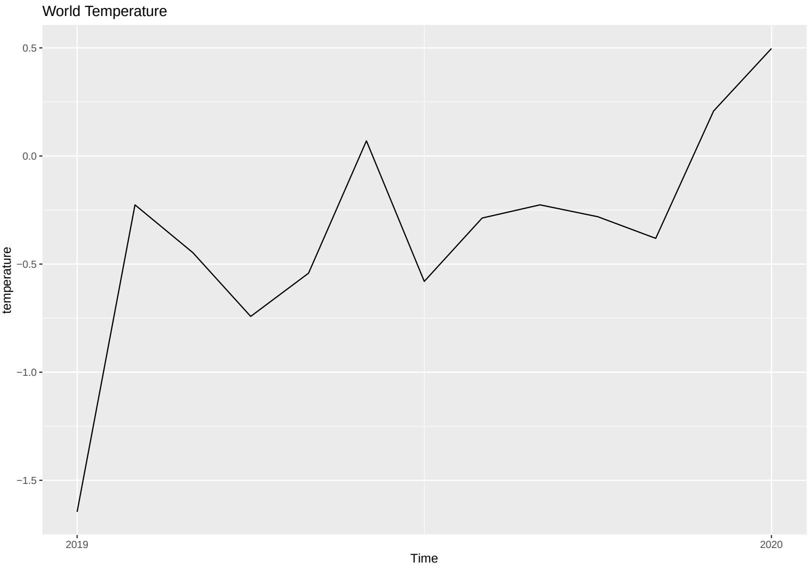
Time Series

1. What is a time series

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

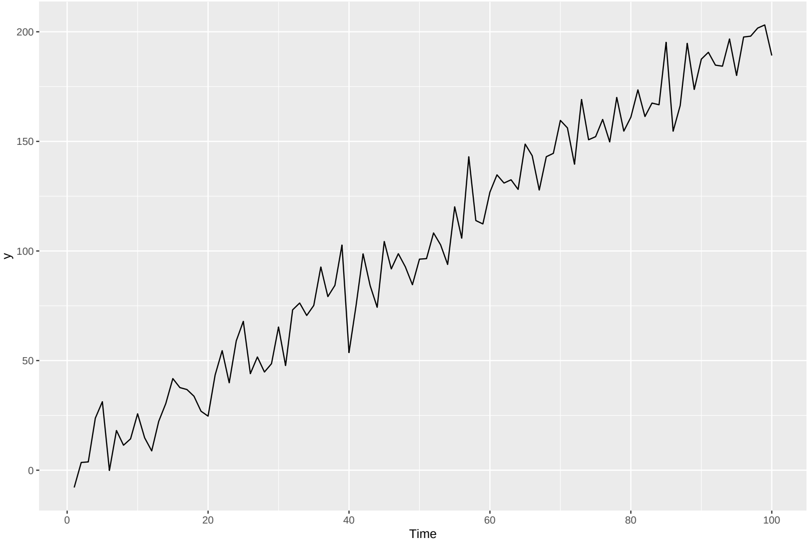

World Temperature
temperature
0.5
0.0
−0.5
−1.0
−1.5
2019
2020
Time

Stationary

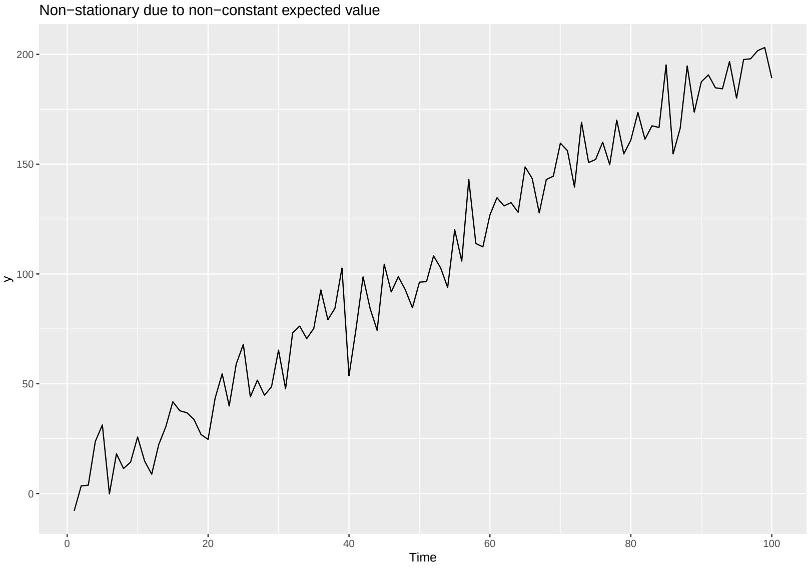
2. Stationary

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

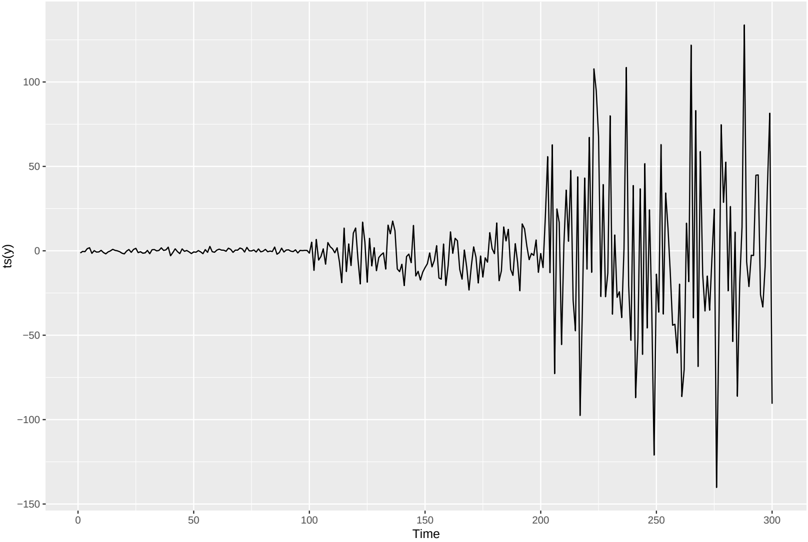
Example



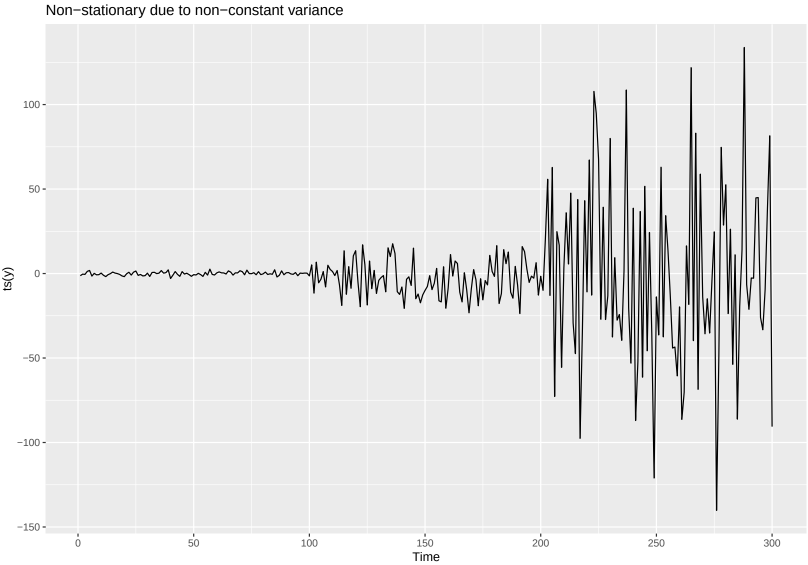
Example


Non−stationary due to non−constant expected value
200
150
