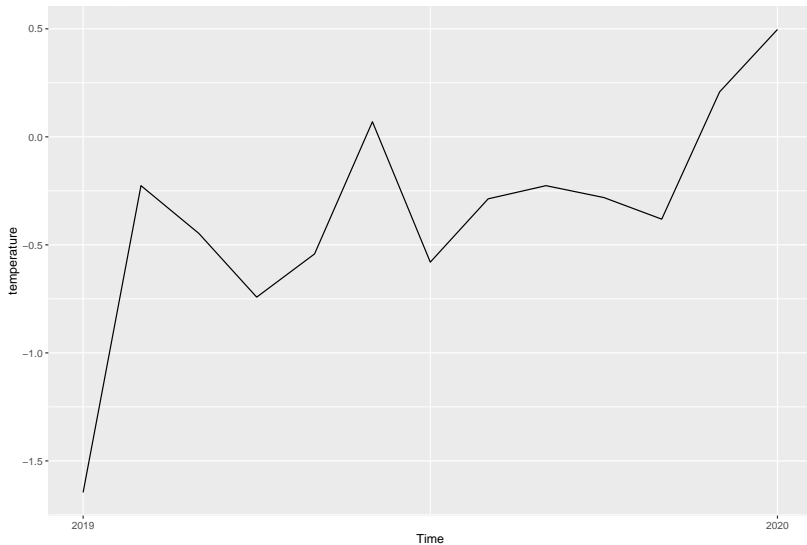


# Time Series

# 1. What is a time series

- ▶ A time series is a sequence of observation taken over time

World Temperature

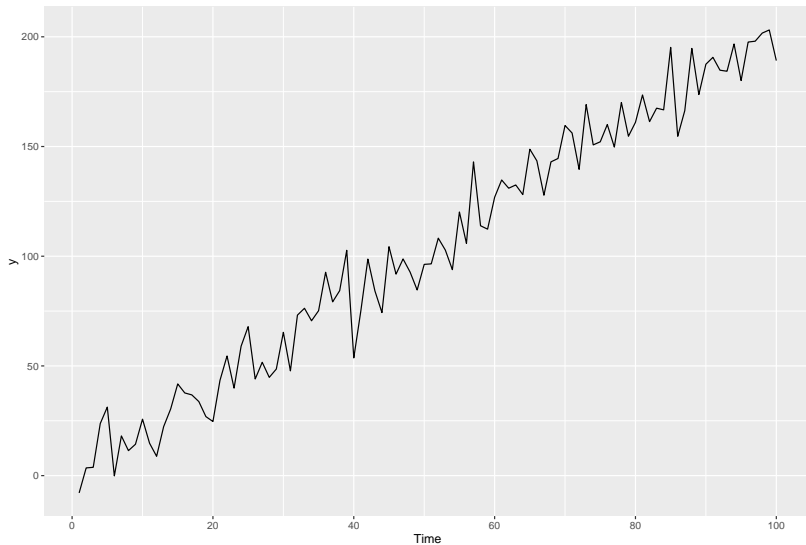


Stationary

## 2. Stationary

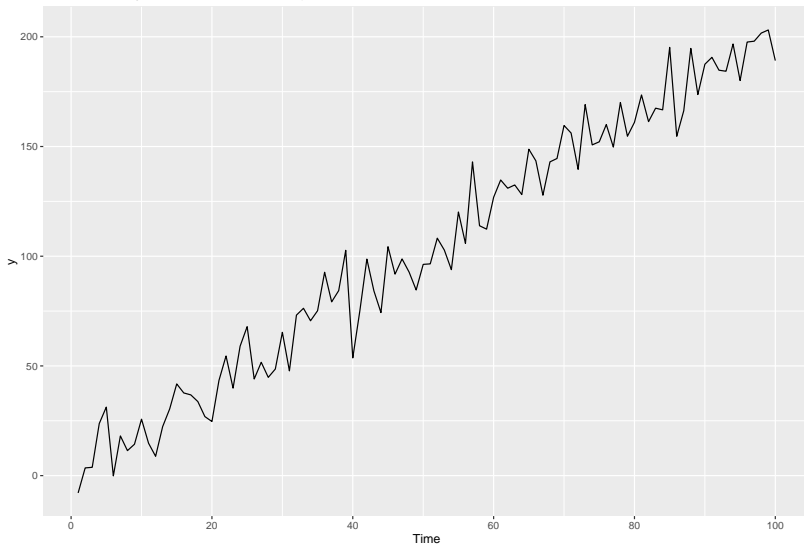
- ▶ A time series  $y_t$  is stationary if
  - ▶  $E(y_t) = \text{constant}$
  - ▶  $Cov(y_t, y_s)$  only depends on the time lag  $|t - s|$
- ▶ If  $y_t$  is stationary then  $Var(y_t) = \text{Constant}$

