustment fixed the problem.
- The ABS has now adopted a new method to

Outline

Moving averages

The simplest estimate of the trend-cycle uses **moving averages**.

Moving averages

The simplest estimate of the trend-cycle uses **moving averages**.

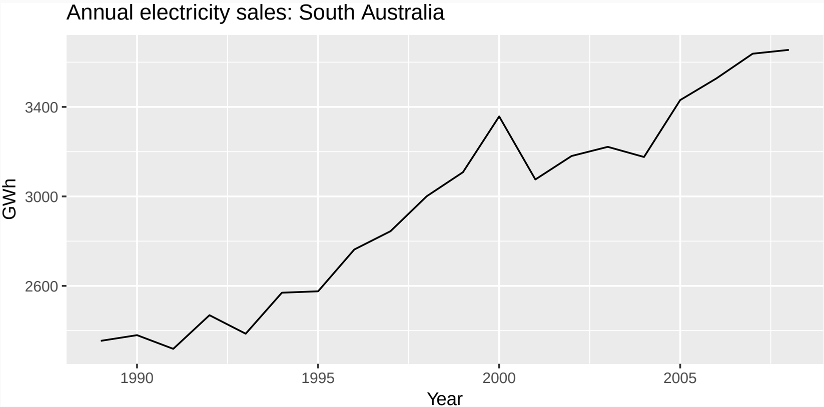
k MA

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

where $m = 2k + 1$.

Moving averages

```r
autoplot(elecsales) + xlab("Year") + ylab("GWh") +
  ggtitle("Annual electricity sales: South Australia")
```



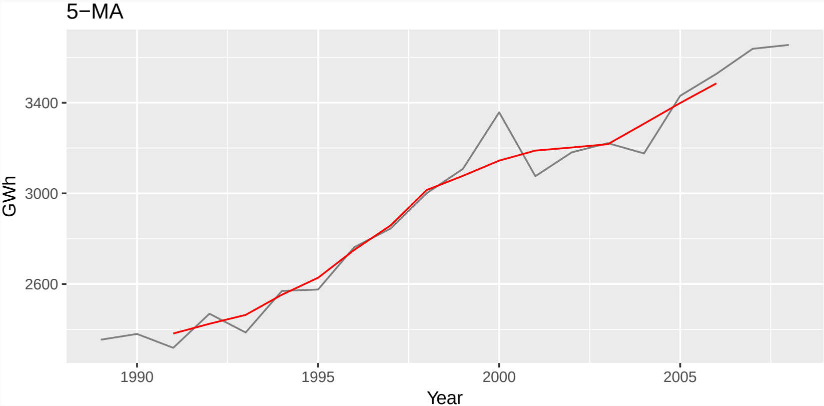
Moving averages

```r
ma5 <- ma(elecsales, 5)
```

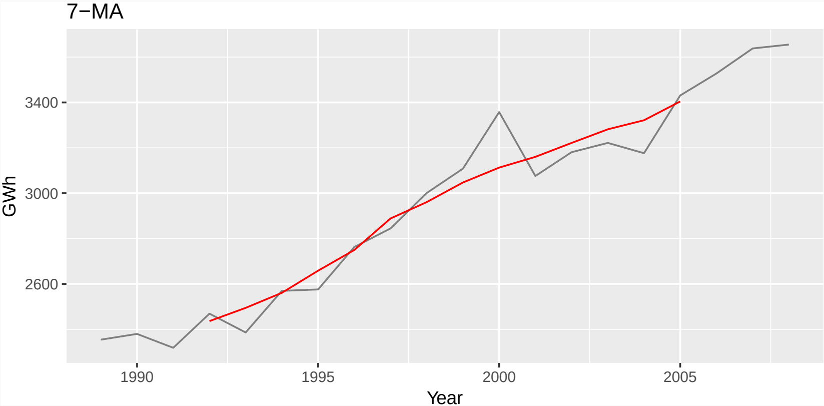
Year	Sales (GWh)	5-MA
1989	2354.34	
1990	2379.71	
1991	2318.52	2381.53
1992	2468.99	2424.56
1993	2386.09	2463.76
1994	2569.47	2552.60
1995	2575.72	2627.70
1996	2762.72	2750.62

