MSCI152: Introduction to Business Intelligence and Analytics

Lecture 14: Forecasting: Moving Averages and Exponential Smoothing

Dr Anna Sroginis

Lancaster University Management School

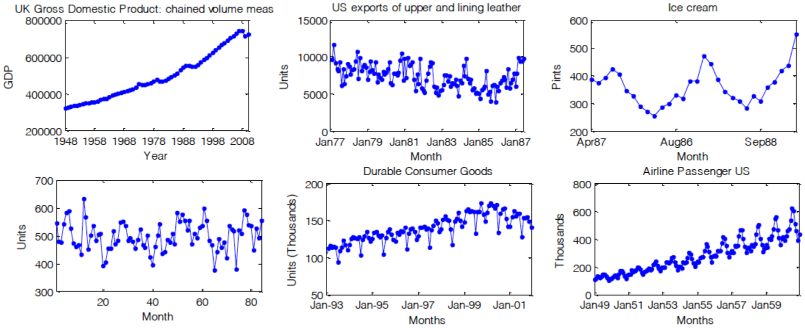
Agenda

1. Time series exploration
   - Identifying the presence of trend
   - Identifying the presence of seasonality
2. Simple forecasting methods
   - Naïve or random walk
   - Average forecast
   - Moving Average
3. Exponential Smoothing

More details can be found in Camms et al., Section 8.3, 8.4, 8.5

Time series components

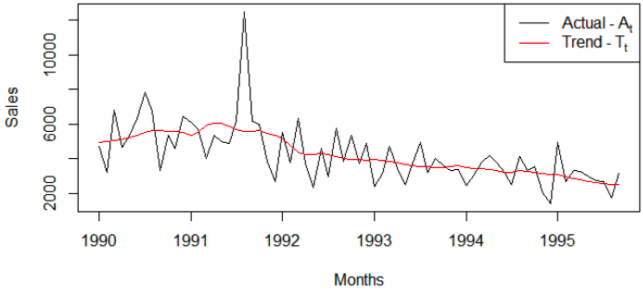
We know that time series can have several different structural forms. We are mainly looking for the **existence** and **behaviour** of: **Trend** (including level/cycle), **Season** and **Irregularities**.



Any components in these time series?

Identifying the presence of trend

To identify trend, we have to fit a smooth line through the time series and observe whether this line is going up or down.



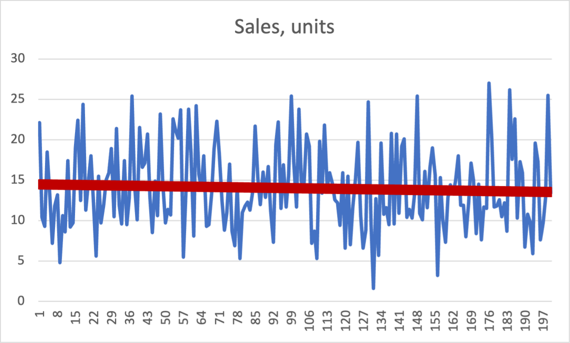
We need to fit a flexible enough smooth line that can capture various versions of trend, cycle or lack of them. For this purpose, we can use the **Centred Moving Average (CMA)**.

Fitting a straight (regression) line, often called a "trendline" is wrong. Why?

Identifying the presence of trend

In Excel, there is **trendline** function.