100
50
0
y
0
20
40
60
80
100
Time

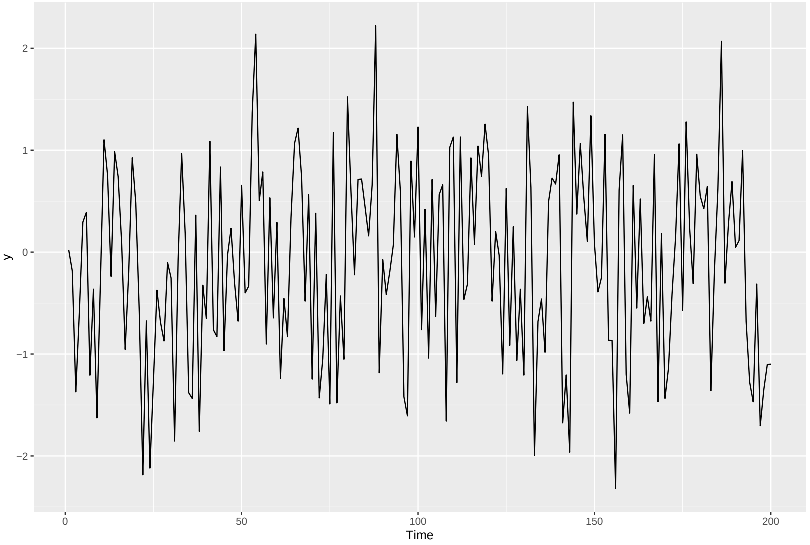
Example



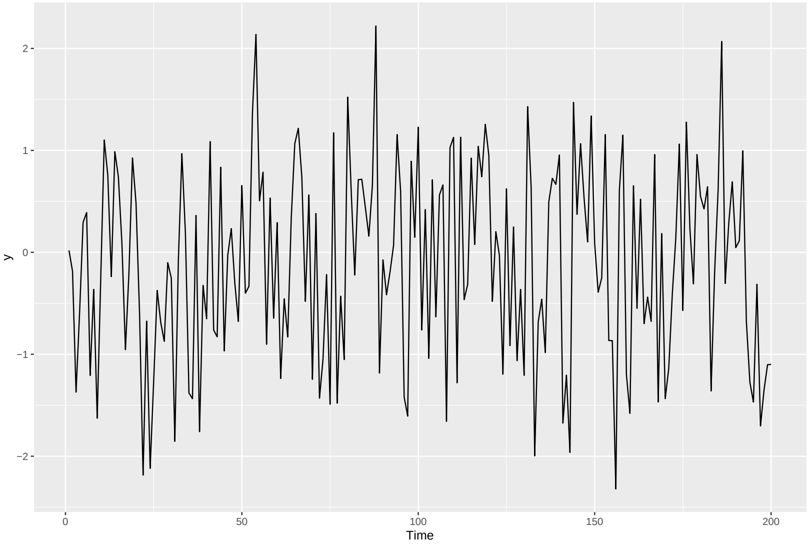
Example



Non−stationary due to non−constant variance
ts(y)
Time
100
50
0
−50
−100
−150
0
50
100
150
200
250
300


y
2
1
0
−1
−2
0
50
100
150
200
Time

A Stationary Time Series

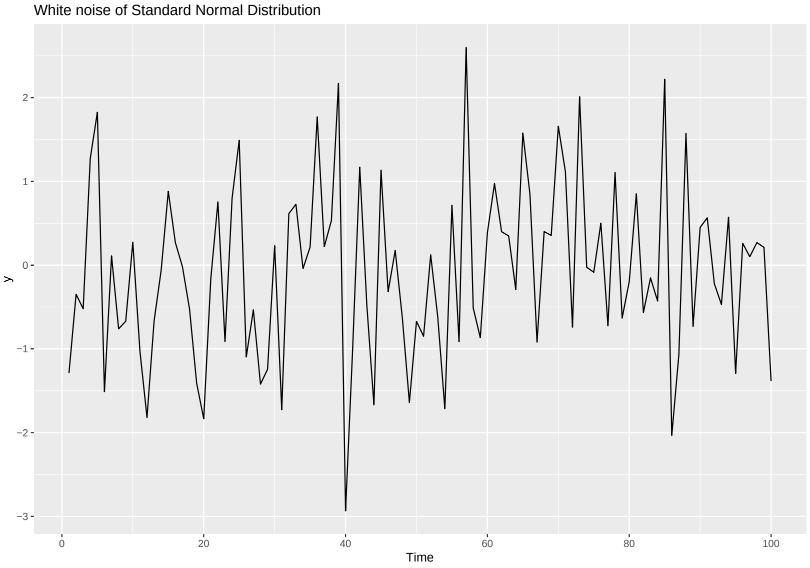


White Noise

3. White Noise

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

Example


White noise of Standard Normal Distribution
y
Time
2
1
0
−1
−2
−3
0
20
40
60
80
100