Moving averages

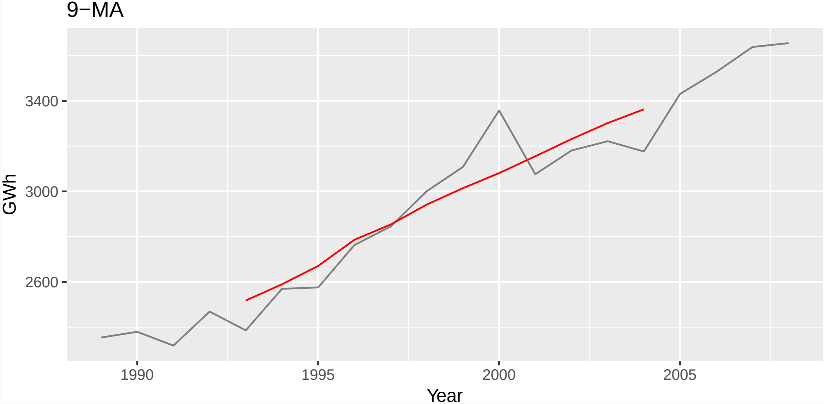
Moving averages



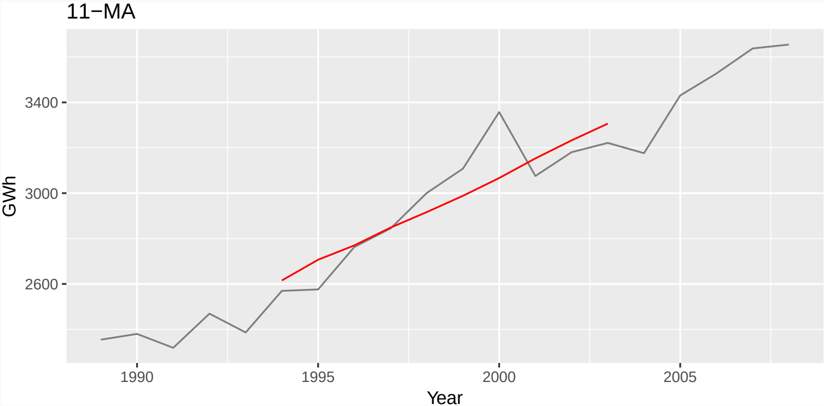
Moving averages



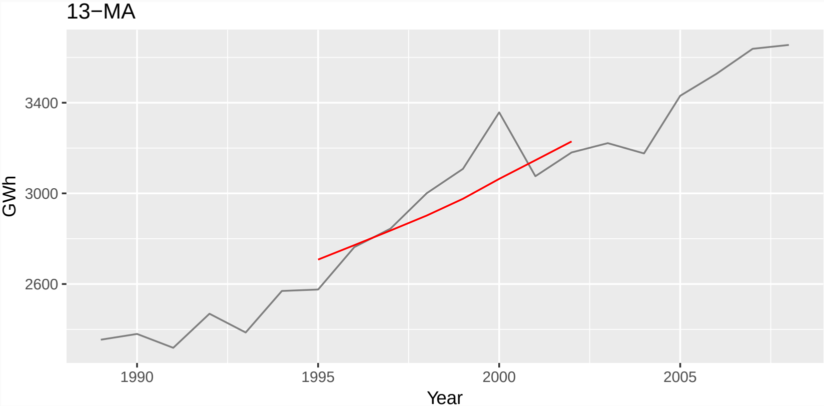
Moving averages



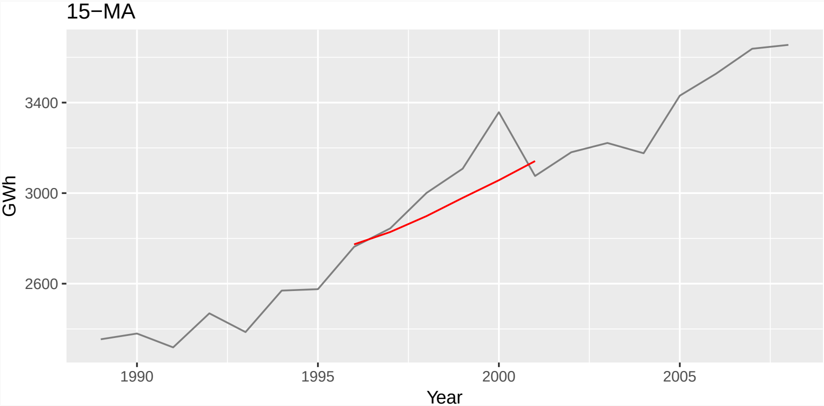
Moving averages



Moving averages



Moving averages



Moving averages

So a moving average is an **average of nearby points**

- observations nearby in time are also likely to be **close in value**.
- average eliminates some **randomness** in the data, leaving a smooth trend-cycle component.

Moving averages

So a moving average is an **average of nearby points**

- observations nearby in time are also likely to be **close in value**.
- average eliminates some **randomness** in the data, leaving a smooth trend-cycle component.

$$\text{3-MA:} \quad \hat{T}_t = (Y_{t-1} + Y_t + Y_{t+1})/3$$

$$\text{5-MA:} \quad \hat{T}_t = (Y_{t-2} + Y_{t-1} + Y_t + Y_{t+1} + Y_{t+2})/5$$

- each average computed by dropping **oldest** observation and including **next** observation.
- averaging **moves** through time series until

Endpoints

Why is there no estimate at ends

- For a 3 MA, there cannot be estimates at time 1 or time n because the observations at time 0 and $n + 1$ are not available.
- Generally: there cannot be estimates at times near the endpoints.

Endpoints

Why is there no estimate at ends

- For a 3 MA, there cannot be estimates at time 1 or time n because the observations at time 0 and $n + 1$ are not available.
- Generally: there cannot be estimates at times near the endpoints.

The order of the MA

- larger order means smoother, flatter curve
- larger order means more points lost at ends
- order = length of season or cycle removes pattern

Centered MA

4 MA:

$$\frac{1}{4}(Y_{t-2} + Y_{t-1} + Y_t + Y_{t+1})$$

or

$$\frac{1}{4}(Y_{t-1} + Y_t + Y_{t+1} + Y_{t+2})$$

Centered MA

4 MA:

$$\frac{1}{4}(Y_{t-2} + Y_{t-1} + Y_t + Y_{t+1})$$

or

$$\frac{1}{4}(Y_{t-1} + Y_t + Y_{t+1} + Y_{t+2})$$

Solution: take a further 2-MA to “centre” result.

Centered MA

4 MA:

$$\frac{1}{4}(Y_{t-2} + Y_{t-1} + Y_t + Y_{t+1})$$

or

$$\frac{1}{4}(Y_{t-1} + Y_t + Y_{t+1} + Y_{t+2})$$

Solution: take a further 2-MA to "centre" result.

$$\begin{aligned} T_t &= \frac{1}{2}\left\{\frac{1}{4}(Y_{t-2} + Y_{t-1} + Y_t + Y_{t+1}) \right. \\ &\qquad \left. + \frac{1}{4}(Y_{t-1} + Y_t + Y_{t+1} + Y_{t+2})\right\} \\ &= \frac{1}{8}Y_{t-2} + \frac{1}{4}Y_{t-1} + \frac{1}{4}Y_t + \frac{1}{4}Y_{t+1} + \frac{1}{8}Y_{t+2} \end{aligned}$$

Centered MA

Year	Data	4-MA	$2 \times$ 4-MA
1992 Q1	443.00		
1992 Q2	410.00		
		451.25	
1992 Q3	420.00		450.00
		448.75	
1992 Q4	532.00		450.12
		451.50	
1993 Q1	433.00		450.25
		449.00	
1993 Q2	421.00		446.50
		444.00	
1993 Q3	410.00		446.00
		448.00	
1993 Q4	512.00		443.00
		438.00	
⋮	⋮	⋮	⋮

Moving average trend

A moving average of the same length as the season removes the seasonal pattern.