# Example

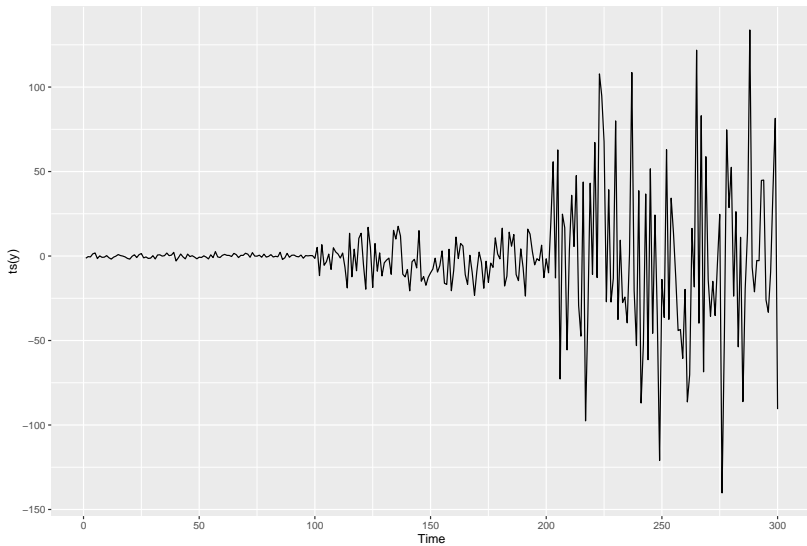


# Example

Non-stationary due to non-constant expected value



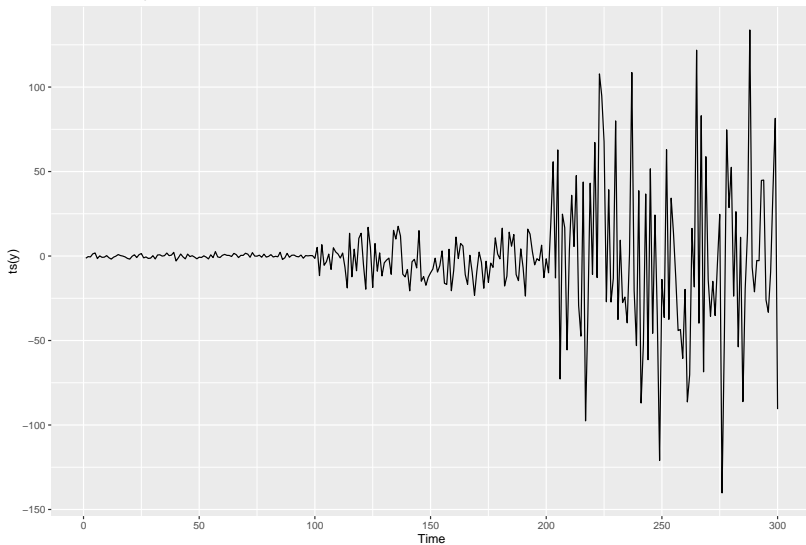
# Example

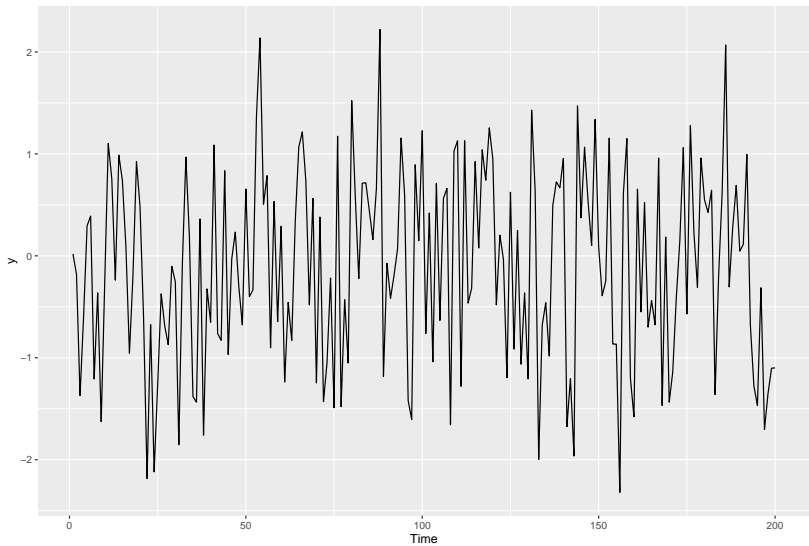




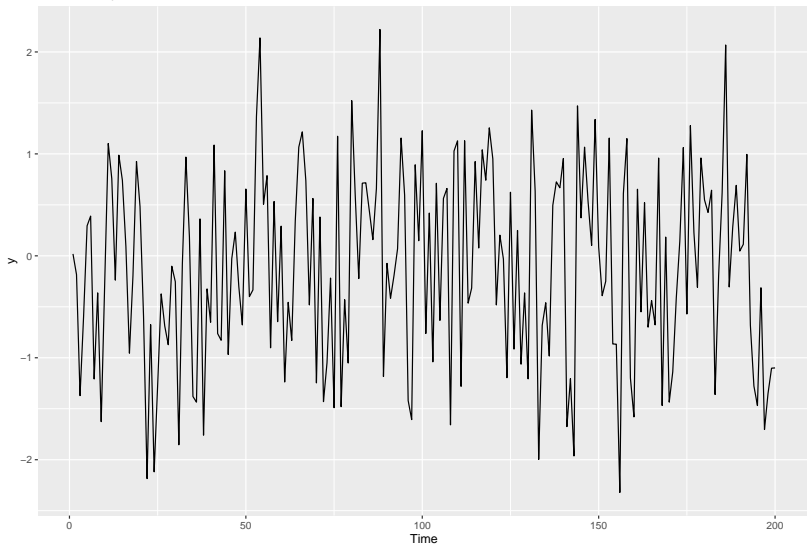
# Example

Non-stationary due to non-constant variance





A Stationary Time Series



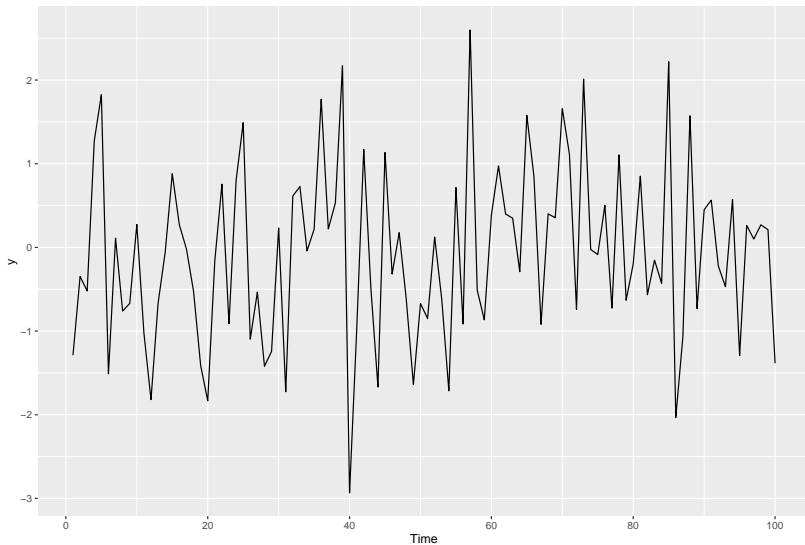
White Noise

### 3. White Noise

- ▶  $y_t$  is a white-noise process (series) if  $y_1, y_2, \dots, y_t, \dots$  are i.i.d random variables from a certain distribution (usually normal)
- ▶ A White noise is stationary