Random Walk

4. Random Walk

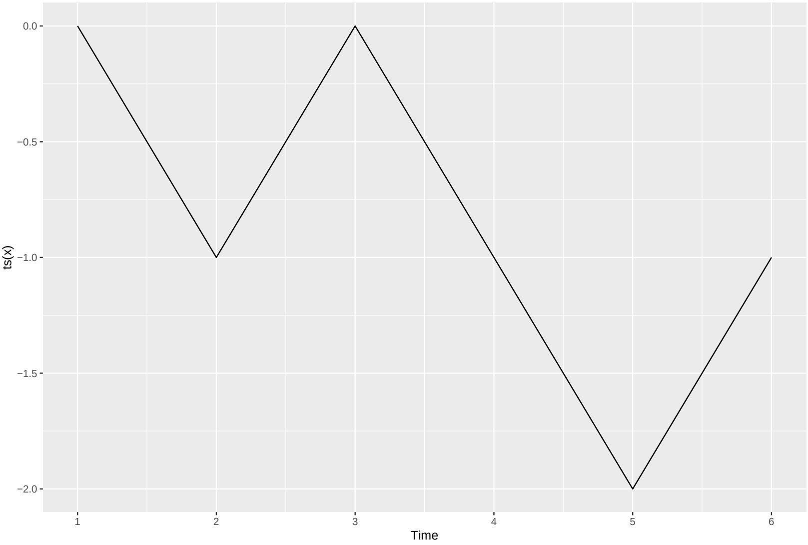
- A time series y_t is called a random walk if

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

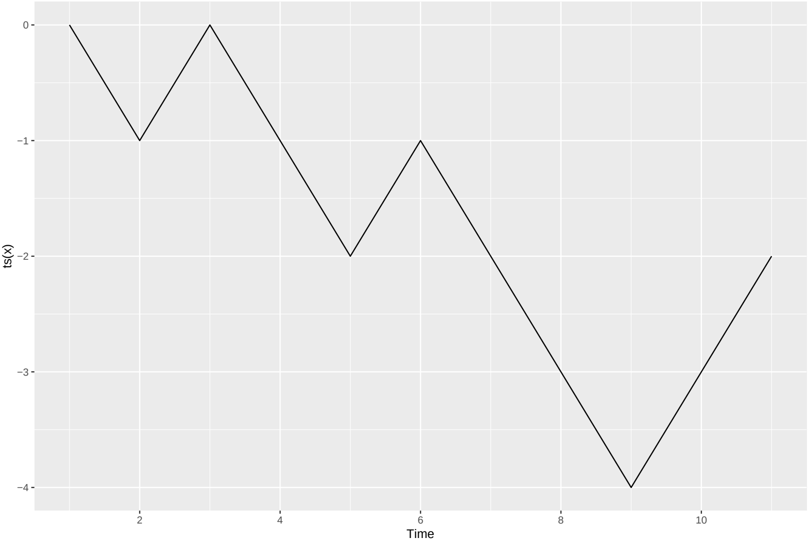
  where c_t is a white-noise
- A random walk can be written as

$$y_t = y_0 + c_1 + c_2 + ... + c_t$$

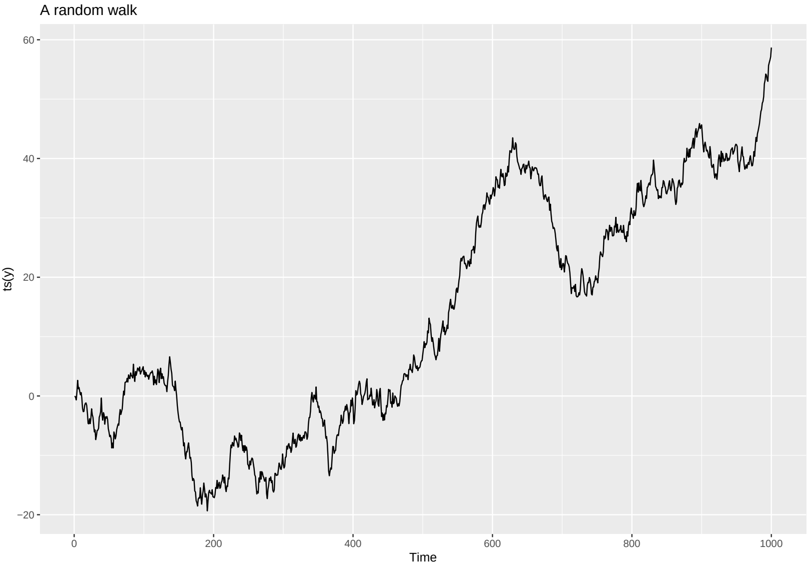
Example



Example



Example



Some Properties

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

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

and

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

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

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

Forecasting with Random Walks

Forecasting with Random Walks

Suppose that we know $y_0, y_1, ..., y_T$ 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 $\bar{c}$

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

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

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

Example

You are given:

i) The random walk model

$$y_t = y_0 + c_1 + c_2 + c_3 + ... + c_t,$$

where $c_i, (i = 1, 2, ..., t)$ 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

$$y_t = y_0 + c_1 + c_2 + c_3 + ... + c_t,$$

where $c_i, (i = 1, 2, ..., t)$ 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

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

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

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

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

$$\begin{aligned}
\hat{y}_5 &= y_4 + \hat{\mu}_c = 12 + 2.25 = 14.25 \\
\hat{y}_6 &= y_5 + \hat{\mu}_c = 15 + 2.25 = 17.25 \\
\hat{y}_7 &= y_6 + \hat{\mu}_c = 21 + 2.2.5 = 23.25
\end{aligned}$$