For quarterly data: use a 2×4 MA

For monthly data: use a 2×12 MA

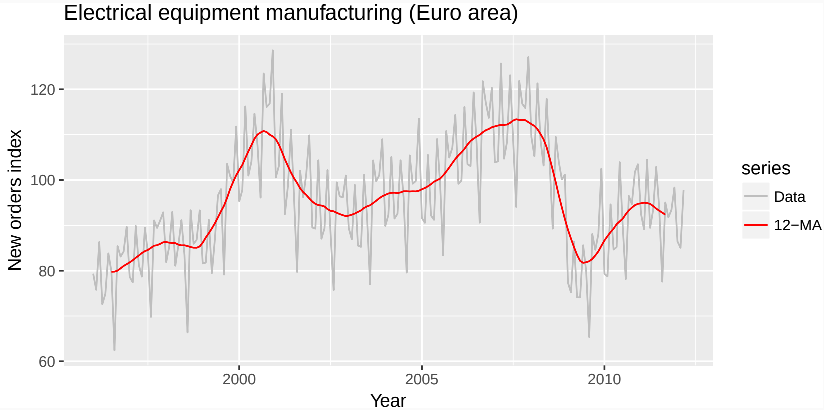
$$\hat{T}_t = \frac{1}{24}Y_{t-6} + \frac{1}{12}Y_{t-5} + \cdots + \frac{1}{12}Y_{t+5} + \frac{1}{24}Y_{t+6}$$

In R:

```
ma(x, order=4, centered=TRUE)
```

- `centered=TRUE` is the default.
- centering makes no difference for odd orders.

Moving average trend



Weighted moving averages

Weighted MA

$$T_t = \sum_{j=-k}^{k} a_j Y_{t+j},$$

where $k = (m-1)/2$ is the half-width and the weights are denoted by $[a_{-k}, \ldots, a_k]$.

Weighted moving averages

Weighted MA

$$T_t = \sum_{j=-k}^{k} a_j Y_{t+j},$$

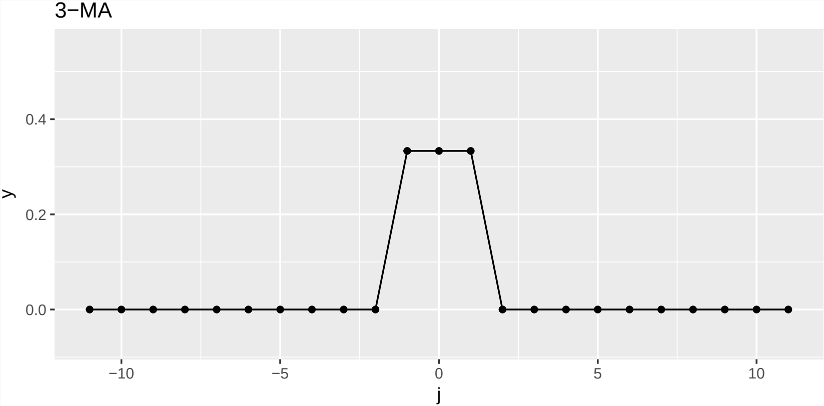
where $k = (m-1)/2$ is the half-width and the weights are denoted by $[a_{-k}, \dots, a_k]$.

- Simple m-MA: all weights equal to $1/m$.
- Require sum of $a_j = 1$ and $a_j = a_{-j}$.
- Weighted MA are smoother.

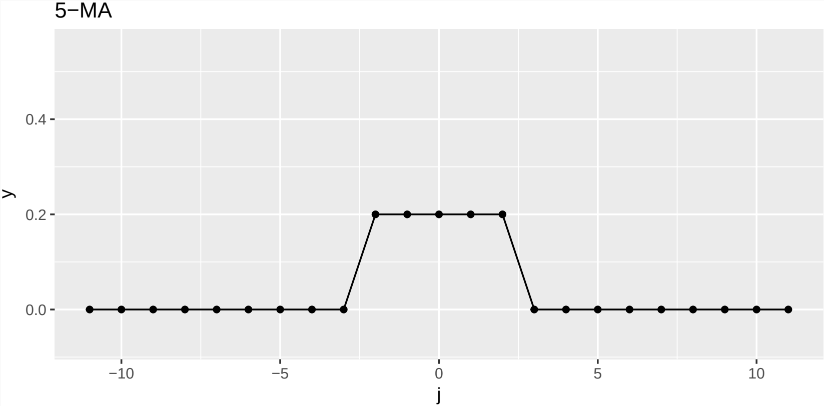
Weighted moving averages

	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_1
3-MA	0.333	0.333										
5-MA	0.200	0.200	0.200									
2x12-MA	0.083	0.083	0.083	0.083	0.083	0.083	0.042					
3x3-MA	0.333	0.222	0.111									
3x5-MA	0.200	0.200	0.133	0.067								
S15-MA	0.231	0.209	0.144	0.066	0.009	-0.016	-0.019	-0.009				
S21-MA	0.171	0.163	0.134	0.094	0.051	0.017	-0.006	-0.014	-0.014	-0.009	-0.003	
H5-MA	0.558	0.294	-0.073									
H9-MA	0.330	0.267	0.119	-0.010	-0.041							
H13-MA	0.240	0.214	0.147	0.066	0.000	-0.028	-0.019					
H23-MA	0.148	0.138	0.122	0.097	0.068	0.039	0.013	-0.005	-0.015	-0.016	-0.011	-0.00