# Example

White noise of Standard Normal Distribution



# Random Walk

## 4. Random Walk

- ▶ A time series  $y_t$  is called a random walk if

$$y_t = y_{t-1} + c_t,$$

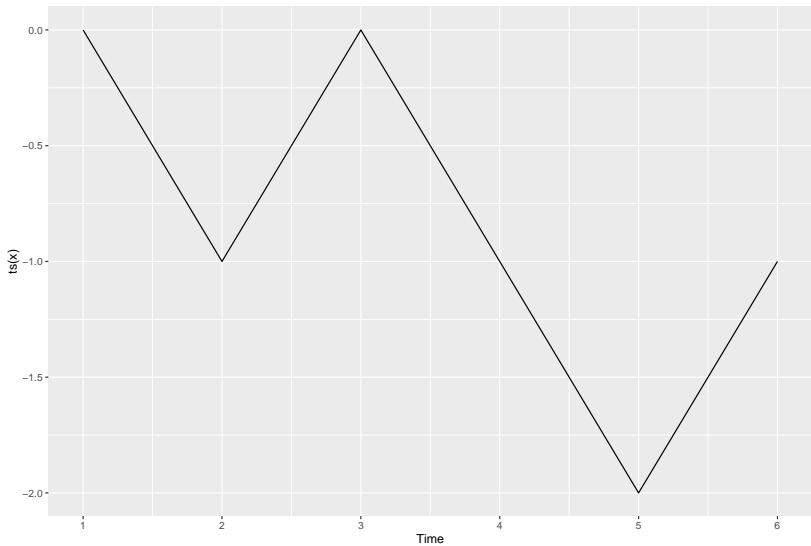
where  $c_t$  is a white-noise

- ▶ A random walk can be written as

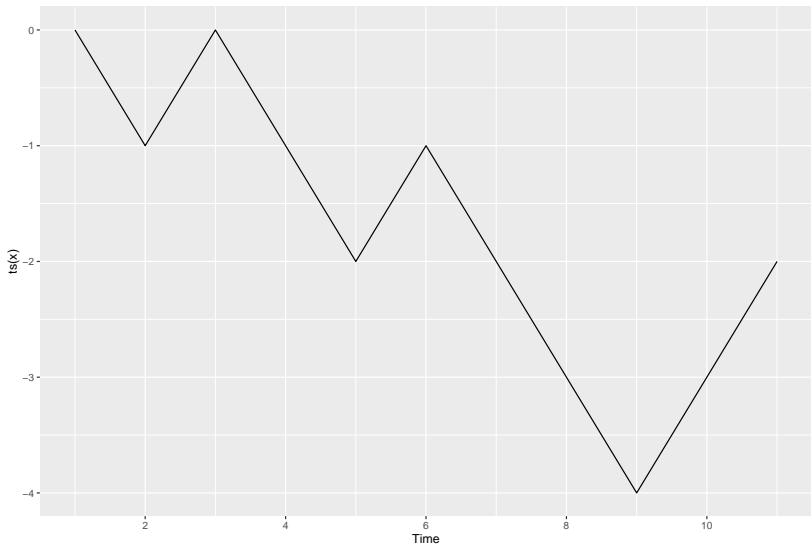
$$y_t = y_0 + c_1 + c_2 + \dots + c_t$$



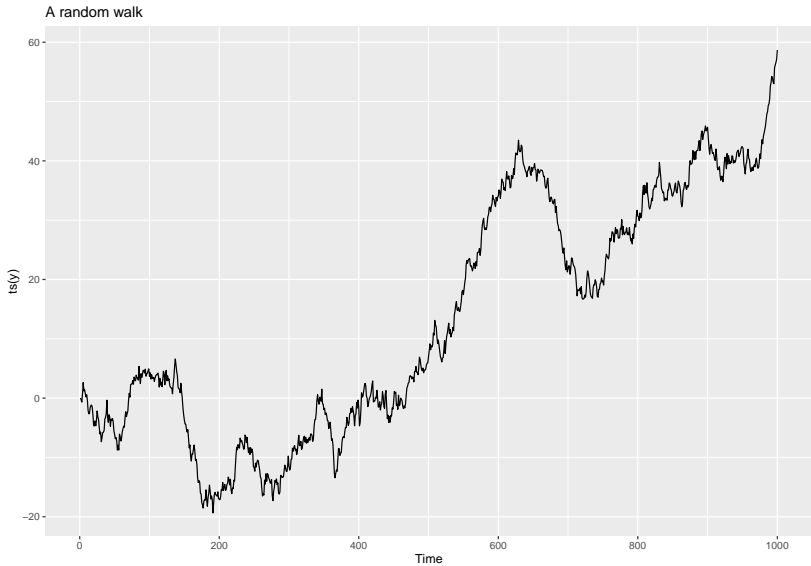
# Example



# Example



# Example



## Some Properties

- ▶ If  $c_t \sim (\mu_c, \sigma_c^2)$ , then

$$E(y_t) = E(y_0 + c_1 + c_2 + \dots + c_t) = y_0 + t\mu_c,$$

and

$$V(y_t) = t\sigma_c^2$$

- ▶ A random walk is non-stationary (unless the associated white-noise is non-random, i.e.  $\mu_c = \sigma_c^2 = 0$ )

$$\text{Cov}(y_t, y_s) = s\sigma_c^2$$

# Forecasting with Random Walks

## Forecasting with Random Walks

Suppose that we know  $y_0, y_1, \dots, y_T$  and we want to forecast  $y_{T+l}$  for some fixed  $l > 0$

- Point forecast: the estimated  $l$  step-ahead is

$$\hat{y}_{T+l} = y_T + l\hat{\mu}_c,$$

where  $\hat{\mu}_c$  is the estimated mean of the white-noise.  $\hat{\mu}_c$  can be  $\bar{c}$

$$\bar{c} = \frac{c_1 + c_2 + \dots + c_T}{T}$$

- The standard error of the forecast is  $s_c\sqrt{l}$ , where  $s_c$  is the estimated standard deviation of  $\sigma_c$ ,

$$s_c^2 = \frac{1}{n-1} \sum_{i=1}^T (c_i - \bar{c})^2$$

## Example

You are given:

i) The random walk model

$$y_t = y_0 + c_1 + c_2 + c_3 + \dots + c_t,$$

where  $c_i$ , ( $i = 1, 2, \dots, t$ ) denote observations from a white noise process.

ii) The following ten observed values of  $c_t$ :

t	1	2	3	4	5	6	7	8	9	10
$y_t$	2	5	10	13	18	20	24	25	27	30

iii)  $y_0 = 0$

Calculate the 9 step-ahead forecast,  $\hat{y}_{19}$ .





## Example

You are given:

- i) The random walk model

$$y_t = y_0 + c_1 + c_2 + c_3 + \dots + c_t,$$

where  $c_i$ , ( $i = 1, 2, \dots, t$ ) denote observations from a white noise process.

- ii) The following ten observed values of  $c_t$ :

t	1	2	3	4	5	6	7	8	9	10
$y_t$	2	5	10	13	18	20	24	25	27	30

- iii)  $y_0 = 0$

Calculate the standard error of the 9 step-ahead forecast,  $\hat{y}_{19}$ .

We have

$$c_t = y_t - y_{t-1} \implies c_1, c_2, \dots, c_{10} = 2, 3, 5, 3, 5, 2, 4, 1, 2, 3$$

$$\implies \bar{c} = \frac{c_1 + c_2 + \dots + c_{10}}{10} = 3$$

$$\implies s_c^2 = \frac{1}{9} \sum_{i=1}^{10} (c_i - 3)^2 = 16/9$$

Hence, the standard error is  $s_c \sqrt{l} = \frac{4}{3} \sqrt{9} = 4$

## Example

You are given the following eight observations from a time series that follows a random walk model:

$t$	0	1	2	3	4	5	6	7
$y_t$	3	5	7	8	12	15	21	22

You plan to fit this model to the first five observations and then evaluate it against the last three observations using one-step forecast residuals. The estimated mean of the white noise process is 2.25.

Calculate the mean error (ME) of the three predicted observations.

We have  $\hat{\mu}_c = 2.25$ . Notice that we are forced to use one-step ahead estimation to calculate  $\hat{y}_5, \hat{y}_6, \hat{y}_7$ . Thus, we need to use  $y_4$  to estimate  $\hat{y}_5$ ,  $y_5$  to estimate  $\hat{y}_6$ , and  $y_6$  to estimate  $\hat{y}_7$ . We have

$$\hat{y}_5 = y_4 + \hat{\mu}_c = 12 + 2.25 = 14.25$$

$$\hat{y}_6 = y_5 + \hat{\mu}_c = 15 + 2.25 = 17.25$$

$$\hat{y}_7 = y_6 + \hat{\mu}_c = 21 + 2.25 = 23.25$$

Hence, the ME error is

$$\begin{aligned} ME &= \frac{1}{3}(y_{15} - \hat{y}_{15} + y_{16} - \hat{y}_{16} + y_{17} - \hat{y}_{17}) \\ &= 15 - 14.25 + 21 - 17.25 + 22 - 23.25 \\ &= 1.083 \end{aligned}$$