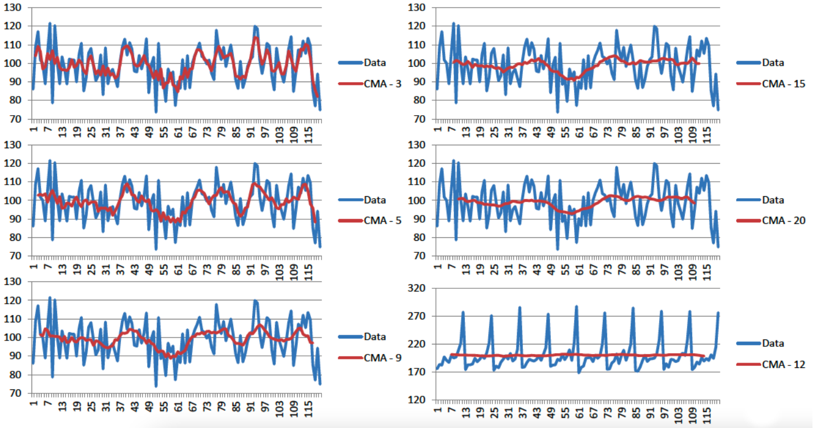
- **trendline** works using least-squares linear regression
- For non-seasonal data, this is appropriate only if the trend is linear and not changing over time
- For seasonal data, it is not appropriate, as the two elements of trend and seasonality need to be decomposed.



Identifying the presence of trend

The centred moving average provides a smoothed version of the time series

- can filter out noise (longer CMA filters more)
- can filter out seasonality (CMA window must be a multiple of the season period).



Calculation of the Centred Moving Average

Period t	Time Series	CMA (3)			Time Series	CMA (4)	
1	86			½	86		
2	109	104.0	Middle value	1	109		
3	117	109.3		1	117	105.3 ?	Middle value
4	102	106.3		1	102	104.5	
5	100	97.0		½	100	100.6	
6	89	98.3			89	101.8	
7	106	105.7			106	101.6	
8	122	102.3			122	102.9	
9	79	107.0			79		
10	120				120		

$$\text{CMA (3)} = \frac{X_1 + X_2 + X_3}{3} \qquad \text{CMA (4)} = \frac{X_1/2 + X_2 + X_3 + X_4 + X_5/2}{4}$$

Careful with even numbers!

The resulting CMA is considered as the value of the trend/cycle (including level) at time t

Centred Moving Average

- 'Moving' because it moves as t changes to t+1
- 'Centred' because it is allocated to the central time period in the calculation
- CMA is **not** a forecast.

For non-seasonal data:

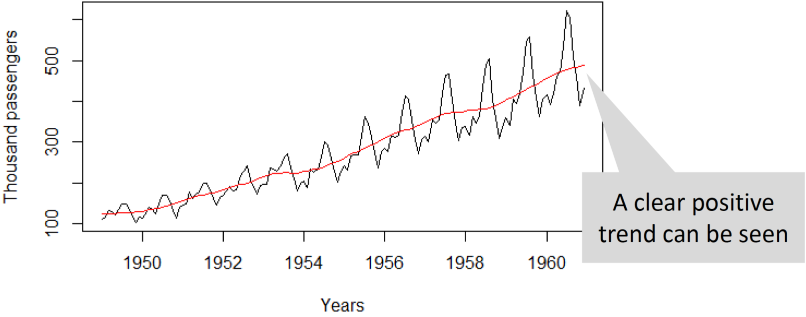
- $CMA(x)_t$ is an estimate of the trend component T_t
- Longer lengths of moving average (x) will tend to reduce the noise but will lose more estimates at the beginning and end of the history.

For seasonal data:

- $CMA(s)_t$ is an estimate of the trend component T_t where s is the length of the seasonal cycle.

Calculating CMA

For example, let us consider the "**monthly** airline passenger" time series.
How long should the CMA be?