Hence, the ME error is

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

Autoregressive model

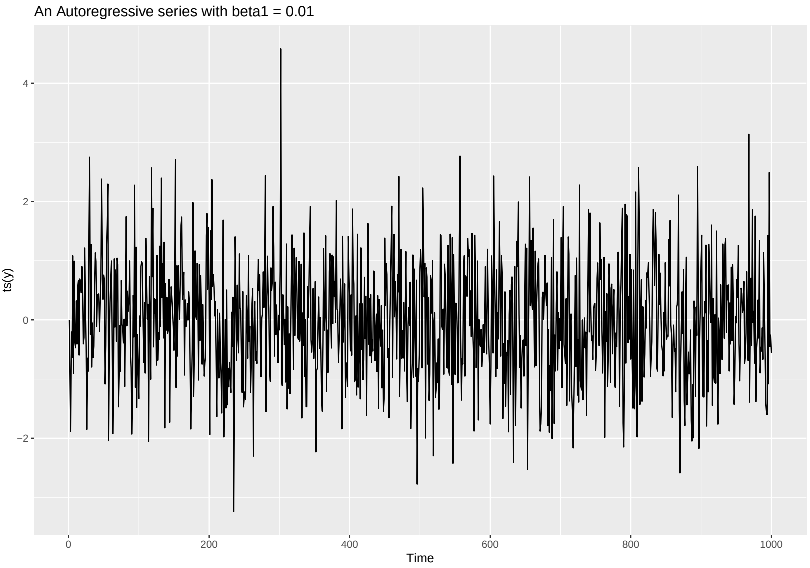
5. Autoregressive model

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

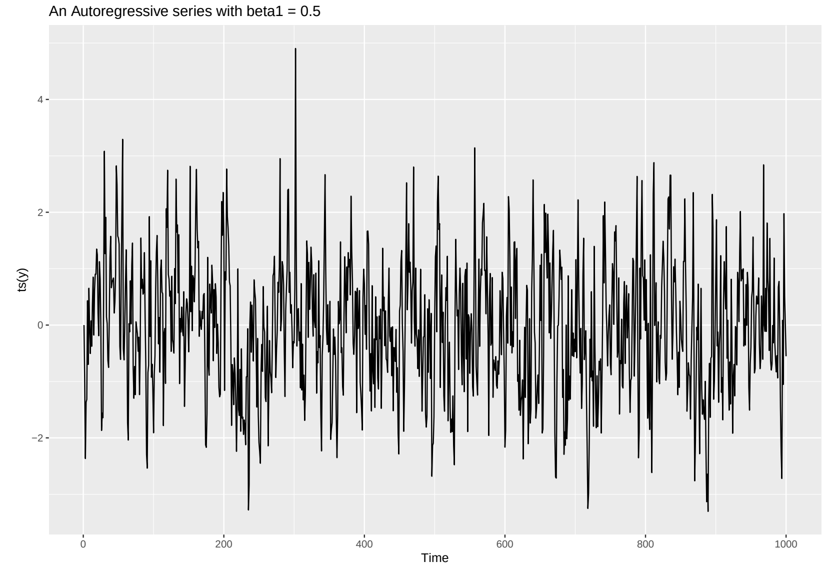
$$y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon_t,$$

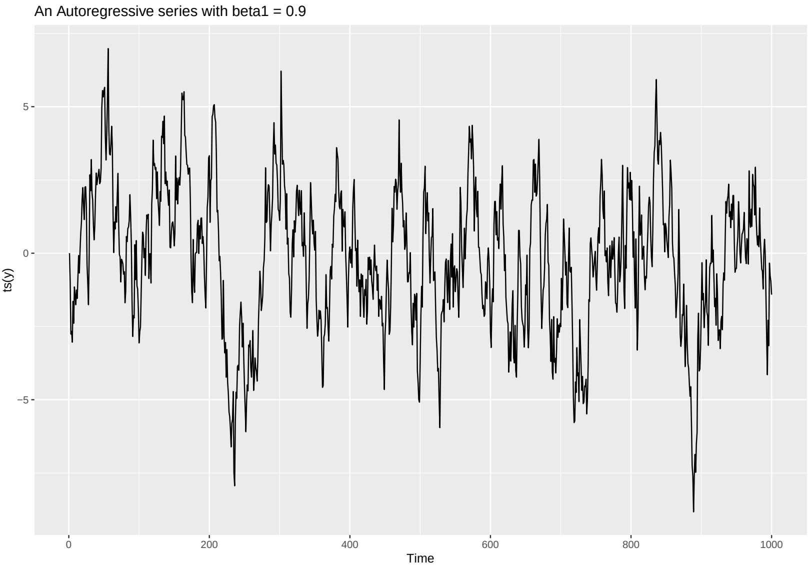
where $|\beta_1| \leq 1$, ϵ_t is a zero mean white-noise process and ϵ_{t+k} is independent of y_t for any $t > 0$ and $k > 0$.