Autoregressive model

## 5. Autoregressive model

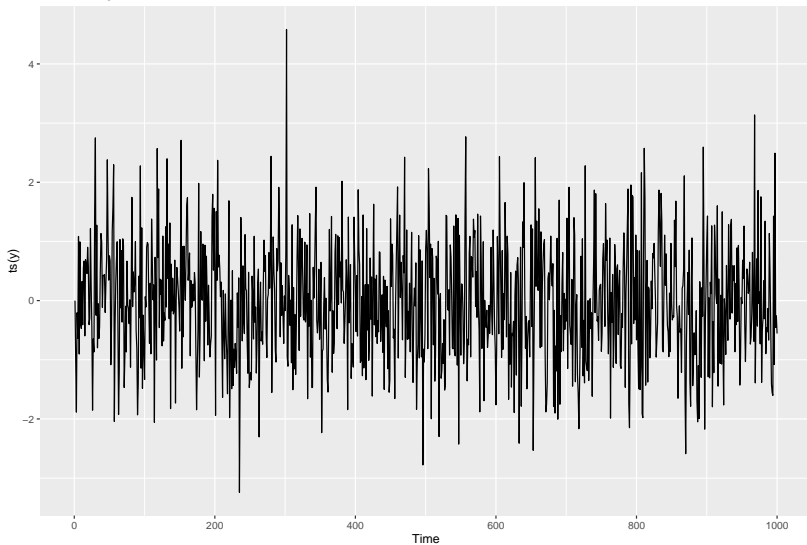
- ▶ A time series  $y_t$  is called a *first-order autoregressive model*, or AR(1) if

$$y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon_t,$$

where  $|\beta_1| \leq 1$ ,  $\epsilon_t$  is a zero mean white-noise process and  $\epsilon_{t+k}$  is independent of  $y_t$  for any  $t > 0$  and  $k > 0$ .

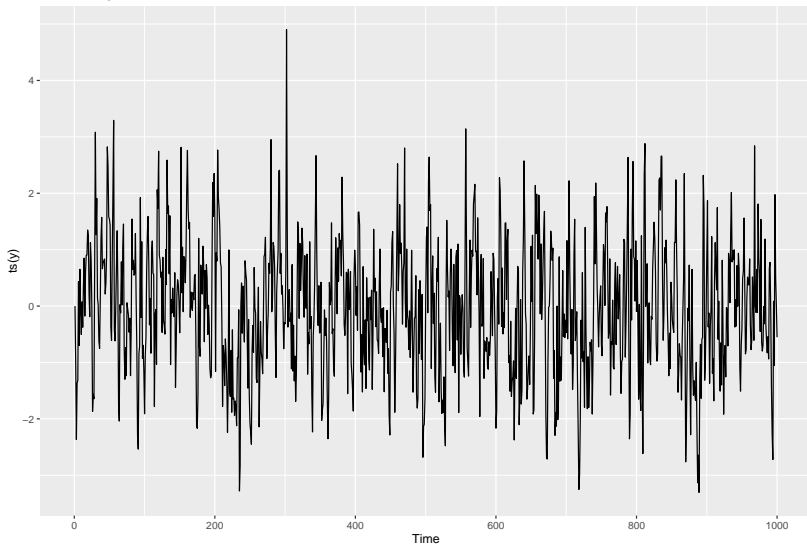
- ▶ When  $\beta_1 = 1$ , AR(1) becomes a random walk model.
- ▶ When  $\beta_1 = 0$ , AR(1) becomes a white noise.
- ▶ when  $|\beta_1| < 1$ , AR(1) is stationary and vice versa

An Autoregressive series with  $\beta_1 = 0.01$

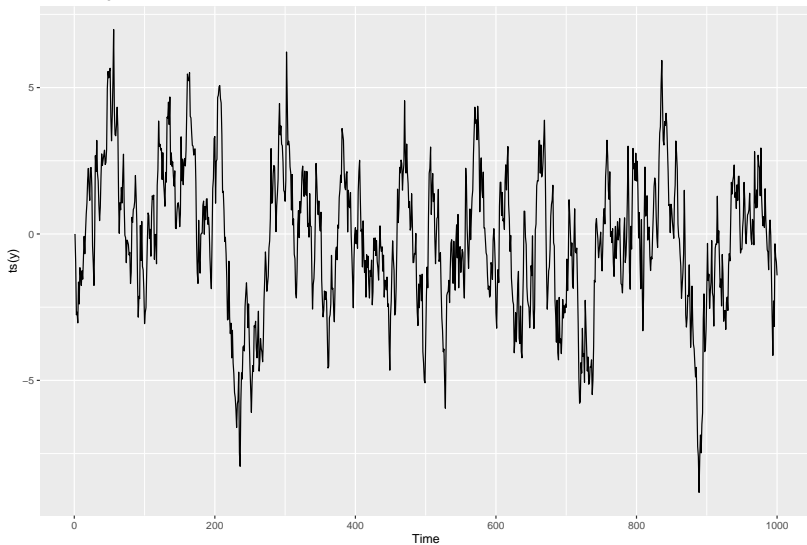




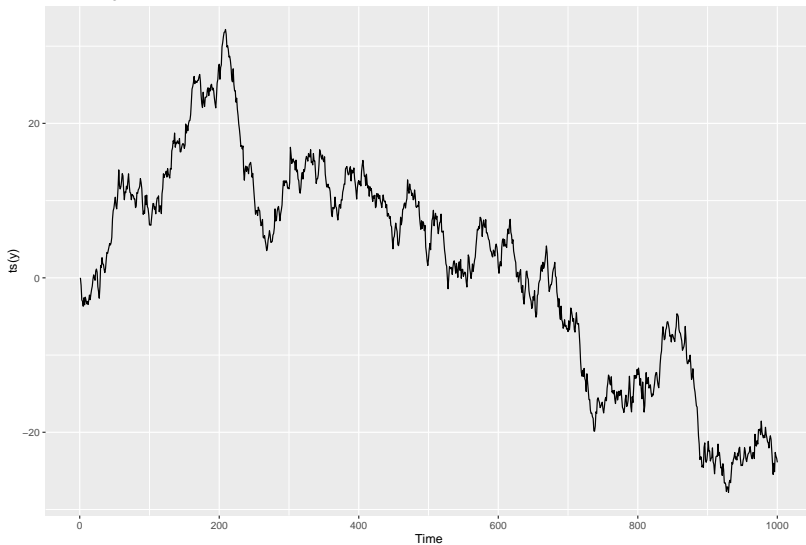
An Autoregressive series with  $\beta_1 = 0.5$



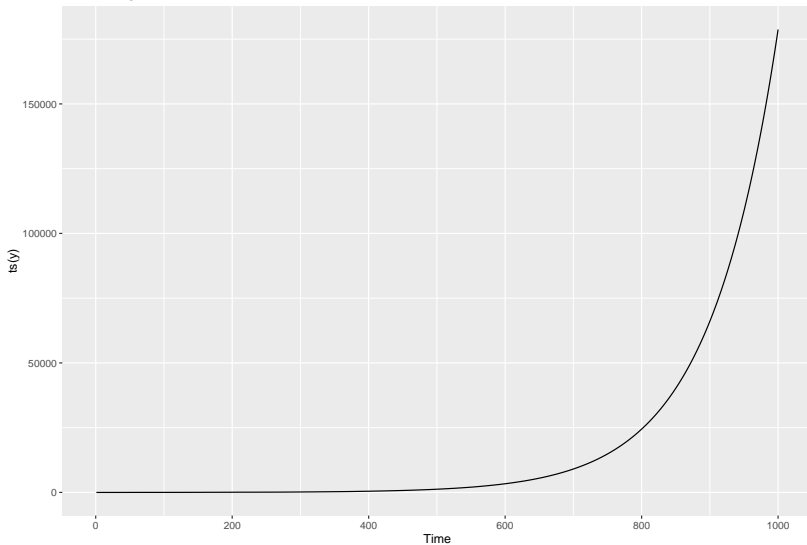
An Autoregressive series with  $\beta_1 = 0.9$