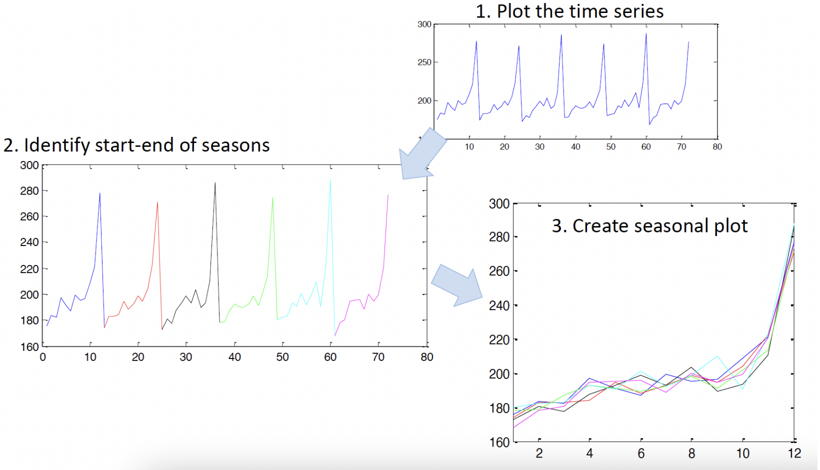


- Think if the time series was seasonal, how many periods would you have in a year? That is the correct length of the CMA.
  - A monthly series has 12 months in a year, so you need a CMA of 12, irrespective of whether the time series is seasonal or not.

Identifying the presence of seasonality

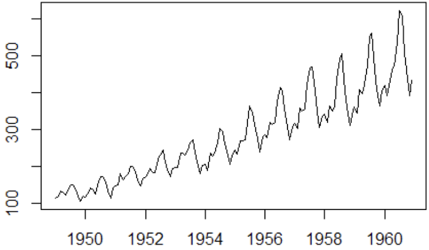
To detect seasonality in a time series, we simply examine whether there is a repeating pattern with the expected frequency.

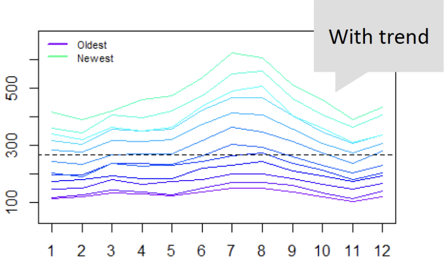
- In Excel, just plot each season separately with a line on the same plot

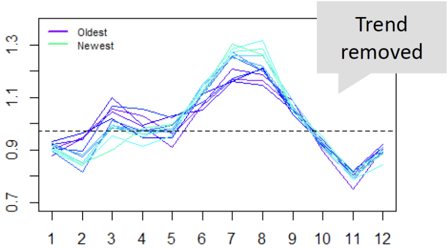


Identifying the presence of seasonality

What happens when there is a trend?



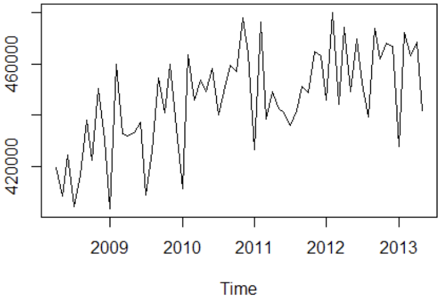


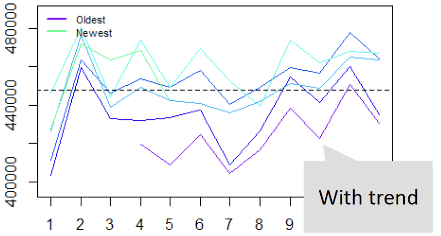


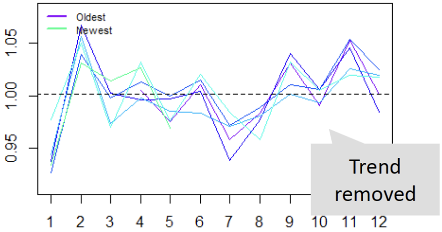
Removing any identified trend first helps to identify seasonality

Identifying the presence of seasonality

What happens when there is a trend?