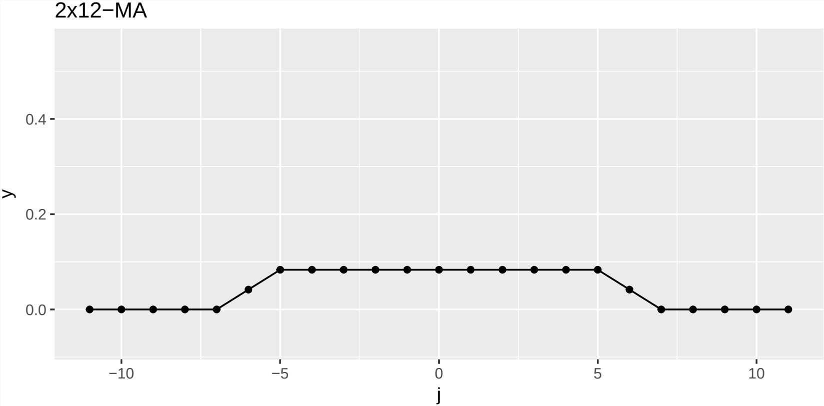
Weighted MA



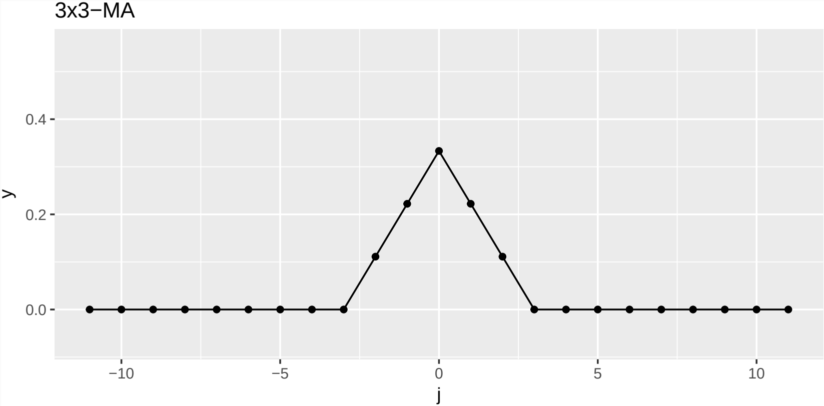
Weighted MA



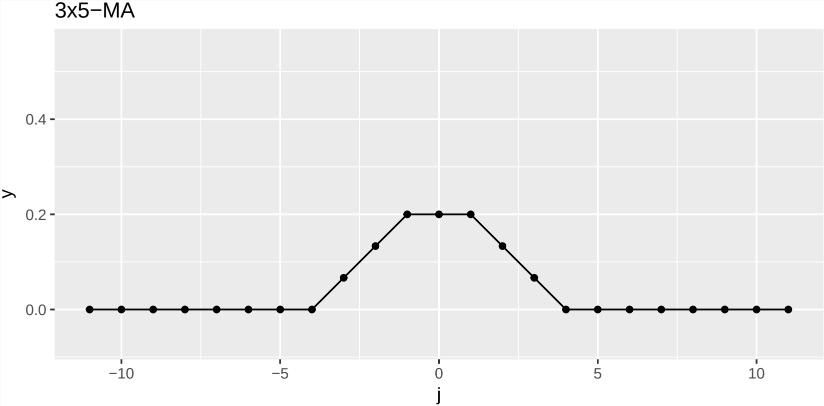
Weighted MA



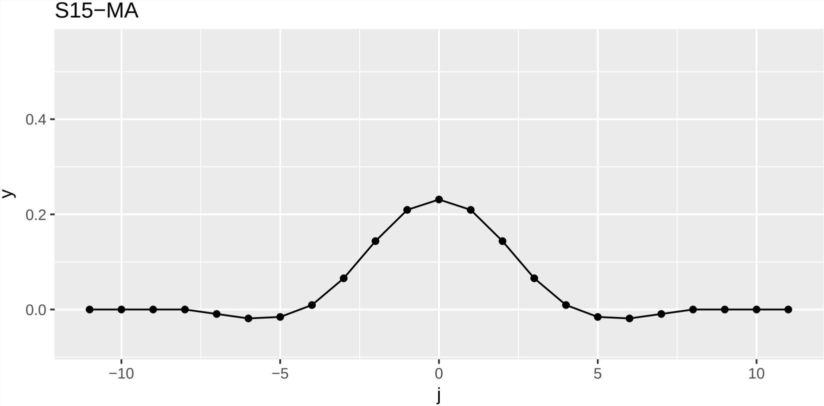
Weighted MA



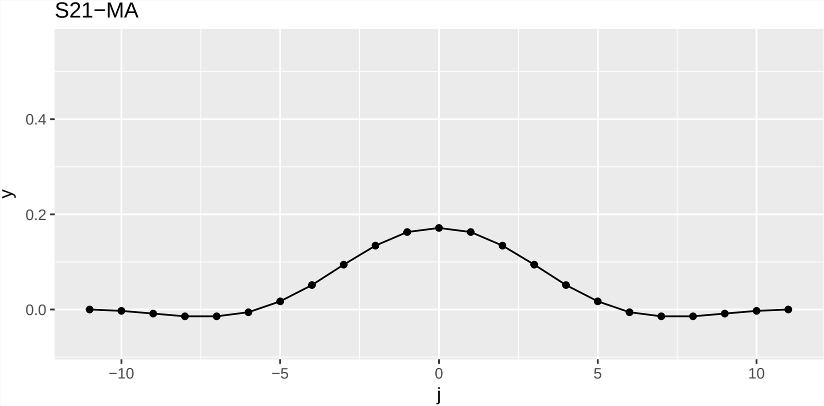
Weighted MA



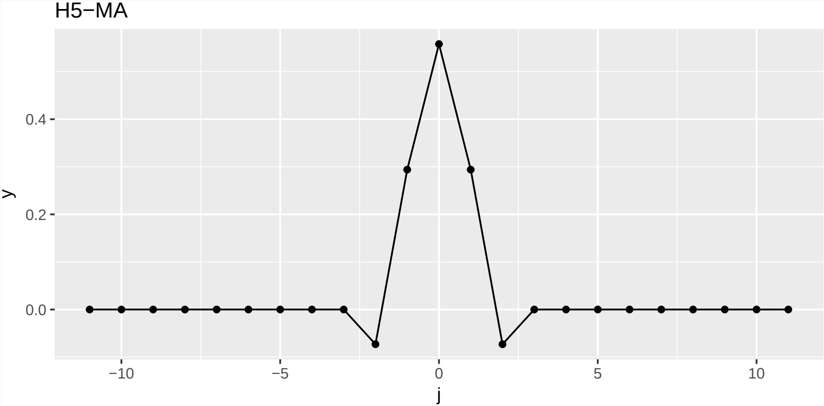
Weighted MA



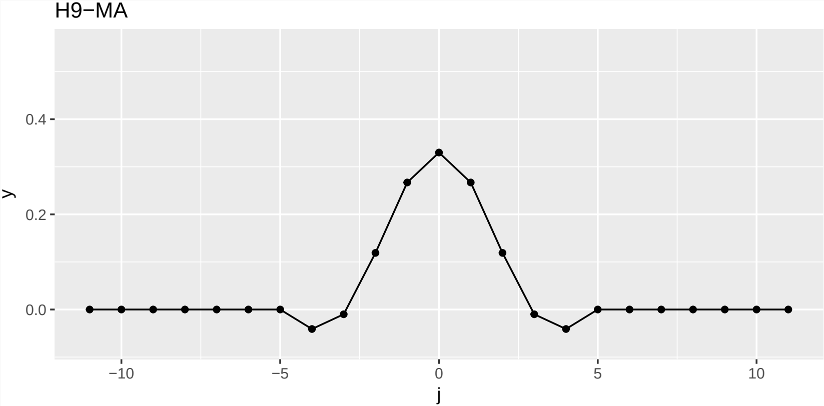
Weighted MA



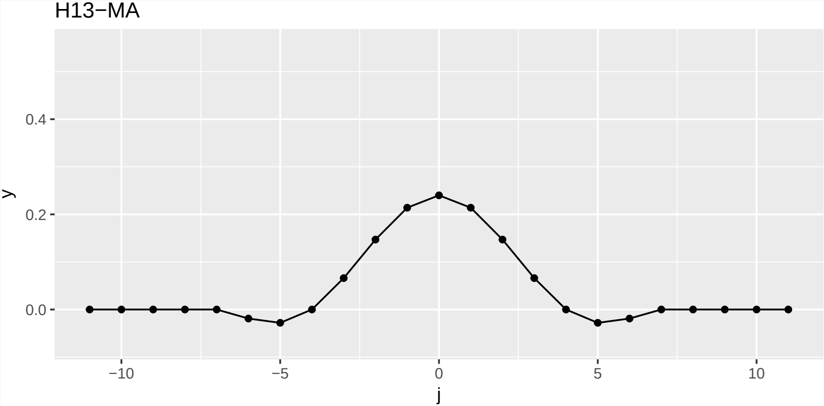
Weighted MA



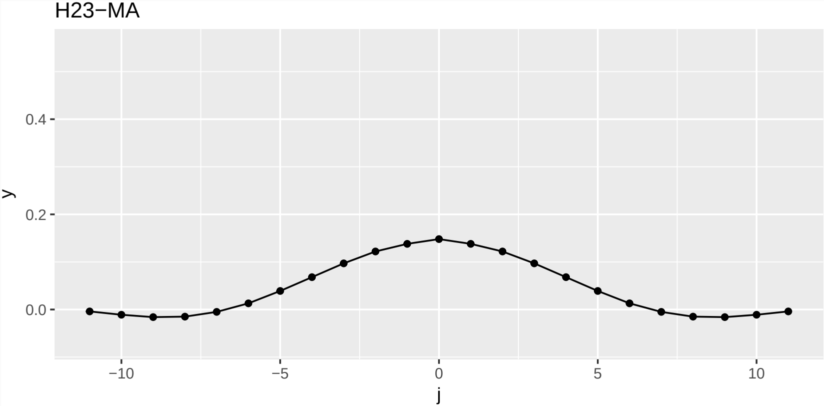
Weighted MA



Weighted MA



Weighted MA



Outline

Trend-cycle

Multiplicative decomposition: $Y_t = T_t S_t R_t = \hat{T}_t \hat{S}_t \hat{R}_t$

Additive decomposition: $Y_t = T_t + S_t + R_t = \hat{T}_t + \hat{S}_t + \hat{R}_t$

Trend-cycle

Multiplicative decomposition: $Y_t = T_t S_t R_t = \hat{T}_t \hat{S}_t \hat{R}_t$