An Autoregressive series with  $\beta_1 = 1$



An Autoregressive series with  $\beta_1 = 1.01$



## Properties: Expectation

- Assume we have a stationary AR(1). Thus,  $E(y_t) = E(y_{t-1})$ .  
Therefore,

$$\begin{aligned} E(y_t) &= E\left(\beta_0 + \beta_1 y_{t-1} + \epsilon_t\right) \\ &= \beta_0 + \beta_1 E(y_{t-1}) \\ &= \beta_0 + \beta_1 E(y_t) \\ \implies E(y_t) &= \frac{\beta_0}{1 - \beta_1} \end{aligned}$$

## Properties: Variance

- Since we have a stationary AR(1),  $V(y_t) = V(y_{t-1})$ .  
Therefore,

$$\begin{aligned} V(y_t) &= V\left(\beta_0 + \beta_1 y_{t-1} + \epsilon_t\right) \\ &= \beta_1^2 V(y_{t-1}) + \sigma_\epsilon^2 \\ &= \beta_1^2 V(y_t) + \sigma_\epsilon^2 \\ \implies V(y_t) &= \frac{\sigma_\epsilon^2}{1 - \beta_1^2} \end{aligned}$$

# Parameter Estimation

- ▶ AR(1) is very similar to linear model where  $y_{t-1}$  play the roles of the predictor and  $y_t$  is the response
- ▶ In linear model, the predictor  $x$  is assumed to be non-random while the predictor  $y_{t-1}$  is non-random in AR(1)
- ▶ We estimate  $\beta_0$  and  $\beta_1$  by minimizing

$$\sum_{t=2}^T \left( y_t - E(y_t | y_{t-1}) \right)^2 = \sum_{t=2}^T \left( y_t - \beta_0 - \beta_1 y_{t-1} \right)^2$$

- ▶ These estimators are called the conditional least squares estimators

The coefficients are estimated by

$$\hat{\beta}_1 = \frac{\sum_{t=2}^T (y_{t-1} - \bar{y})(y_t - \bar{y})}{\sum_{t=2}^T (y_t - \bar{y})^2}$$
$$\hat{\beta}_0 = \bar{y}(1 - \hat{\beta}_1)$$

The only parameter left to estimate is the error variance,  $\sigma_\epsilon^2$ , (error mean is zero), which can be estimated by  $s^2$

$$s^2 = \frac{\sum_{t=2}^T (e_t - \bar{e})^2}{T - 3}$$

where  $e_t = y_t - (\hat{\beta}_0 - \hat{\beta}_1 y_{t-1})$ .



## Example

You are given the following six observed values of the autoregressive model of order one time series

$$y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon_t, \text{ with } Var(\epsilon_t) = \sigma^2.$$

$t$	1	2	3	4	5
$y_t$	1	3	5	8	13

Calculate  $\hat{\beta}_1$  using the conditional least squares method.



# Forecasting with $AR(1)$

- ▶ Suppose we have the AR(1) time series with known  $\beta_0$  and  $\beta_1$ . If these parameters are unknown we can estimate them by the formula in the previous slices.
- ▶ We use the following formulas to for forecasting

$$\hat{y}_{T+1} = \beta_0 + \beta_1 y_T$$

$$\hat{y}_{T+k} = \mu + \beta_1^k (y_T - \mu)$$

where  $\mu = \frac{\beta_0}{1-\beta_1}$ .

## Example

You are given

$$y_t = .3y_{t-1} + 4 + \epsilon$$

$$y_T = 7$$

Calculate the three step ahead forecast of  $y_{T+3}$



# Smoothing

## 6. Smoothing

- ▶ Smoothing is usually done to reveal the series patterns and trends.



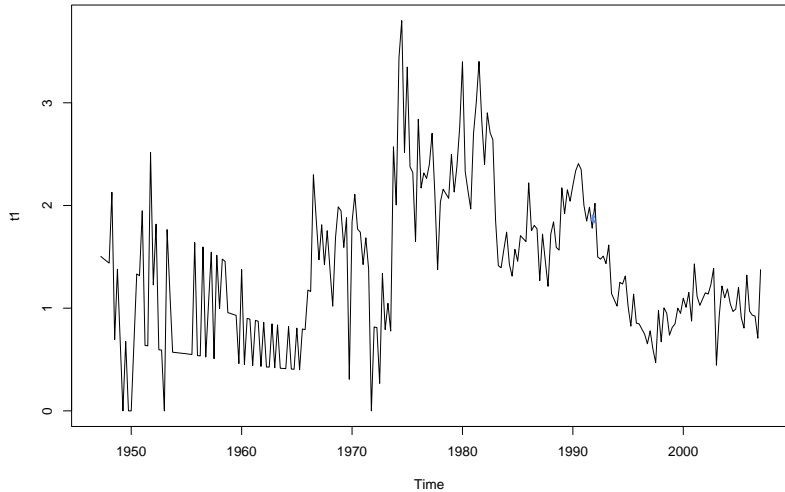
# Simple Moving Average Smoothing

- ▶ Moving Average (MA) creates a new series by averaging the most recent observations from the original series.
- ▶ MA(k) creates  $s_t$  as follows

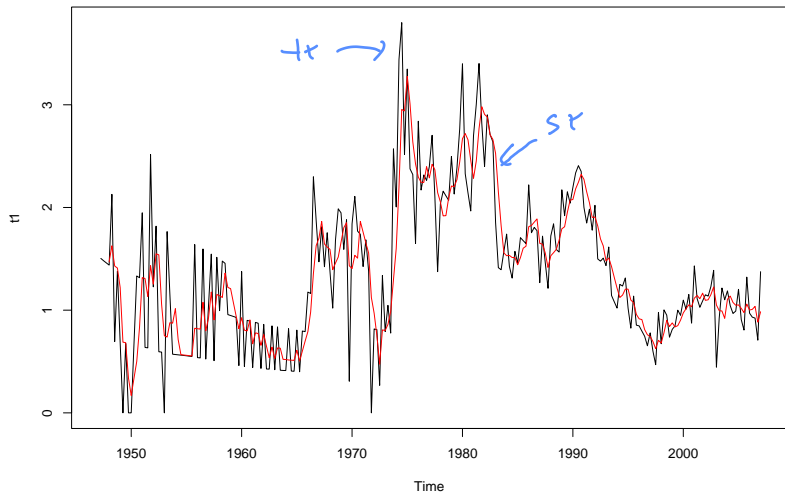
*smoothing* ↘  $s_t = \frac{y_t + y_{t-1} + \dots + y_{t-k+1}}{k}$       ↙ *original*