- 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

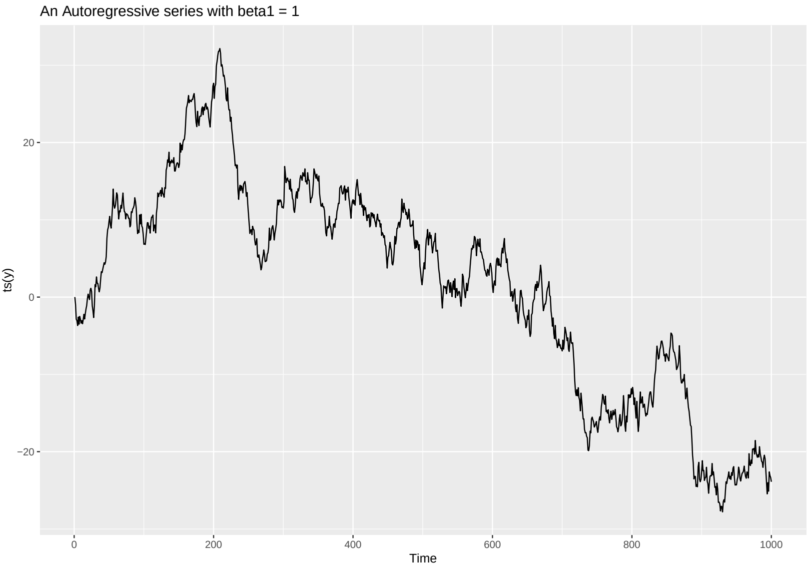


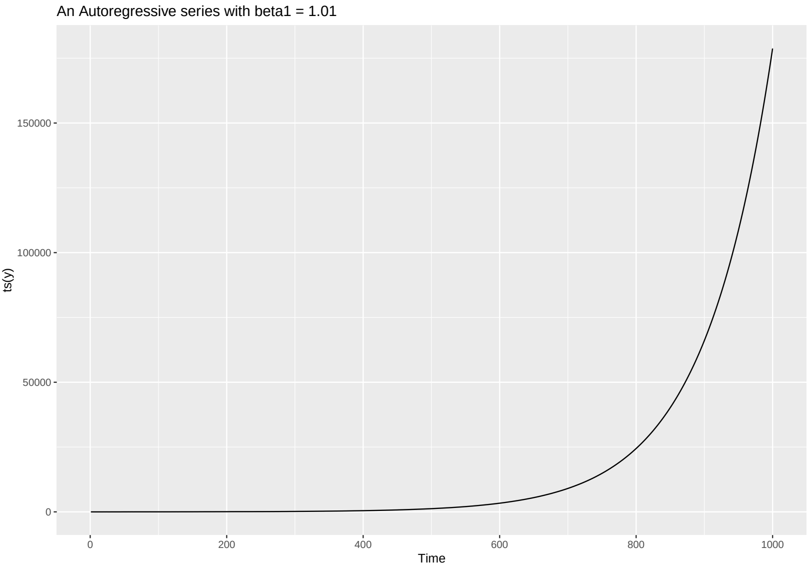
An Autoregressive series with beta1 = 0.01
ts(y)
4
2
0
−2
0
200
400
600
800
1000
Time


An Autoregressive series with beta1 = 0.5
ts(y)
4
2
0
−2
0
200
400
600
800
1000
Time



An Autoregressive series with beta1 = 0.9
5
0
−5
ts(y)
0
200
400
600
800
1000
Time


An Autoregressive series with beta1 = 1
ts(y)
20
0
−20
0
200
400
600
800
1000
Time



An Autoregressive series with beta1 = 1.01
ts(y)
150000
100000
50000
0
0
200
400
600
800
1000
Time

Properties: Expectation

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

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

Properties: Variance

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

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

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 β_0 and β_1 by minimizing

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

- 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, σ_ϵ^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 β_0 and β_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

$$\begin{aligned} y_t &= .3y_{t-1} + 4 + \epsilon \\ y_T &= 7 \end{aligned}$$

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

Smoothing

6. Smoothing

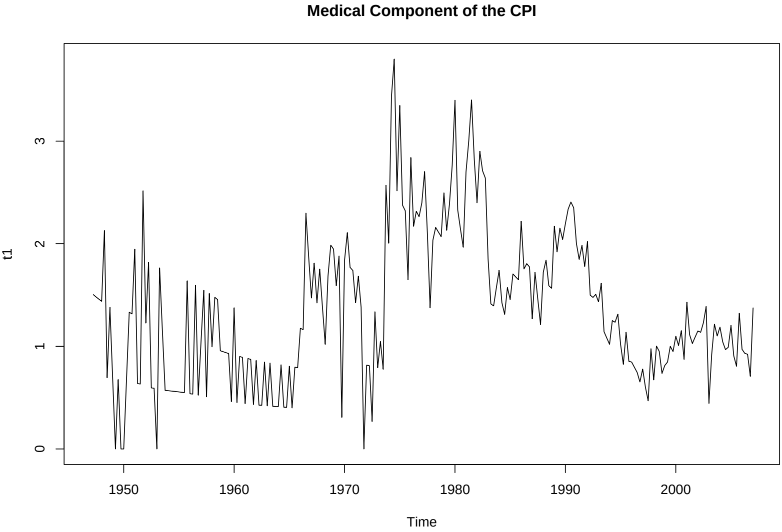
- Smoothing is usually done to reveal the series patterns and trends.

Simple Moving Average Smoothing

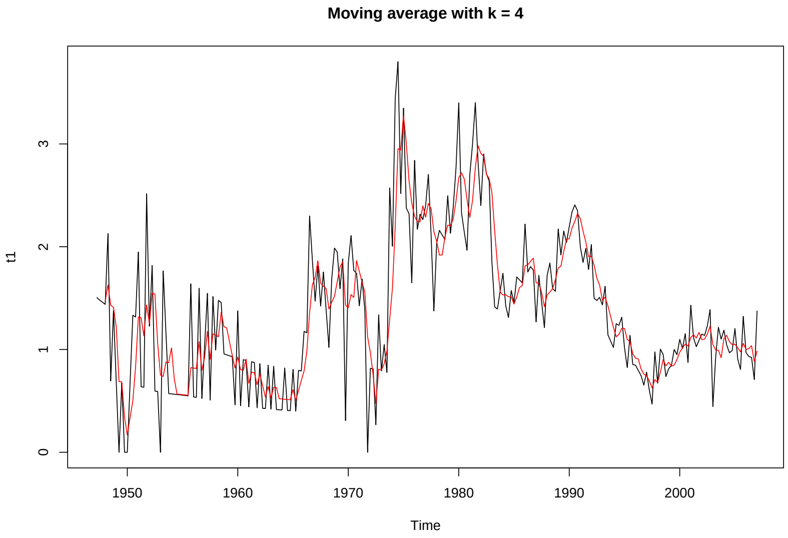
- Moving Average (MA) creates a new series by averaging the most recent observations from the original series.
- MA(k) creates s_t as follows.

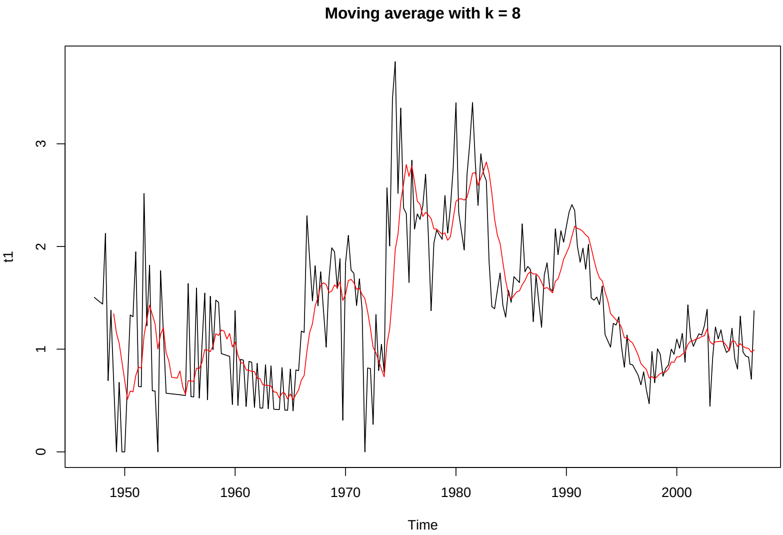
$$s_t = \frac{y_t + y_{t-1} + ... + y_{t-k+1}}{k}$$

