THE UNIVERSITY OF TEXAS AT AUSTIN
McCOMBS SCHOOL OF BUSINESS

Time Series: Smoothing & Moving Averages

Lecture 19

STA 371G

Trend & Seasonality in Time Series

- Last time we said that autoregression requires that the time series be stationary, but most time series are not stationary.

Trend & Seasonality in Time Series

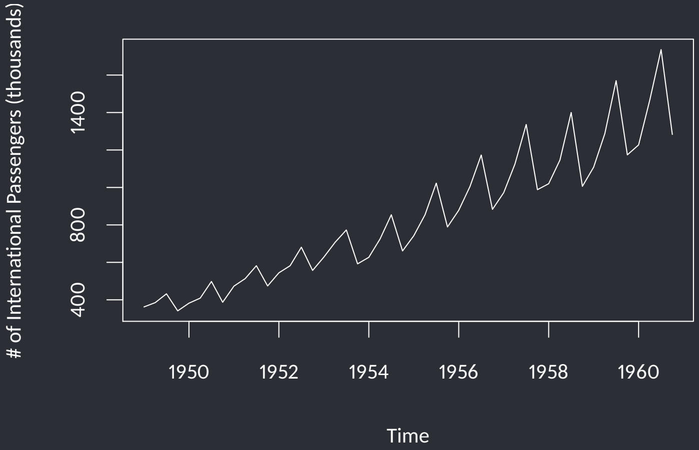
- Last time we said that autoregression requires that the time series be stationary, but most time series are not stationary.
- Two main factors that make data non-stationary: trends and seasonality.

Trend & Seasonality in Time Series

- Last time we said that autoregression requires that the time series be stationary, but most time series are not stationary.
- Two main factors that make data non-stationary: trends and seasonality.
- Sales, economic, activity, etc. data often show strong seasonality.

Airline Passengers 1949-1960

```
# Convert data into time series, starting from 1st quarter of 1949.
air <- ts(air_passengers$number, start=c(1949,1), frequency=4)
# Frequency: 4 data points per year (this is quarterly data)
plot(air, ylab="# of International Passengers (thousands)")
```



What is seasonality, exactly?

- Seasonality: The presence of variations that occur at specific regular intervals in a season.

What is seasonality, exactly?

- Seasonality: The presence of variations that occur at specific regular intervals in a season.
- In this case, a season is a year.

What is seasonality, exactly?

- Seasonality: The presence of variations that occur at specific regular intervals in a season.
- In this case, a season is a year.
- The frequency, the number of observations in a season, is 4.

What is seasonality, exactly?

- Seasonality: The presence of variations that occur at specific regular intervals in a season.
- In this case, a season is a year.
- The frequency, the number of observations in a season, is 4.
- The data behaves similarly in the same quarter of different years: every year, the number of passengers peaks in Q3 and dips in Q4.

Exploring the data

- Overall, there is an upward trend in the number of airline passengers from 1949-1960.

Exploring the data

- Overall, there is an upward trend in the number of airline passengers from 1949-1960.
- The seasonality in the data makes it difficult to see how the trend changes over years: is the number is increasing at the same rate every year, or does the rate vary in different years?

Exploring the data

- Overall, there is an upward trend in the number of airline passengers from 1949-1960.
- The seasonality in the data makes it difficult to see how the trend changes over years: is the number is increasing at the same rate every year, or does the rate vary in different years?
- The trend in the data makes it difficult to see the impact of seasonality in the data: is Q2 higher than the Q1 because it is a more active traveling period of the year or is it just that more people are using air travel as time goes by?

Exploring the data

- Overall, there is an upward trend in the number of airline passengers from 1949-1960.
- The seasonality in the data makes it difficult to see how the trend changes over years: is the number is increasing at the same rate every year, or does the rate vary in different years?
- The trend in the data makes it difficult to see the impact of seasonality in the data: is Q2 higher than the Q1 because it is a more active traveling period of the year or is it just that more people are using air travel as time goes by?
- Smoothing the data helps identify the trend and seasonal effects.