- ▶ Larger  $k$  will smooth the series more strongly

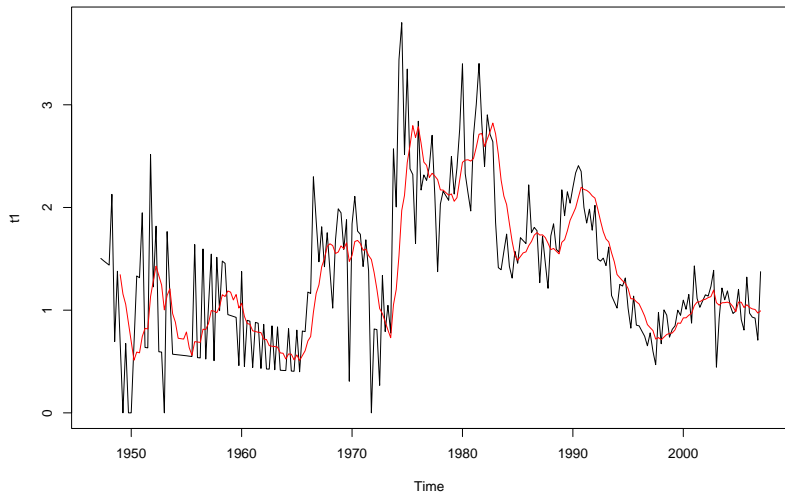
### Medical Component of the CPI



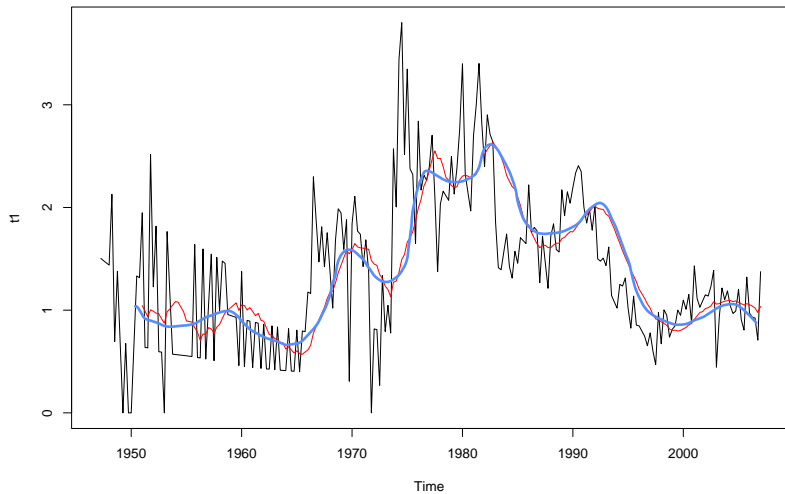
Moving average with  $k = 4$



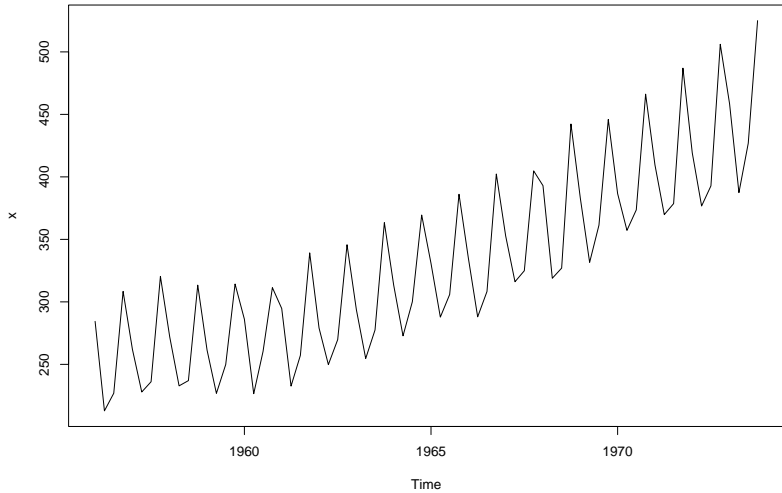
**Moving average with  $k = 8$**



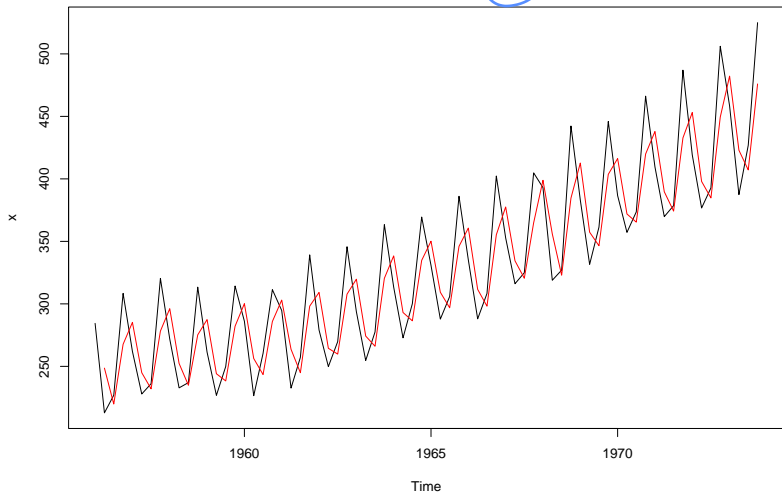
Moving average with  $k = 16$



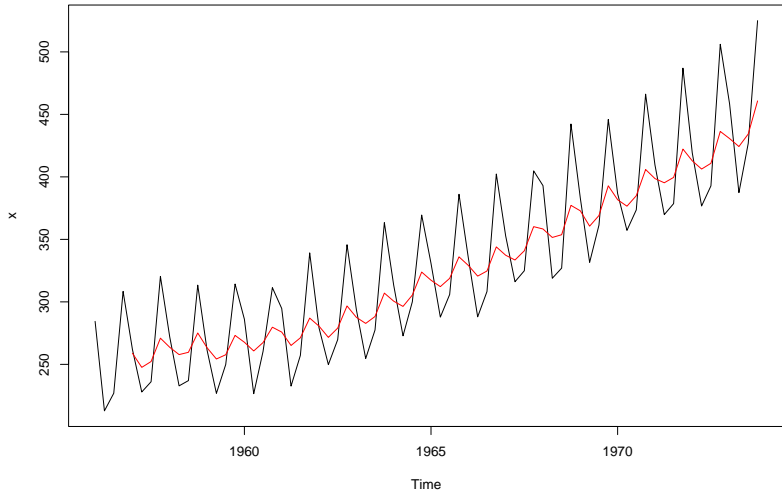
**Original Series**



Moving average with  $k = 2$

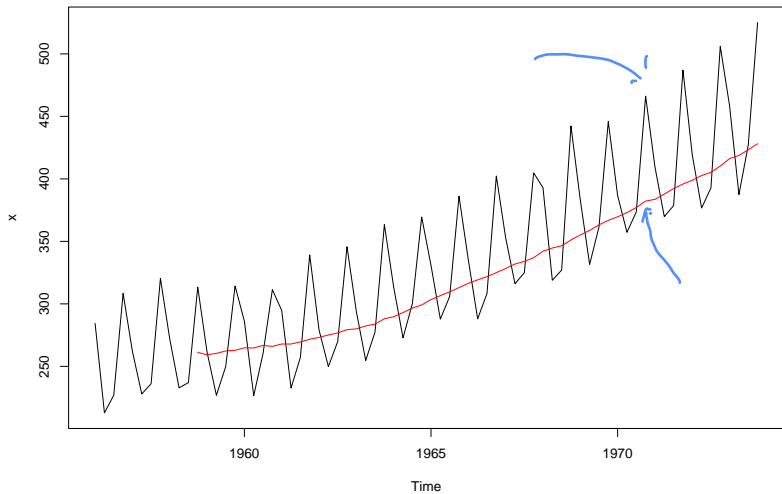


**Moving average with  $k = 5$**





Moving average with  $k = 12$



# Forecasting

- ▶ We can use smoothing for forecasting
- ▶ We have

$$\begin{aligned}\hat{s}_t &= \frac{y_t + y_{t-1} + \dots + y_{t-k+1}}{k} \\&= \frac{y_t + y_{t-1} + \dots + y_{t-k+1} + y_{t-k} - y_{t-k}}{k} \\&= \frac{y_t + \left( y_{t-1} + \dots + y_{t-k+1} + y_{t-k} \right) - y_{t-k}}{k} \\&= \frac{y_t + k\hat{s}_{t-1} - y_{t-k}}{k} \\&= \hat{s}_{t-1} + \frac{y_t - y_{t-k}}{k}\end{aligned}$$