Removing any identified trend first helps to identify seasonality

Agenda

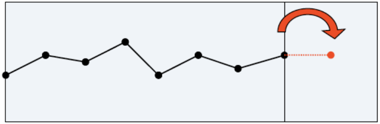
1. Time series exploration
   - Identifying the presence of trend
   - Identifying the presence of seasonality
2. Simple forecasting methods
   - Naïve or random walk
   - Average forecast
   - Moving Average
3. Exponential Smoothing

Naïve (or Random Walk)

What is the simplest forecast you can think of for a time series?

That it will remain the same! This is called the **Naïve** forecast (or **Random Walk**)

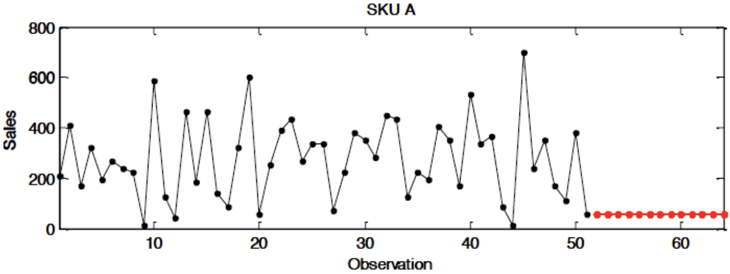
$$F_{t+1} = A_t$$



- Observe that it has no parameters to set and requires only 1 observation.
- It is the simplest possible forecast we can produce.
- For this reason, it is a very good benchmark. If we cannot best it no need for more complex forecasts!

Naïve forecasts

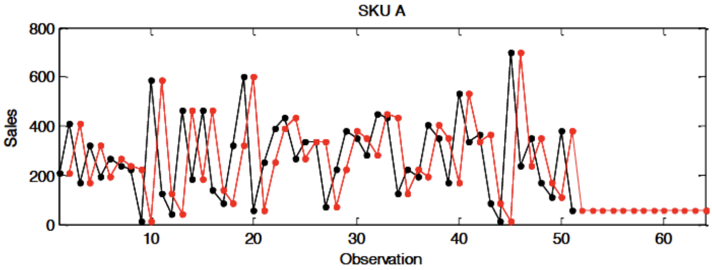
A naïve forecast for SKU A for the next 13 weeks (3 months planning lead time):



- The forecast is a straight line - always equal to the last observation
- Is this a good forecast?

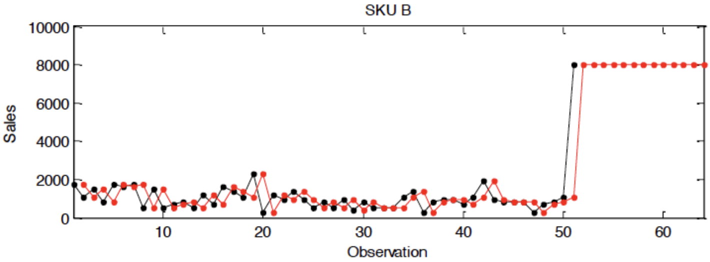
Naïve forecast

Let us add Naïve forecasts for each observation:



- We can see that the forecast always lags behind by one month.

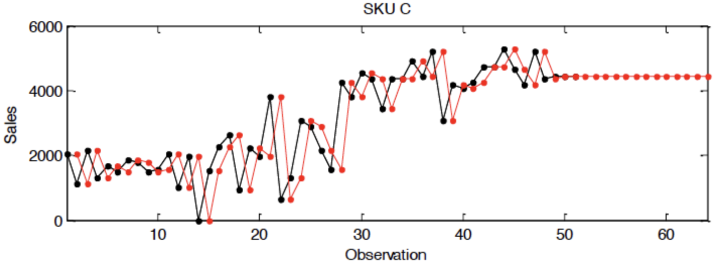
What about this product?



- The naïve is not robust against outliers, which fully affect its forecasts.