Smoothing the data: one-sided moving average

A (one-sided) moving average of span 4 (MA) is a simple average of all observations over the previous 4 time periods, and can be used to forecast the next time period.

$$y_t : \text{Number of passengers traveled at time } t$$

One-sided moving average at time t:

$$\hat{y}_{t+1} = \frac{1}{4}y_t + \frac{1}{4}y_{t-1} + \frac{1}{4}y_{t-2} + \frac{1}{4}y_{t-3}$$

Each quarter has equal weight in the moving average.

Smoothing the data: one-sided moving average

	Y49Q1	Y49Q2	Y49Q3	Y49Q4	Y50Q1	Y50Q2	...
t	1	2	3	4	5	6	...
y_t	362	385	432	341	382	409	...
$\hat{y}_t$	NA	NA	NA	NA	380	385	...

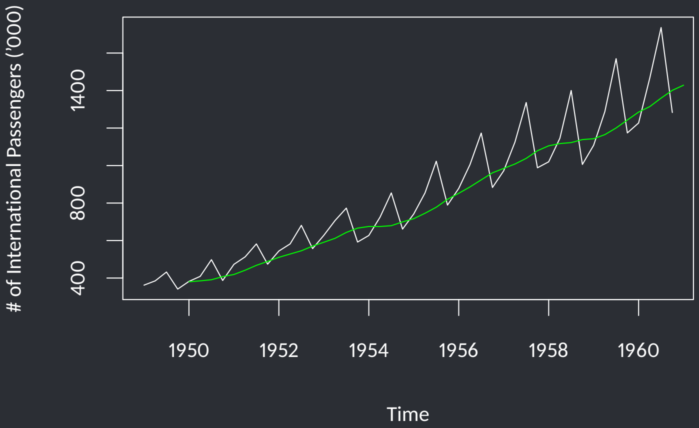
Smoothing the data: one-sided moving average

	Y49Q1	Y49Q2	Y49Q3	Y49Q4	Y50Q1	Y50Q2	...
t	1	2	3	4	5	6	...
y_t	362	385	432	341	382	409	...
$\hat{y}_t$	NA	NA	NA	NA	380	385	...

$$\begin{aligned}
\hat{y}_5 &= \frac{1}{4}y_1 + \frac{1}{4}y_2 + \frac{1}{4}y_3 + \frac{1}{4}y_4 \\
&= \frac{1}{4}362 + \frac{1}{4}385 + \frac{1}{4}432 + \frac{1}{4}341 \\
&= 380
\end{aligned}$$

Smoothing the data: one-sided moving average

```
# One sided moving average
air_ma_one <- (lag(air, -1) + lag(air, -2) + lag(air, -3) + lag(air, -4))/4
plot(air, ylab="# of International Passengers ('000)")
lines(air_ma_one, col='green')
```



Smoothing the data

Smoothed data better shows the slow-downs in the trend around 1954 and 1959.

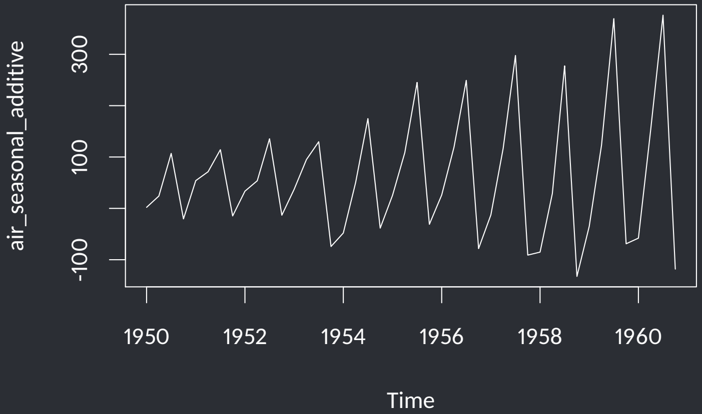
Let’s also see the effect of the seasonality by eliminating the trend in the data. This is called detrending.