# Forecasting

- ▶ If there is no trend in  $y_t$  the second term  $(y_t - y_{t-k})/k$  can be ignored
- ▶ Forecasting  $l$  lead time into future by

$$\hat{y}_{T+l} = \hat{s}_T$$

- ▶ If there is a linear trend in a series, we can use the double moving average to estimate the trend (slope)

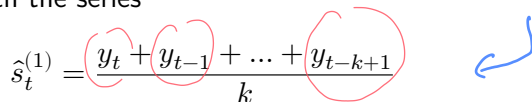
Double MA

## 7. Double MA


- ▶ Linear trend time series:

$$y_t = \beta_0 + \beta_1 t + \epsilon_t$$

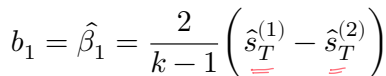
- ▶ Step 1: Smooth the series

$$\hat{s}_t^{(1)} = \frac{y_t + y_{t-1} + \dots + y_{t-k+1}}{k}$$


- ▶ Step 2: Smooth the smoothed series

$$\hat{s}_t^{(2)} = \frac{\hat{s}_t^{(1)} + \hat{s}_{t-1}^{(1)} + \dots + \hat{s}_{t-k+1}^{(1)}}{k}$$


- ▶ Step 3: Calculate the trend

$$b_1 = \hat{\beta}_1 = \frac{2}{k-1} \left( \hat{s}_T^{(1)} - \hat{s}_T^{(2)} \right)$$


# Forecasting

- Forecasting  $l$  lead time into future by

$$\hat{y}_{T+l} = \hat{s}_T + b_1 \cdot l$$

You are given the following time series

$t$	1	2	3	4	5
$y_t$	1	3	5	8	13

- ① Forecasting  $y_7$  using simple moving average with  $k = 2$
- ② Forecasting  $y_7$  using double moving average with  $k = 2$

$$s_t^{(1)} = \left\{ \begin{array}{cccc} 2 & 4 & 6.5 & 10.5 \\ \downarrow & \downarrow & \downarrow & \downarrow \\ \hat{s}_2^{(1)} & \hat{s}_3^{(1)} & \hat{s}_4^{(1)} & \hat{s}_5^{(1)} \end{array} \right\}$$

$$\hat{y}_7 = \hat{s}_5^{(1)} = \boxed{10.5}$$

$$x_i = \{1 \quad 3 \quad 5 \quad 8 \quad 13\}$$

$$\hat{S}_T^{(0)} = \{ \underbrace{2, 4, 6.5, 10.5} \} \leftarrow$$

$$\hat{S}_T^{(2)} = \{ 3, 5.25, 8.5 \} \leftarrow$$

$$k=2$$

The slope  $b_1 = \frac{2}{k-1} \left[ \hat{S}_T^{(0)} - \hat{S}_T^{(2)} \right] = \frac{2}{2-1} [10.5 - 8.5]$

$$= 4$$

$$\hat{y}_7 = \hat{S}_T^{(0)} + b_1 \cdot l$$

$$l = 7 - 5 = 2$$

$$= 10.5 + 4 \cdot 2 = \boxed{18.5}$$

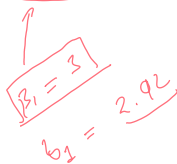


## Example

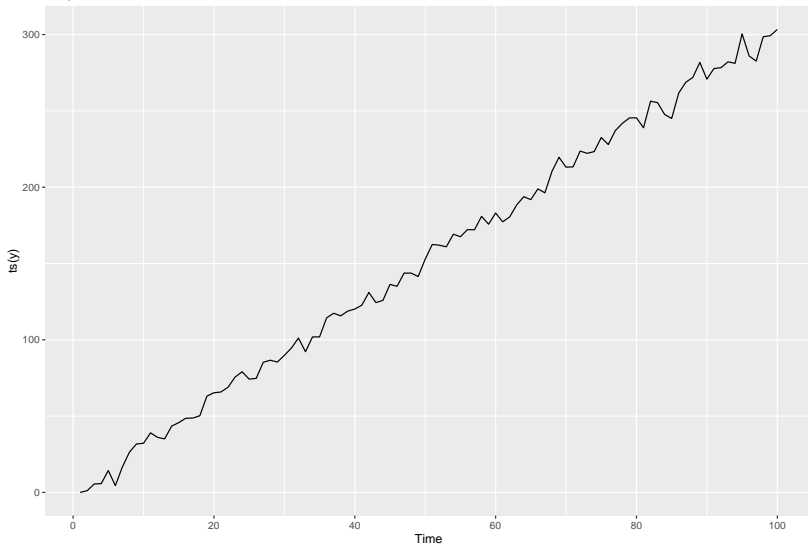
- We simulate 100 data points ( $T = 100$ ) of

$$y_t = 1 + 3t + \epsilon,$$

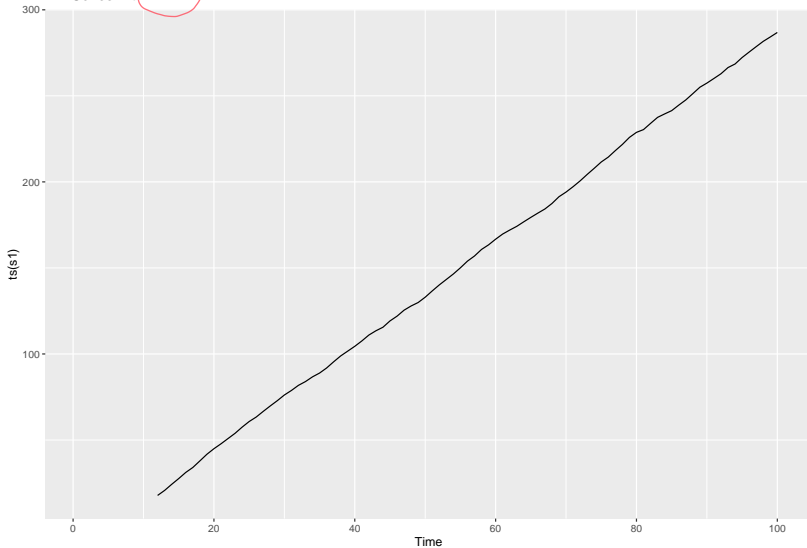
where,  $\epsilon \sim N(0, 5^2)$ .