Naïve forecast

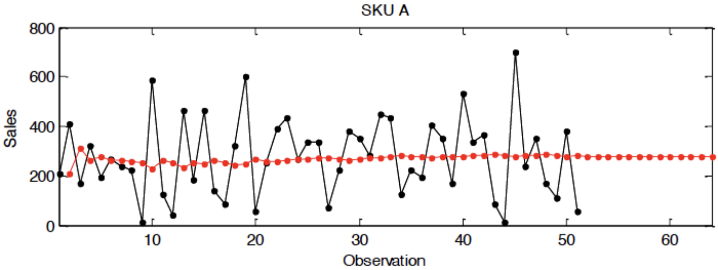
And this product?



- The naïve is very fast at picking up level shifts and other dramatic changes.

Remember that the naïve has a memory of only 1 observation.

The average has long memory and the random movements of the noise will be cancelled out:

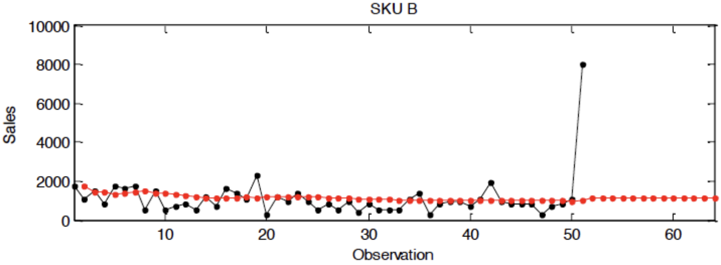


- The more available data, the better the correct level is estimated.
- Gives equal importance to all observations.

In contrast to the naïve, the arithmetic mean is very robust against noise for various degrees of it.

Average forecast

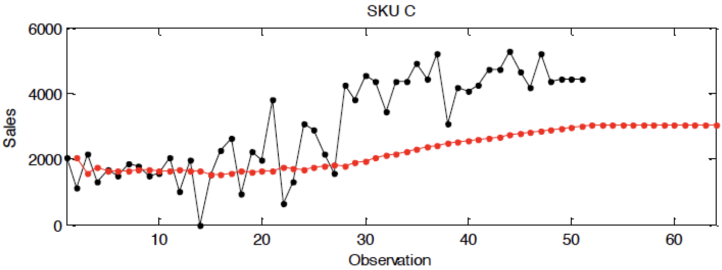
What about this product?



- The arithmetic mean is very resilient against outliers and still provides a good estimation of the level.

Average forecast

And this product?



- However, due to its long memory it does not adjust to changes in the level, producing very poor forecasts when this happens.

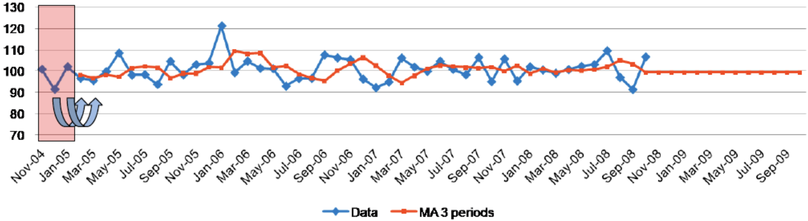
Moving Average

So far,

- Naïve: memory too short
- Arithmetic Mean: memory too long

Moving Average allows us to select the appropriate memory (length of the average).

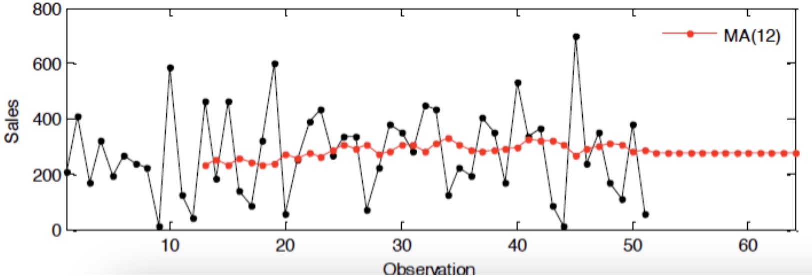
For example, Moving Average of 3 periods is calculated:



$$(\text{Nov} + \text{Dec} + \text{Jan})/3 = \text{Feb} \Rightarrow (101 + 91 + 102)/3 = 98$$
$$(\text{Dec} + \text{Jan} + \text{Feb})/3 = \text{Mar} \Rightarrow (91 + 102 + 97)/3 = 96.67$$