Smoothing the data

When trend and seasonality are additive:

$$\text{Data} = \text{Trend} + \text{Seasonality} + \text{Randomness}$$

```
air_seasonal_additive <- air - air_ma_one
plot(air_seasonal_additive)
```

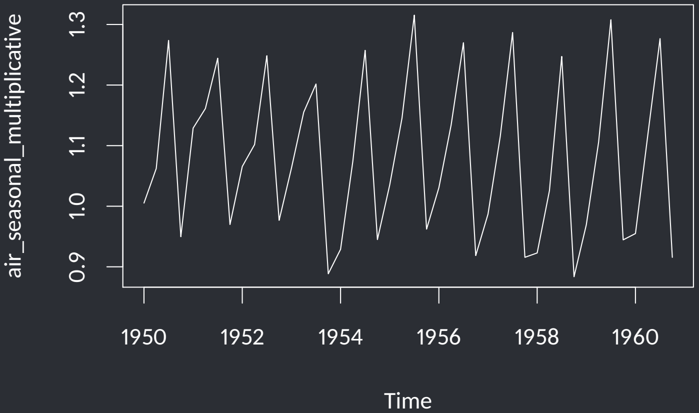


Smoothing the data

When trend and seasonality are multiplicative:

$$\text{Data} = \text{Trend} \times \text{Seasonality} \times \text{Randomness}$$

```r
air_seasonal_multiplicative <- air / air_ma_one
plot(air_seasonal_multiplicative)
```



Forecasting using Simple Exponential Smoothing

So far, we have smoothed the data to better observe the trend and the seasonality, and to make predictions about future values.

The problem with moving averages is that it ignores previous quarters beyond the last 4. We could extend the span to a larger number—but then we weight less recent data as much as more recent data!

Simple exponential smoothing (SES) is a way to deal with these issues.

Simple Exponential Smoothing

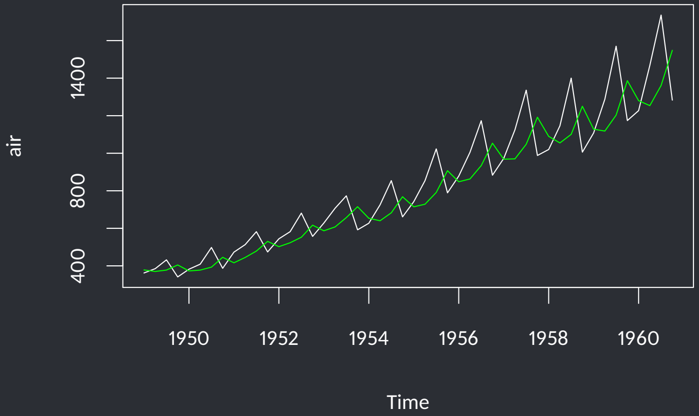
We'll use *every* previous data point to forecast the next one, but using a weight that decreases exponentially:

$$\begin{aligned}\hat{y}_{t+1} &= \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \cdots \\ &= \sum_{k=0}^{t-1} \alpha(1-\alpha)^k y_{t-k}.\end{aligned}$$

$\hat{y}_{t+1}$ carries a portion of all past observations; the more recent the observation is, the more weight it has (since $(1-\alpha)^k \to 0$ as k increases).

Simple Exponential Smoothing

```
library(forecast)
air_ses_05 <- ses(air, alpha=0.5)
plot(air)
lines(air_ses_05$fitted, col='green')
```

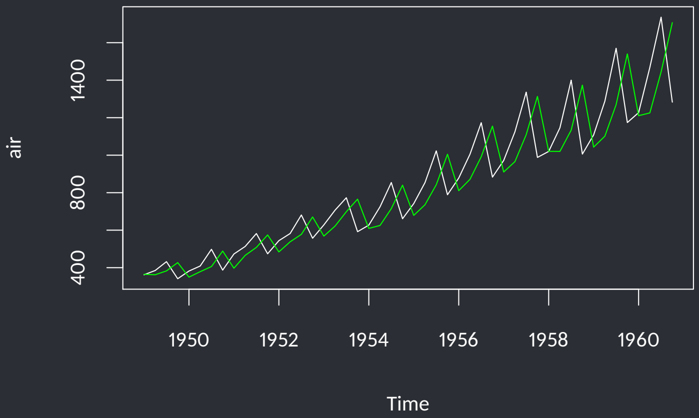


Simple Exponential Smoothing

- α determines how aggresively the forecast will react to the observed error.
- A larger α means a more aggressive adjustment (i.e. closer forecast to the observed value), but can also cause “overfitting.”

Simple Exponential Smoothing

```
air_ses_09 <- ses(air, alpha = 0.9)
plot(air)
lines(air_ses_09$fitted, col='green')
```



Simple Exponential Smoothing

- Regardless of the size of the α parameter, there seems to be always a delay in the forecast.