$$\beta_1 = 3$$
$$b_1 = 2.92$$

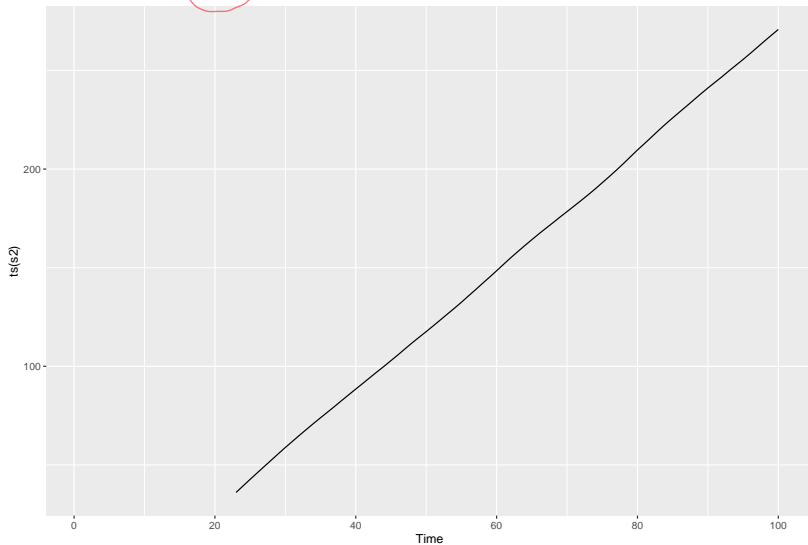
Original Series



MA Series with  $k = 12$



Double MA Series with  $k = 12$



- ▶ Using the above steps, the estimated trend is  $b_1 = 2.92$
- ▶ The forecast for the next points from  $y_{100}$  is

$$\hat{y}_{100+l} = \hat{s}_{100} + b_1 \cdot l = \hat{s}_{100} + 2.92 \cdot l$$

# Exponential Smoothing

# Exponential Smoothing

- ▶ MA distributes the weight equally to the recent observations
- ▶ Exponential Smoothing controls the weights of the recent observations by  $w$

$$\hat{s}_t = \frac{\underbrace{y_t}_{\downarrow} + \underbrace{w}_{\downarrow}y_{t-1} + \underbrace{w^2}_{\downarrow}y_{t-2} + \dots + \underbrace{w^t}_{\downarrow}y_0}{1/(1-w)}$$

- ▶ Smaller  $w$  ( $w \rightarrow 0$ ) gives higher weights to the more recent observations
- ▶ Smaller  $w$  smooths the series more lightly.

# Exponential Smoothing

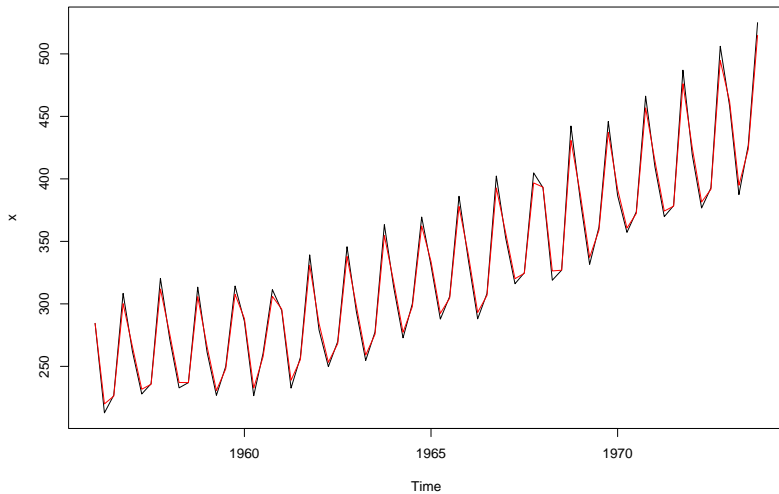
► We have

$$\begin{aligned}\hat{s}_t &= \hat{s}_{t-1} + (1 - w)(y_t - \hat{s}_{t-1}) \\ &= (1 - w)y_t + w\hat{s}_{t-1}\end{aligned}$$

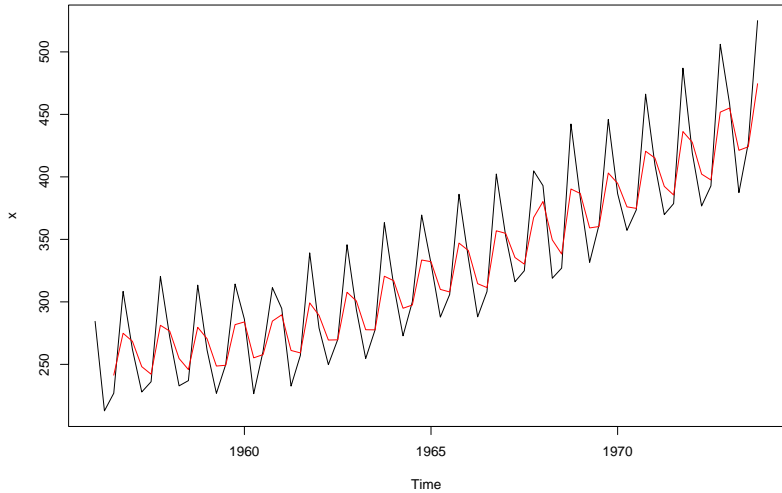
► When  $w \rightarrow 0$ ,  $\hat{s}_t \rightarrow y_t$ , or little smoothing has taken



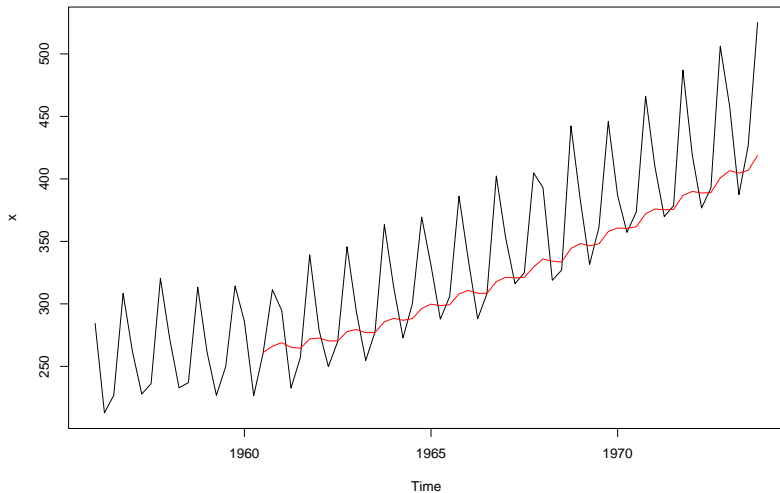
# Exponential Smoothing with $w = 0.1$



Exponential Smoothing with  $w = 0.5$



# Exponential Smoothing with $w = 0.9$



# Double Exponential Smoothing

# Double Exponential Smoothing

We can use double smoothing to identify the trend and forecast linear trend time series as follows.

► Step 1: Create a smoothed series:  $\hat{s}_t^{(1)} = (1 - w)y_t + w\hat{s}_{t-1}^{(1)}$

► Step 2: Create a double smoothed series:

$$\hat{s}_t^{(2)} = (1 - w)\hat{s}_t^{(1)} + w\hat{s}_{t-1}^{(2)}$$

► Step 3: Estimate the trend:

$$b_1 = \frac{1 - w}{w}(\hat{s}_T^{(1)} - \hat{s}_T^{(2)})$$

► Step 4: Forecast

$$\hat{y}_{T+l} = 2\hat{s}_T^{(1)} - \hat{s}_T^{(2)} + b_1 \cdot l$$

## Example

You are given the following time series

$t$	1	2	3	4	5
$y_t$	1	3	5	8	13

- ▶ Forecasting  $y_7$  using exponential smoothing with  $w = .8$
- ▶ Forecasting  $y_7$  using double exponential smoothing with  $w = .8$