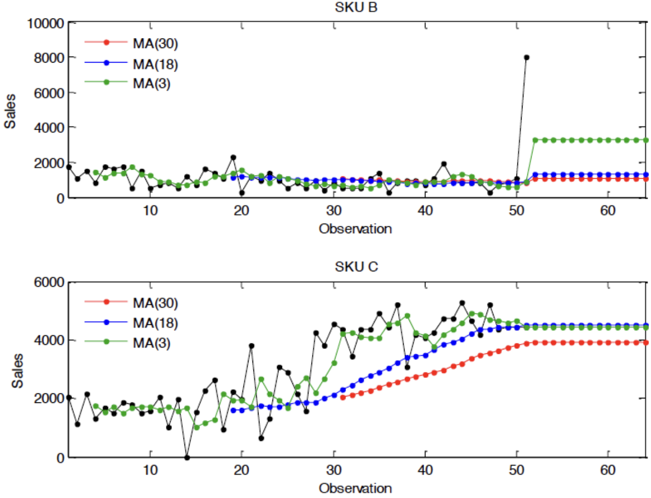
Moving Average

- Its variable length allows us to control how reactive we are to new information and how robust we are against noise.
- Gives equal importance to all k observations.



- We need to choose the moving average that gives us a smooth estimate of the level, here MA(12). But do not use excessive moving average lengths (it is too insensitive to new information).

Moving Average



Long averages result in robust forecasts against noise and outliers. **Short averages** result in very reactive forecasts that adjust quickly to new information.

Centred Moving Average vs Moving Average

Centred Moving Average

- CMA is a descriptive tool. It is a **fitted** value (e.g. CMA Jan, Feb, Mar = fitted value for February)
- Requires future information (it is centred).
- Difference in calculation between even and odd lengths.

Moving Average

- MA is a predicitve tool. It is a **forecasted** value (e.g. SMA Jan, Feb, Mar = forecasted value for April)
- Does not require any future information.
- No difference in the calculation of even or odd lengths.
- Is not a good estimation of the trend.

Summary so far

1. Naive
    - Cannot filter noise
    - Very short memory
2. Arithmetic Mean
    - Very long memory $\Rightarrow$ does not forget history
    - Sometimes not reactive enough due to its long memory
    - All observations are weighted equally
3. Moving Average
    - We can control how reactive the forecast is by changing its length
    - All observations are weighted equally

Agenda

Exponential Smoothing

Moving Average puts equal weight on each value, but we can argue that if next month's forecast is to be based on the previous 12 months' observations, **more weight should be placed on the more recent observations**.

Exponential smoothing uses a weighted average of past time series values as a forecast.

The exponential smoothing model is as follows:

$$\hat{y}_{t+1} = \alpha y_t + (1 - \alpha)\hat{y}_t \tag{1}$$

where

- $\hat{y}_{t+1}$ = forecast of the time series for period $t + 1$
- y_t = actual value of the time series in period t
- $\hat{y}_t$ = forecast of the time series for period t
- α = smoothing constant ($0 <= \alpha <= 1$)

Exponential Smoothing

- More recent observations are more important, therefore weighted more heavily.
- But weights must add up to 100% (or 1)