Additive decomposition: $Y_t = T_t + S_t + R_t = \hat{T}_t + \hat{S}_t + \hat{R}_t$

- Estimate $\hat{T}$ using $(2 \times m)$-MA if m is even. Otherwise, estimate $\hat{T}$ using m-MA
- Compute de-trended series
  - Multiplicative decomposition: $y_t/\hat{T}_t$
  - Additive decomposition: $y_t - \hat{T}_t$

Trend-cycle

Multiplicative decomposition: $Y_t = T_t S_t R_t = \hat{T}_t \hat{S}_t \hat{R}_t$

Additive decomposition: $Y_t = T_t + S_t + R_t = \hat{T}_t + \hat{S}_t + \hat{R}_t$

- Estimate $\hat{T}$ using $(2 \times m)$-MA if m is even. Otherwise, estimate $\hat{T}$ using m-MA
- Compute de-trended series
  - Multiplicative decomposition: $y_t / \hat{T}_t$
  - Additive decomposition: $y_t - \hat{T}_t$

De-trending

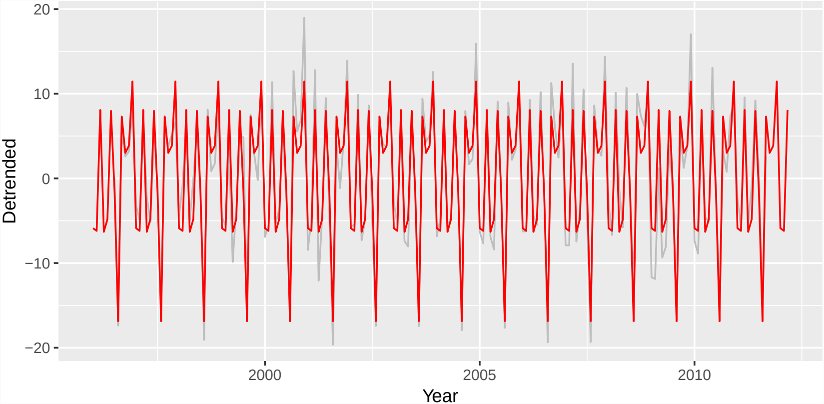
Remove smoothed series $\hat{T}_t$ from Y to leave S and E.