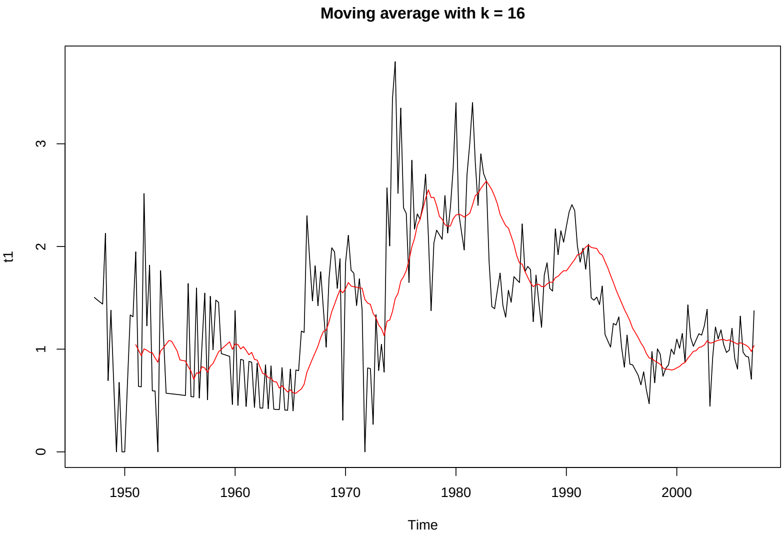
- Larger k will smooth the series more strongly



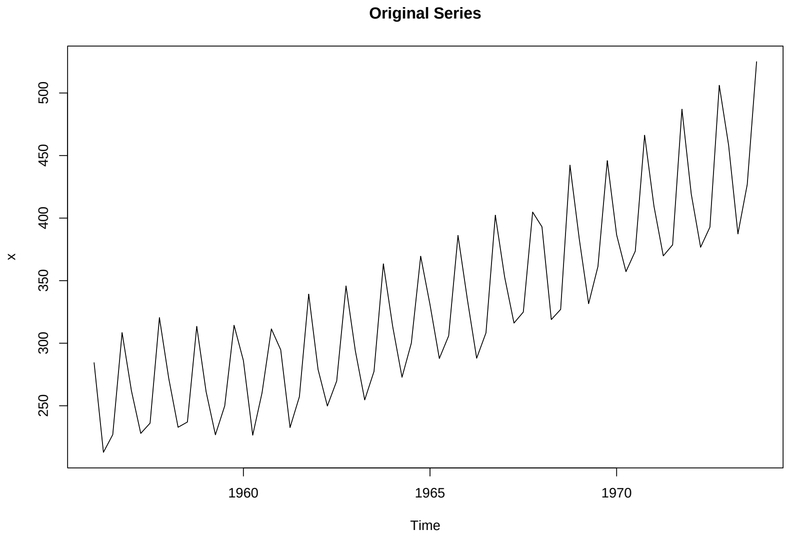
Medical Component of the CPI
t1
0
1
2
3
1950
1960
1970
1980
1990
2000
Time