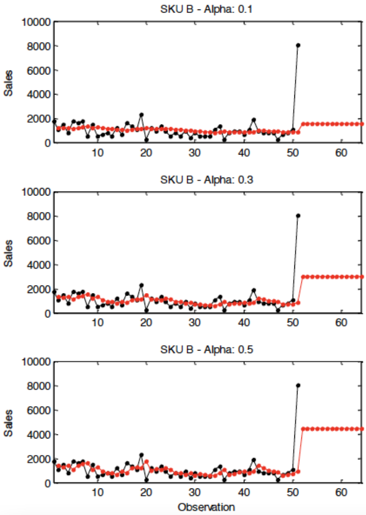
The **exponential smoothing** formula can be read as:

- the forecast is α times the most recent observation and $(1-\alpha)$ times all the previous information.
  - A low α: more emphasis on older information
  - A high α: more emphasis on newer information

Therefore the smoothing parameter α controls how reactive is the forecast to new information.

Empirically, **alpha must be $>$ 0**, but often **smaller than 0.5**
An appropriate alpha results in a **smooth in-sample fit** that goes through the middle of our time series.

Single Exponential Smoothing

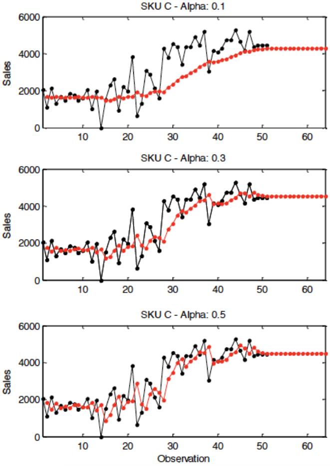


Alpha = 0.1: less sensitive to last observation (outlier). Robust.

Alpha = 0.3: more sensitive to the outlier . Less robust.

Alpha = 0.5: even more sensitive to the outlier. Not robust.

Single Exponential Smoothing



Alpha = 0.1: too slow to adjust to the new level of sales.

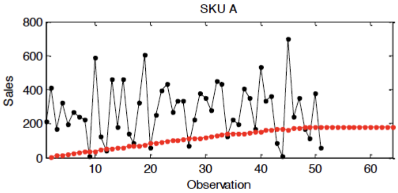
Alpha = 0.3: a good balance between reactivity and robustness to noise.

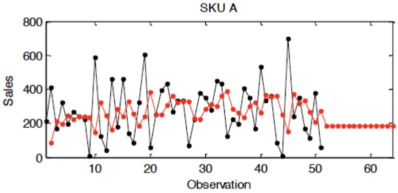
Alpha = 0.5: forecast reacts quickly, but it does not filter out noise adequately.

Single Exponential Smoothing

What do we need to produce an exponential smoothing forecast?

- We need the latest **Actual** and a **Forecast** for the previous time period and the **alpha** parameter, which we choose.
- We cannot calculate a forecast, unless we have an **initial forecast**, where does the first forecast come from?
  - Use a different forecasting model for the first forecast (e.g., naive, simple Moving Average etc.)
  - It is important to have a reasonable initial value!



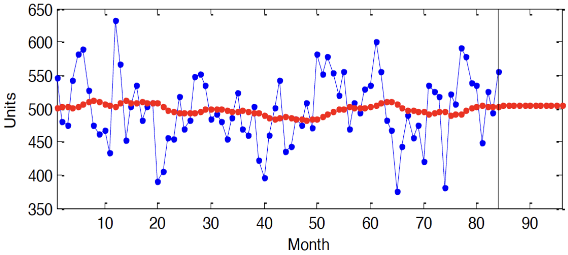


To counter the poor initial value we are forced to use higher smoothing parameters, which in turn do not filter the noise adequately.

Does a straight line forecast make sense?

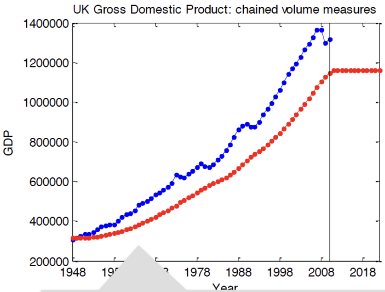
Yes! Several reasons:

1. The time series has no trend or season, therefore the only structure is level.
2. Noise cannot be forecasted. A "wiggly" line trying to forecast noise will only reduce accuracy.
3. The last level estimate: the forecast going forward contains all the useful information up to that point, therefore the straight line going forward does too.