- Multiplicative model: $\frac{Y}{\hat{T}} = \frac{\hat{T} \times \hat{S} \times \hat{R}}{\hat{T}} = \hat{S} \times \hat{R}$
- Additive model: $Y - \hat{T} = (\hat{T} + \hat{S} + \hat{R}) - \hat{T} = \hat{S} + \hat{R}$

Estimating seasonal component

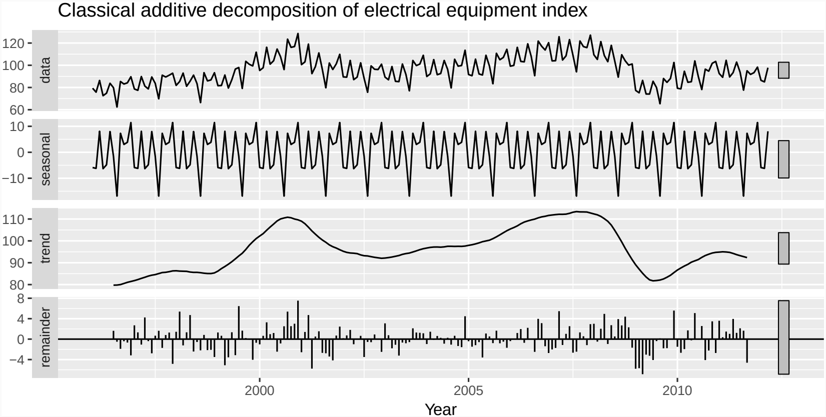
- Seasonal index for each quarter is estimated as an **average** of the detrended series for that quarter of successive years.
- If necessary, adjust the seasonal indices so they add to m (multiplicative) or 0 (additive).
- The seasonal component $\hat{S}_t$ simply consists of replications of the seasonal indices.

Seasonal patterns



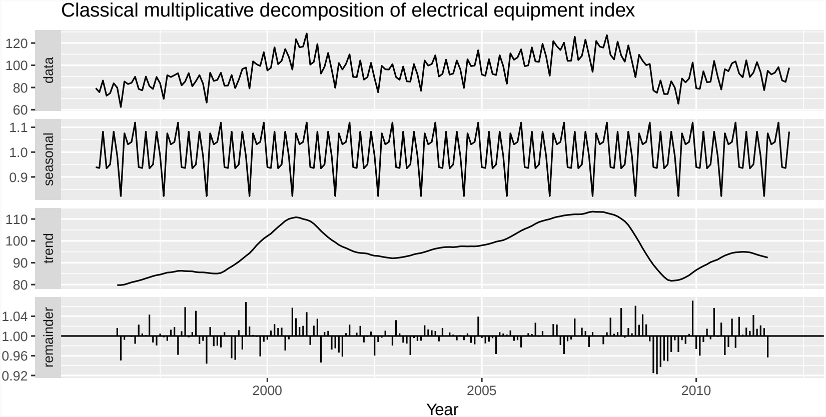
Seasonal patterns

```r
fit <- decompose(elecequip, type="additive")
autoplot(fit)
```



Seasonal patterns

```
fit <- decompose(elecequip, type="multiplicative")
autoplot(fit)
```



Remainder component

Multiplicative decomposition: $\hat{R}_t = Y_t/(\hat{T}_t\hat{S}_t)$

Additive decomposition: $\hat{R}_t = Y_t - \hat{T}_t - \hat{S}_t$

Classical decomposition

- Choose additive or multiplicative depending on which gives the most stable components.
- For multiplicative model, this method of estimation is known as **ratio-to-moving-average method**
- In R: `decompose()`.

Helper functions

- `seasonal()` extracts the seasonal component
- `trendcycle()` extracts the trend-cycle component
- `remainder()` extracts the remainder

Comments on classical decomposition

- Estimate of trend is unavailable for first few and last few observations.
- Seasonal component repeats from year to year. May not be realistic.
- Not robust to outliers.
- Newer methods designed to overcome these problems.

History of time series decomposition

- Classical method originated in 1920s.
- Census II method introduced in 1957. Basis for X-11 method and variants (including X-12-ARIMA, X-13-ARIMA)
- STL method introduced in 1983
- TRAMO/SEATS introduced in 1990s.

History of time series decomposition

- Classical method originated in 1920s.
- Census II method introduced in 1957. Basis for X-11 method and variants (including X-12-ARIMA, X-13-ARIMA)
- STL method introduced in 1983
- TRAMO/SEATS introduced in 1990s.

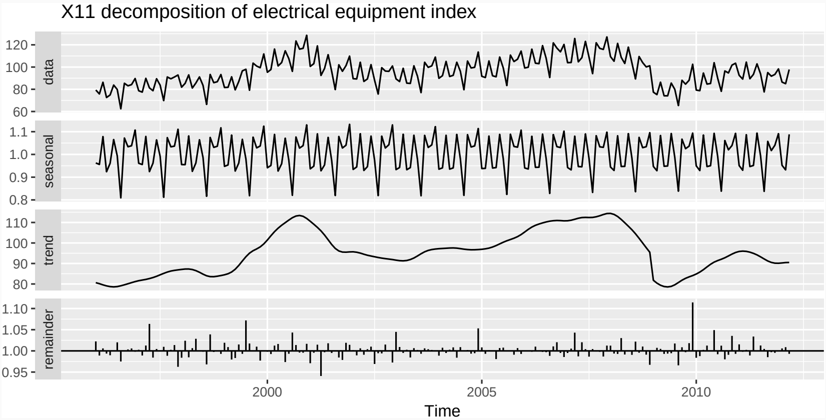
National Statistics Offices

- ABS uses X-12-ARIMA
- US Census Bureau uses X-13-ARIMA-SEATS
- Statistics Canada uses X-12-ARIMA
- ONS (UK) uses X-12-ARIMA

Outline

X-11 decomposition

```r
library(seasonal)
fit <- seas(elecequip, x11="")
autoplot(fit)
```



(Dis)advantages of X-11

Advantages

- Smoother trend estimate
- Allows estimates at end points
- Relatively robust to outliers
- Allows changing seasonality
- Very widely tested on economic data over a long period of time.

(Dis)advantages of X-11

Advantages

- Smoother trend estimate
- Allows estimates at end points
- Relatively robust to outliers
- Allows changing seasonality
- Very widely tested on economic data over a long period of time.