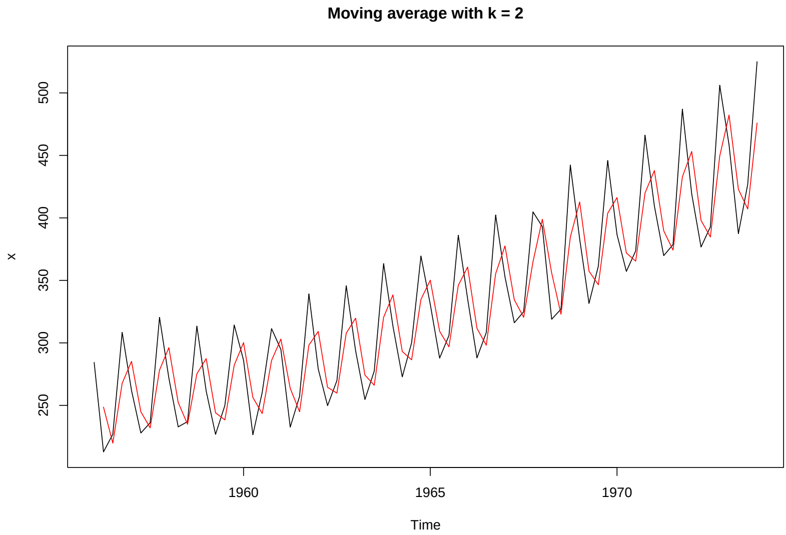

Moving average with k = 4
t1
0
1
2
3
Time
1950
1960
1970
1980
1990
2000

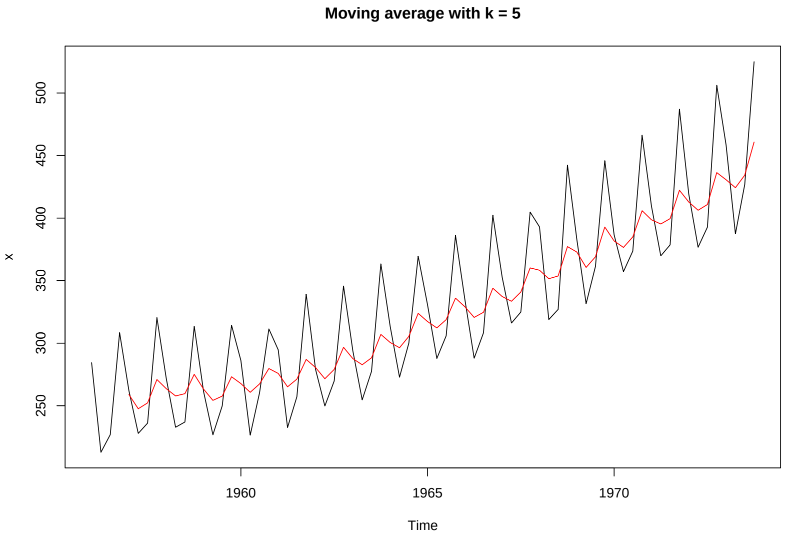

Moving average with k = 8
t1
0
1
2
3
Time
1950
1960
1970
1980
1990
2000

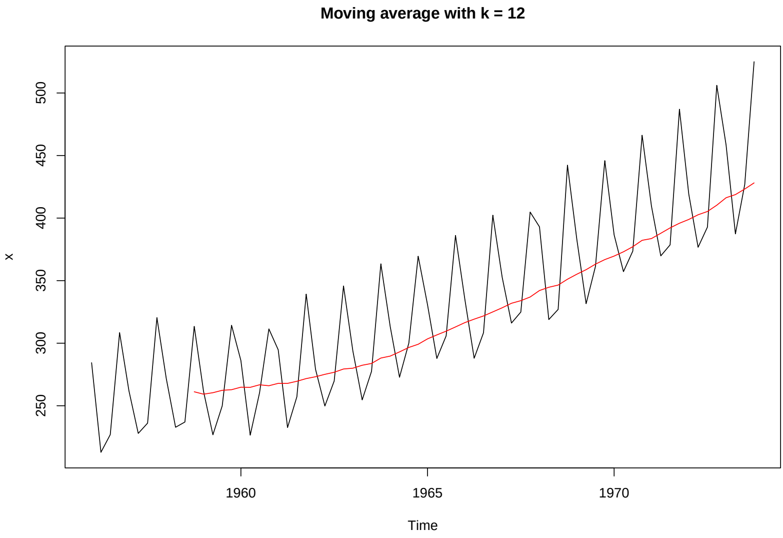


Moving average with k = 16
t1
0
1
2
3
1950
1960
1970
1980
1990
2000
Time


Original Series
x
500
450
400
350
300
250
1960
1965
1970
Time


Moving average with k = 2
x
Time
250
300
350
400
450
500
1960
1965
1970


Moving average with k = 5
x
500
450
400
350
300
250
1960
1965
1970
Time


Moving average with k = 12
x
250
300
350
400
450
500
1960
1965
1970
Time

Forecasting

- We can use smoothing for forecasting
- We have

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

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

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

- Step 2: Smooth the smoothed series

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

- 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

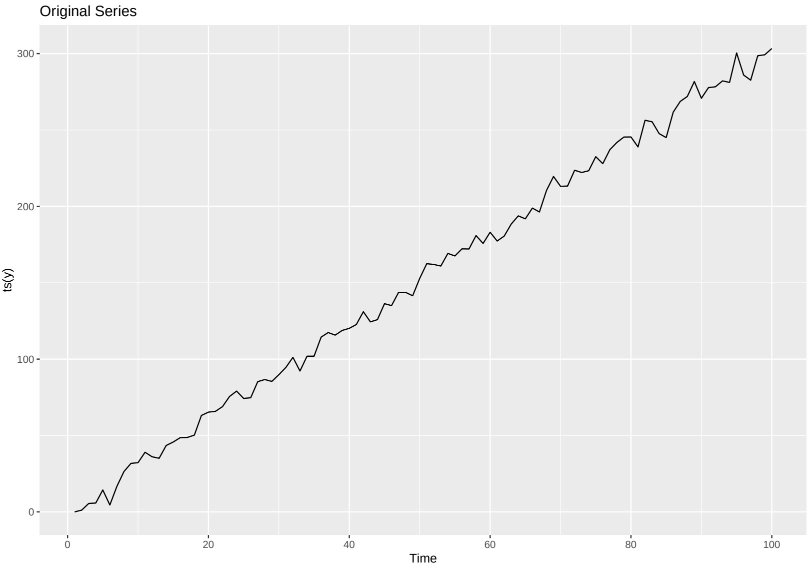
- Forecasting y_7 using simple moving average with $k = 2$
- Forecasting y_7 using double moving average with $k = 2$

Example

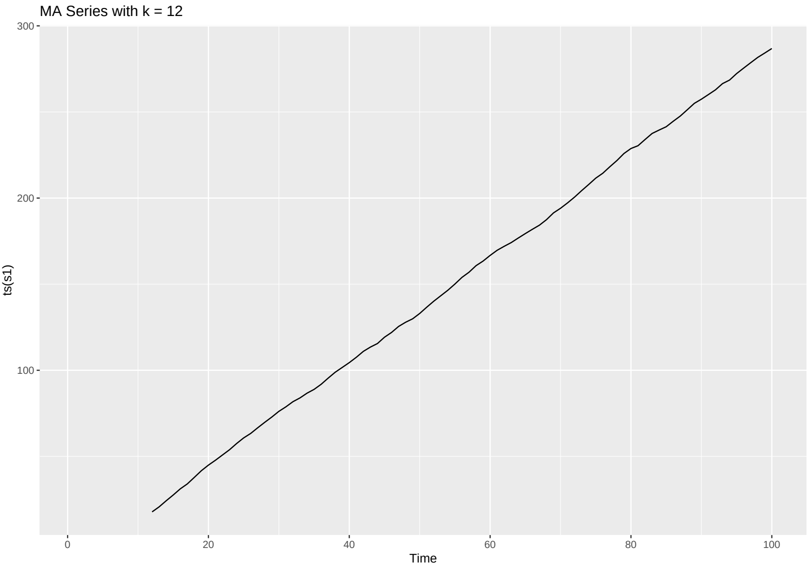
- We simulate 100 data points ($T = 100$) of