Simple Exponential Smoothing

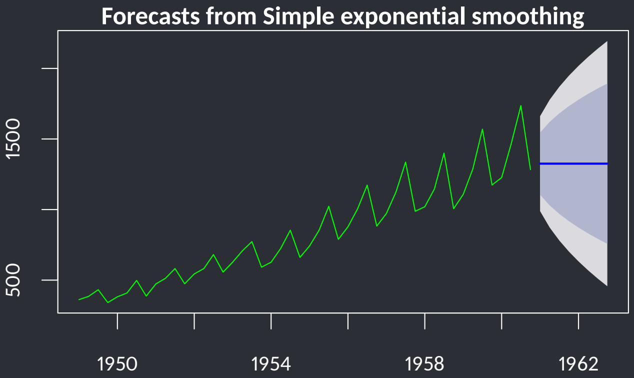
- Regardless of the size of the α parameter, there seems to be always a delay in the forecast.
- This is because the SES method is for stationary data, i.e., when there is no trend or seasonality.

Simple Exponential Smoothing

- Regardless of the size of the α parameter, there seems to be always a delay in the forecast.
- This is because the SES method is for stationary data, i.e., when there is no trend or seasonality.
- When trend and seasonality exist, SES (like MA) predicts the same value for all future observations after a certain point.

Simple Exponential Smoothing

```
air_ses_09 <- ses(air, alpha = 0.9, h=8)
plot(air_ses_09, col='green')
```



Simple Exponential Smoothing

```
air_ses_09

        Point Forecast     Lo 80    Hi 80    Lo 95    Hi 95
1961 Q1       1325.377 1105.3915 1545.363 988.9380 1661.817
1961 Q2       1325.377 1029.4167 1621.338 872.7445 1778.010
1961 Q3       1325.377  969.2991 1681.456 780.8027 1869.952
1961 Q4       1325.377  917.9579 1732.797 702.2830 1948.472
1962 Q1       1325.377  872.3988 1778.356 632.6064 2018.148
1962 Q2       1325.377  831.0206 1819.734 569.3240 2081.431
1962 Q3       1325.377  792.8480 1857.907 510.9440 2139.811
1962 Q4       1325.377  757.2343 1893.520 456.4776 2194.277
```

What if there is a trend and/or seasonality?

- We can use Holt's Method (when there is a trend) or the Holt-Winter's Method (when there is also seasonality), which extend SES.

What if there is a trend and/or seasonality?

- We can use Holt's Method (when there is a trend) or the Holt-Winter's Method (when there is also seasonality), which extend SES.
- Holt's Method estimates the trend, and smooths it (with smoothing parameter β).

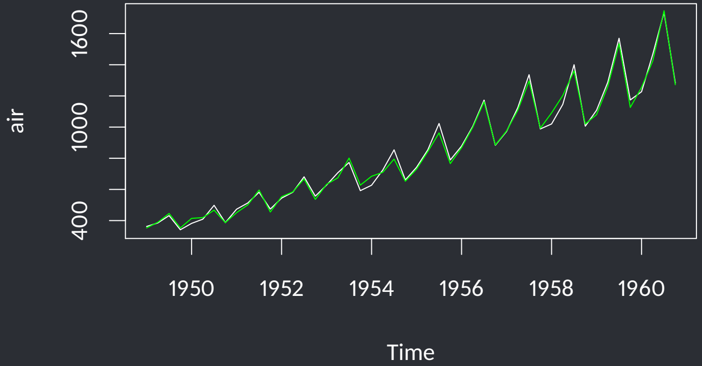
What if there is a trend and/or seasonality?

- We can use Holt's Method (when there is a trend) or the Holt-Winter's Method (when there is also seasonality), which extend SES.
- Holt's Method estimates the trend, and smooths it (with smoothing parameter β).
- Winter's Method estimates the seasonal components, and smooths them (with smoothing parameter γ).

What if there is a trend and/or seasonality?

We can use `holt` when there is a trend, or `hw` when there is both a trend and seasonality. You can specify α, β, and/or γ, or omit them and let R find the optimal values that minimize prediction error:

```
air_hw <- hw(air, seasonal='multiplicative')
plot(air)
lines(air_hw$fitted, col='green')
```



```r
plot(air_hw, col='white')
```