Disadvantages

- No prediction/confidence intervals
- Ad hoc method with no underlying model
- Only developed for quarterly and monthly data

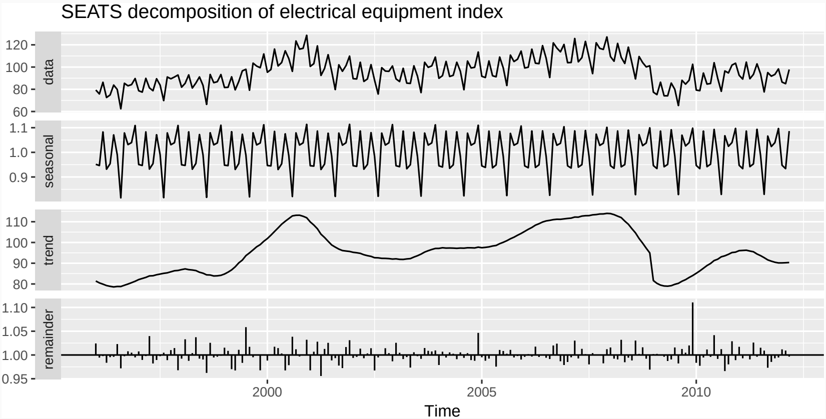
Extensions: X-12-ARIMA and X-13-ARIMA

- The X-11, X-12-ARIMA and X-13-ARIMA methods are based on Census II decomposition.
- These allow adjustments for trading days and other explanatory variables.
- Known outliers can be omitted.
- Level shifts and ramp effects can be modelled.
- Missing values estimated and replaced.
- Holiday factors (e.g., Easter, Labour Day) can be estimated.

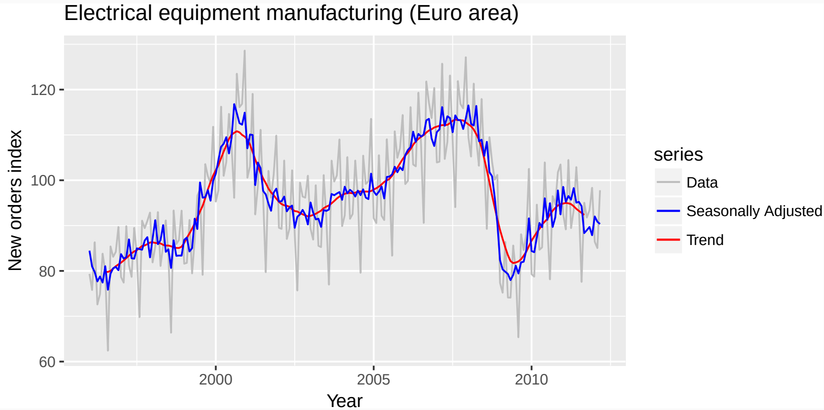
Outline

SEATS decomposition

```r
library(seasonal)
fit <- seas(elecequip)
autoplot(fit)
```

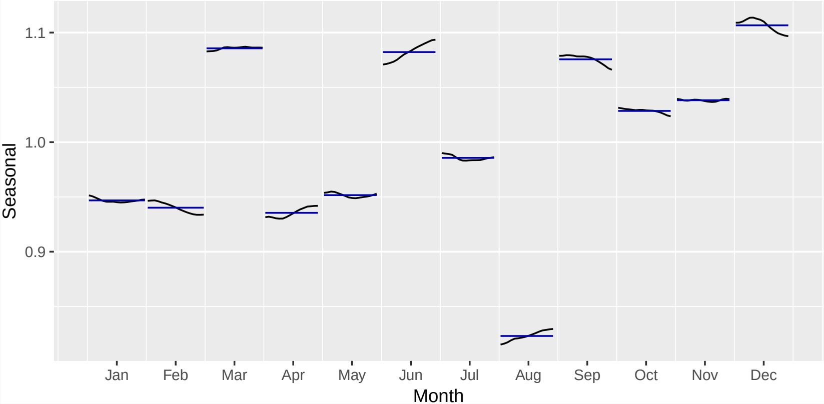


SEATS decomposition



SEATS decomposition

```
ggsubseriesplot(seasonal(fit)) + ylab("Seasonal")
```



(Dis)advantages of SEATS

Advantages

- Model-based
- Smooth trend estimate
- Allows estimates at end points
- Allows changing seasonality
- Developed for economic data

(Dis)advantages of SEATS

Advantages

- Model-based
- Smooth trend estimate
- Allows estimates at end points
- Allows changing seasonality
- Developed for economic data

Disadvantages

- Only developed for quarterly and monthly data

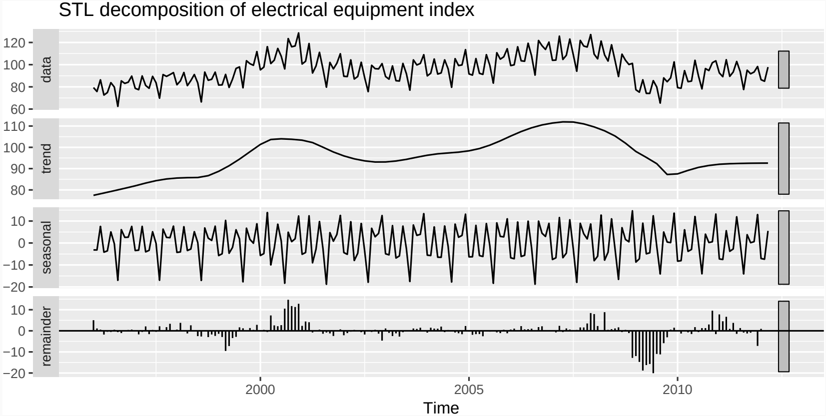
Outline

STL decomposition

- STL: “Seasonal and Trend decomposition using Loess”,
- Very versatile and robust.
- Unlike X-12-ARIMA, STL will handle any type of seasonality.
- Seasonal component allowed to change over time, and rate of change controlled by user.
- Smoothness of trend-cycle also controlled by user.
- Robust to outliers
- Not trading day or calendar adjustments.
- Only additive