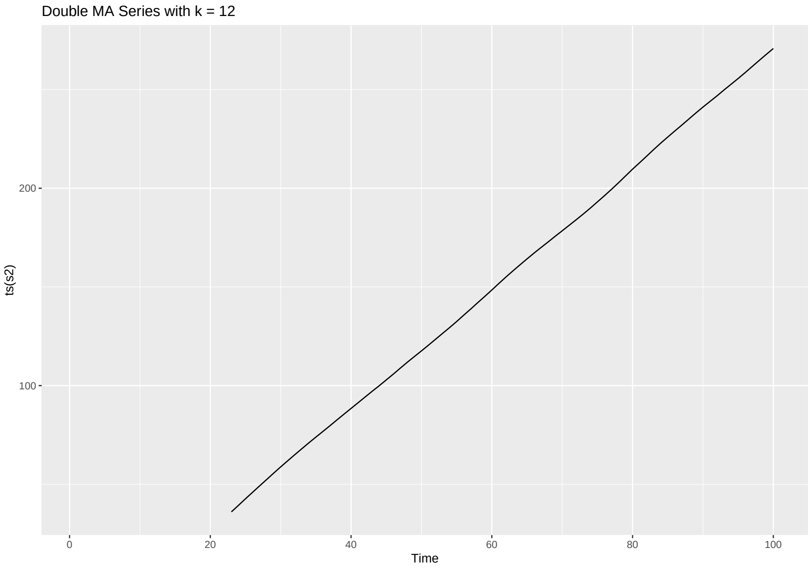
$$y_t = 1 + 3t + \epsilon,$$

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



Original Series
ts(y)
300
200
100
0
0
20
40
60
80
100
Time


MA Series with k = 12
300
200
100
ts(s1)
0
20
40
60
80
100
Time


Double MA Series with k = 12
200
100
ts(s2)
0
20
40
60
80
100
Time

- 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

$$\hat{s}_t = \frac{y_t + wy_{t-1} + w^2 y_{t-2} + ... + w^t y_0}{1/(1-w)}$$

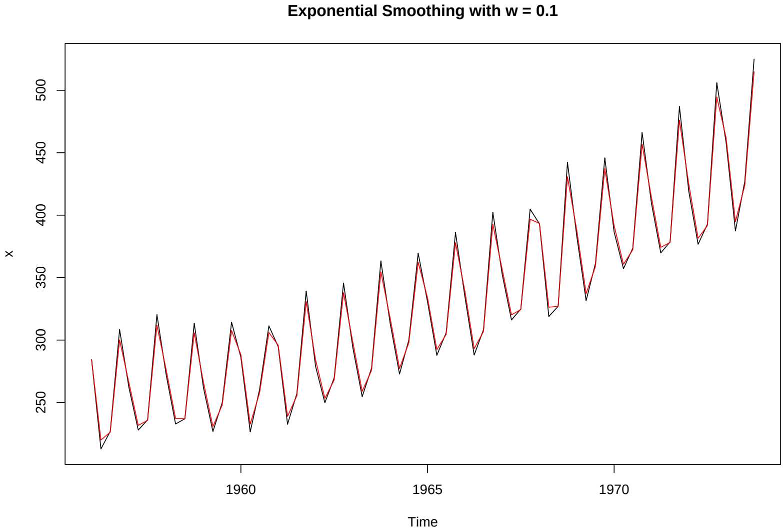
- Smaller w ($w \to 0$) gives higher weights to the more recent observations
- Smaller w smooths the series more lightly.

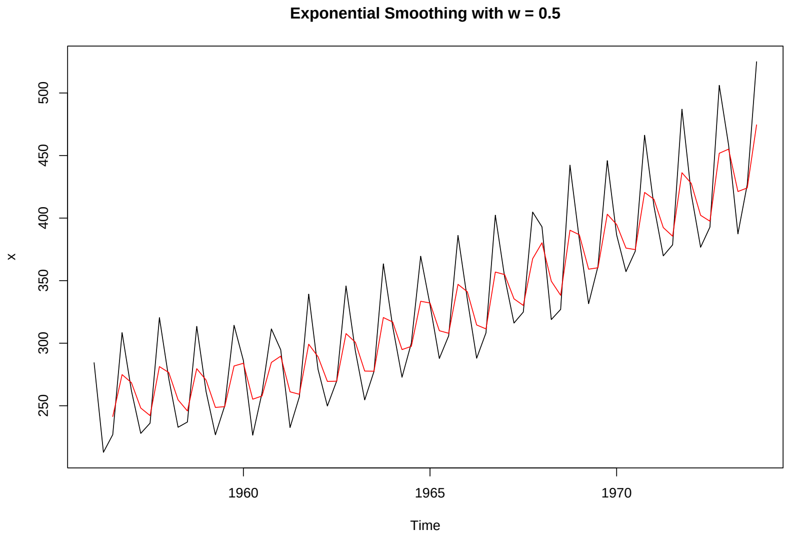
Exponential Smoothing

- We have

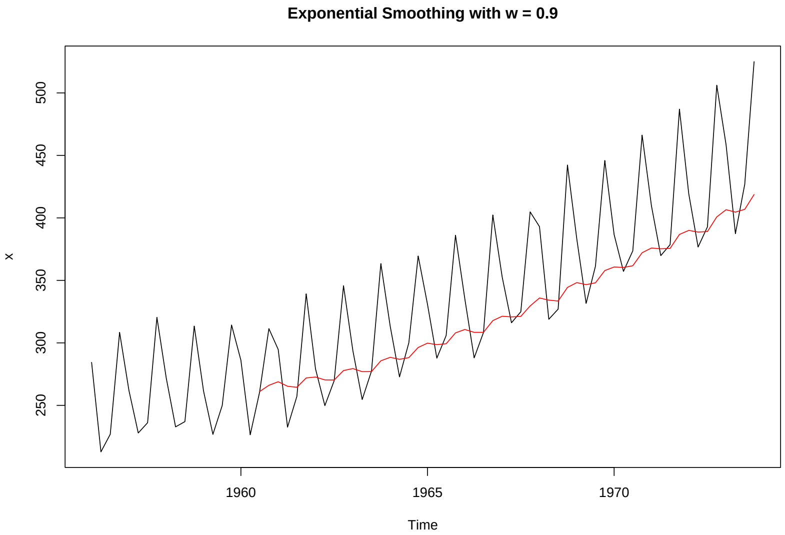
$$\begin{aligned}\hat{s}_t &= \hat{s}_{t-1} + (1-w)(y_t - \hat{s}_{t-1}) \\ &= (1-w)y_t + w\hat{s}_{t-1}\end{aligned}$$

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


Exponential Smoothing with w = 0.1
x
250
300
350
400
450
500
Time
1960
1965
1970



Exponential Smoothing with w = 0.5
x
500
450
400
350
300
250
1960
1965
1970
Time


Exponential Smoothing with w = 0.9
x
Time
250
300
350
400
450
500
1960
1965
1970

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 $w = .8$