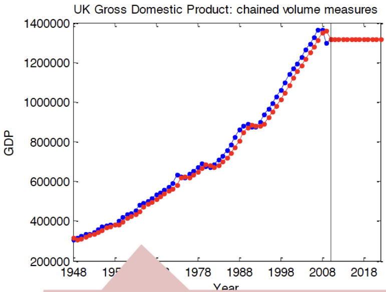
Single exponential smoothing cannot model trended time series

Alpha = 0.2



The method cannot output trend, a reasonable alpha provides a poor fit and bad forecast.

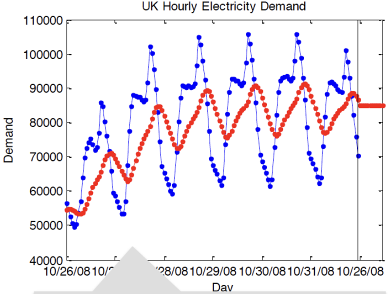
Alpha = 0.7



A very high alpha seems to provide a good fit, however the forecast is poor and the noise is not smoothed as the average is too short!

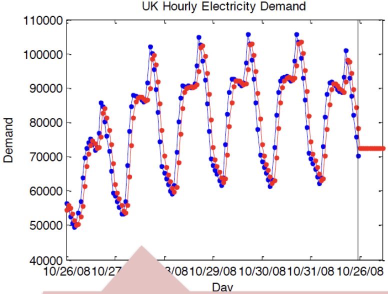
Single exponential smoothing cannot model seasonal time series

Alpha = 0.2



The method cannot output seasonality resulting in a poor fit and bad forecast.

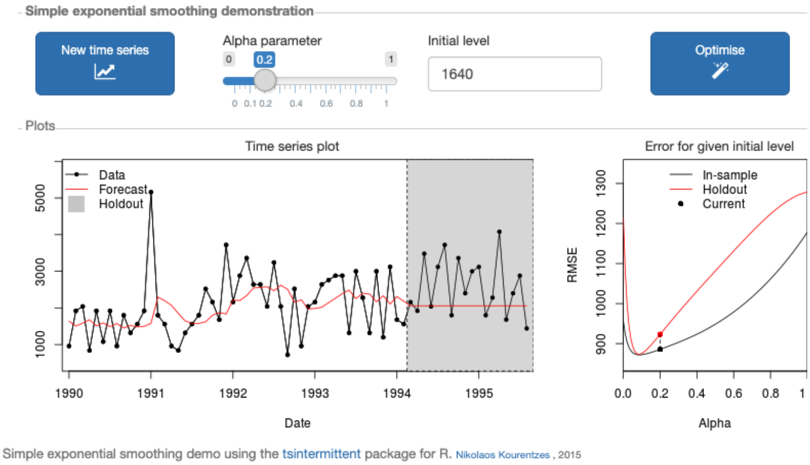
Alpha = 0.7



A very high alpha cannot compensate, forecasts lags by 1 observation and the noise is not smoothed.

Simple Exponential Smoothing: Online example

You can use the interactive tool at https://kourentzes.shinyapps.io/shinySES for testing SES with different series/parameters.



Other types of exponential smoothing

There are many variations of exponential smoothing:

- If there is a trend but no seasonality $\Rightarrow$ Holt's method
- If there is seasonality $\Rightarrow$ Winters' method
- If there are a trend and seasonality $\Rightarrow$ Holt-Winters models

But you can choose MSCI 381 Business Forecasting to learn more!

Wrap up

Here we:

- Introduced CMA, naive, moving averages and ETS;
- Defined main components of time series

Next time:

- **Dashboards and modelling in practice**