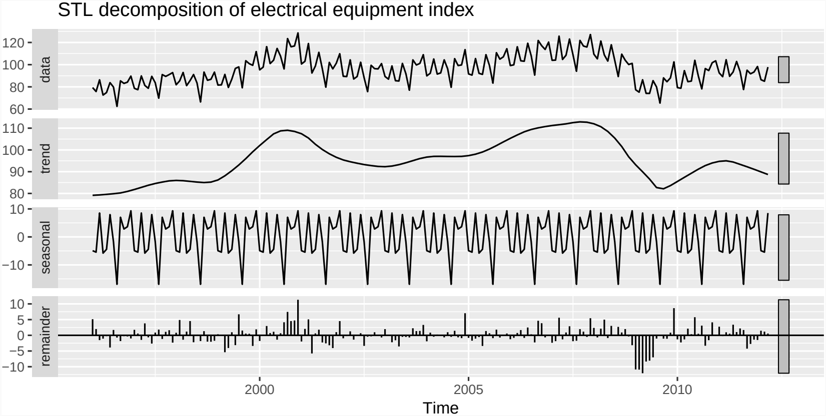
STL decomposition

```r
fit <- stl(elecequip, s.window=5, robust=TRUE)
autoplot(fit) +
  ggtitle("STL decomposition of electrical equipment in
```



STL decomposition

```
fit <- stl(elecequip, s.window="periodic", robust=TRUE)
autoplot(fit) +
  ggtitle("STL decomposition of electrical equipment in
```



STL decomposition

```
stl(elecequip,s.window=5)

stl(elecequip, t.window=15,
  s.window="periodic", robust=TRUE)
```

- `t.window` controls wiggliness of trend component.
- `s.window` controls variation on seasonal component.

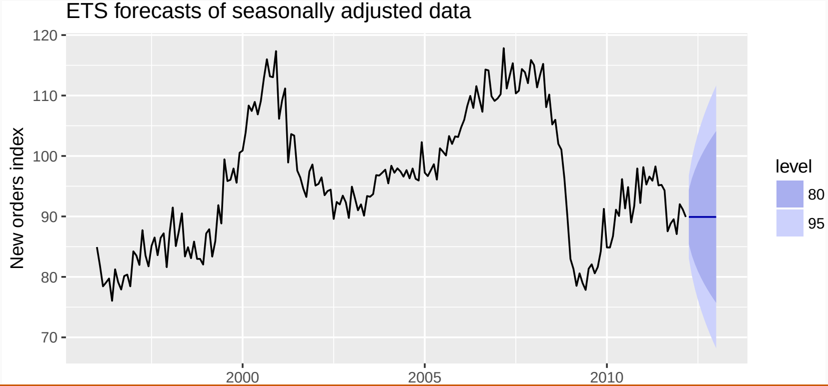
Outline

Forecasting and decomposition

- Forecast seasonal component by repeating the last year
- Forecast seasonally adjusted data using non-seasonal time series method. E.g., ETS or ARIMA
  - Holt's method — next topic
  - Random walk with drift model
- Combine forecasts of seasonal component with forecasts of seasonally adjusted data to get forecasts of original data.
- Sometimes a decomposition is useful just for understanding the data before building a

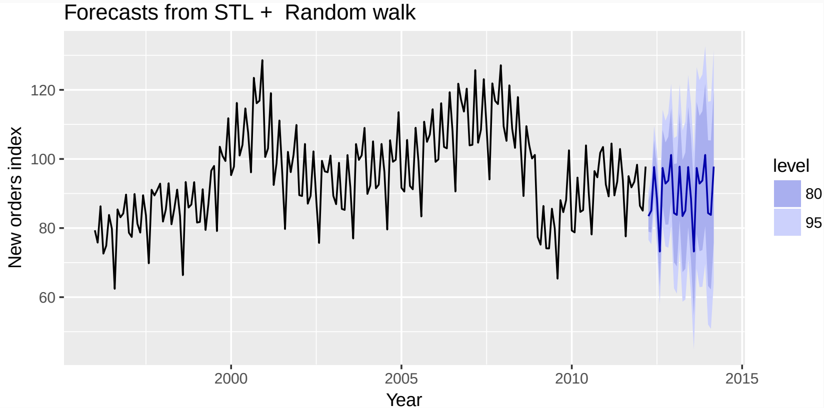
Electrical equipment

```r
fit <- stl(elecequip, t.window=13, s.window="periodic")
fit %>% seasadj %>% naive %>%
  autoplot() + ylab("New orders index") +
  ggtitle("ETS forecasts of seasonally adjusted data")
```



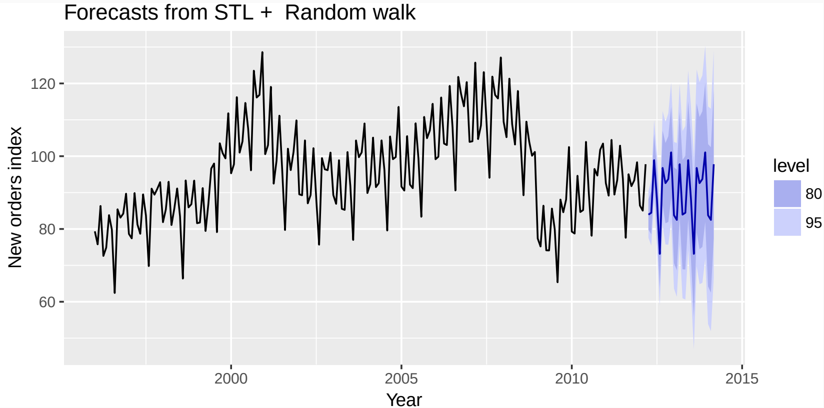
Electrical equipment

```
fcast <- forecast(fit, method="naive")
autoplot(fcast) + ylab("New orders index") + xlab("Yea
```



Forecasting and decomposition

```
elecequip %>% stlf(method='naive') %>% autoplot() +
  ylab("New orders index") + xlab("Year")
```



Decomposition and prediction intervals

- It is common to take the prediction intervals from the seasonally adjusted forecasts and modify them with the seasonal component.
- This ignores the uncertainty in the seasonal component estimate.
- It also ignores the uncertainty in the future seasonal pattern.