

# Forecasting Methods

---

- ⇒ What is forecasting ?
  - ⇒ Why is forecasting important ?
  - ⇒ How can we evaluate a future demand ?
  - ⇒ How do we make mistakes ?
-

## Contents

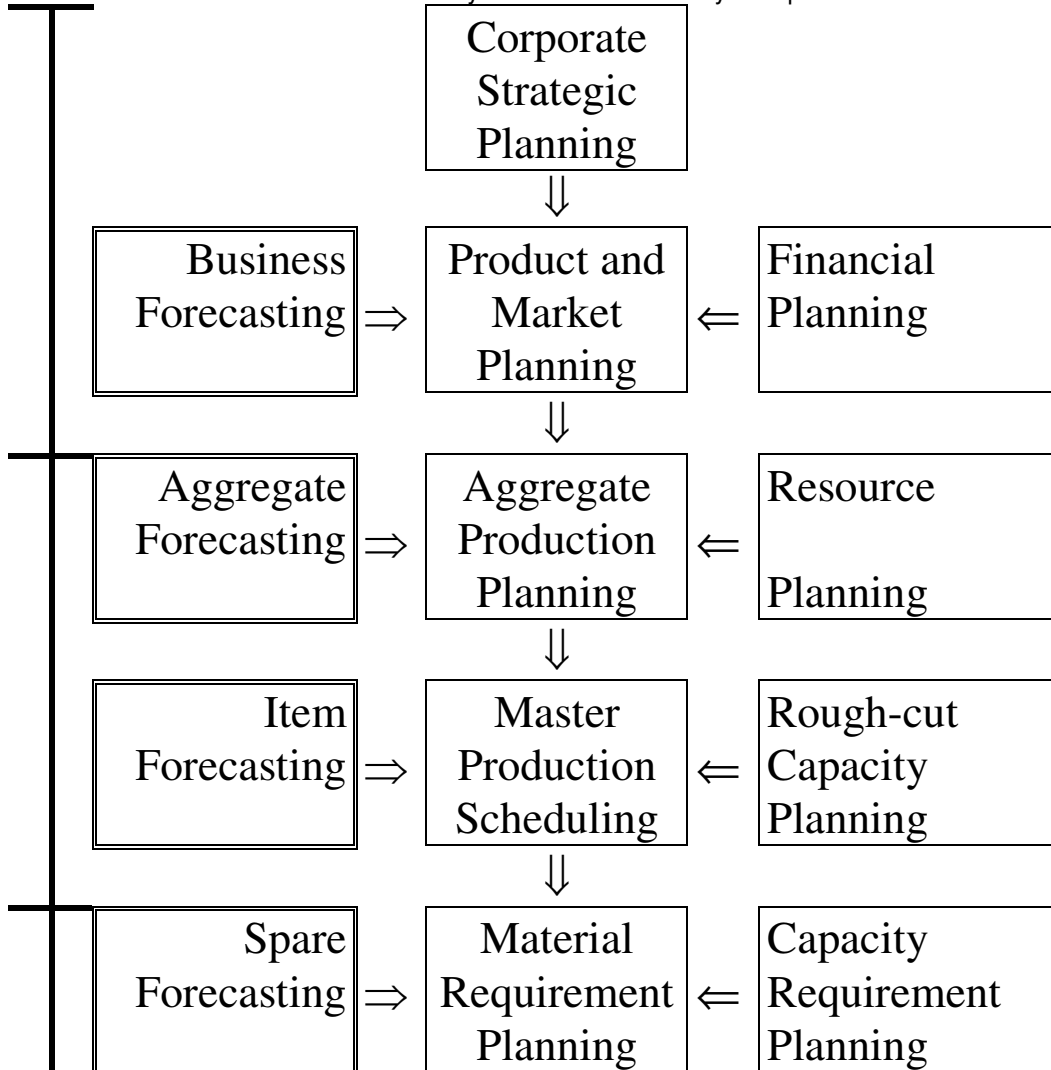
<b>1. FRAMEWORK OF PLANNING DECISIONS .....</b>	<b>2</b>
<b>2. FORECASTING.....</b>	<b>3</b>
2.1 CHARACTERISTICS.....	3
2.2 TYPE OF FORECASTS: .....	4
<b>3. TIME SERIES METHODS.....</b>	<b>5</b>
3.1 EXAMPLES.....	5
3.2 PRINCIPLES.....	7
<b>4 STATIONARY TIME SERIES.....</b>	<b>8</b>
4.1 MOVING AVERAGE:.....	8
4.2 WEIGHTED AVERAGE: .....	9
4.3 EXP. SMOOTHING: .....	9
4.4 COMPARISON BETWEEN MA AND ES .....	10
<b>5 TREND-BASED TIME SERIES.....</b>	<b>11</b>
5.1 LINEAR REGRESSION .....	12
5.2 DOUBLE EXP. SMOOTHING (HOLT).....	13
5.3 COMPARISON ( LIN. REGRESSION - 2ES).....	15
<b>6. SEASONAL TIME SERIES .....</b>	<b>16</b>
6.1 CMA METHOD .....	17
6.2 WINTERS' METHOD .....	21
<b>7. EVALUATION OF FORECAST .....</b>	<b>22</b>
<b>8. OTHER METHODS .....</b>	<b>23</b>
<b>9. WHICH METHOD TO USE ?.....</b>	<b>23</b>

---

Makridakis S. and S.C. Wheelwright, *Forecasting Methods for Management*, John Wiley & Sons, 1989.

# 1. Framework of planning decisions

Let us first remember where the inventory control decisions may take place.



Here are the successive decision steps by which a company finally manufactures a product. These steps have been analyzed in details in the previous chapters. Roughly, the strategic planning defines which products to manufacture, how and where. The aggregate planning looks at the intermediate term and selects the best policy to cope with the fluctuation of the global demand. Finally, the master production schedule defines the production activities in shorter terms (a few weeks) and per products.

Each of these planning steps requires some kind of forecasting which differs by the term which is considered and the product scope.

The business forecasting considers the long term. It focuses on product lines. The aggregate forecasting considers the aggregated (in terms of products) demand for each of the 12 -24 coming months. The item forecasting is an estimation of the demand for each item in the coming weeks. The need for spares is also required for the MRP.

## ⇒ Demand Forecasting

The independent demand is driven by the market. You do not know exactly how many cars will be sold in the next period. Here you need some forecasting.

The dependent demand is driven by the demand of another product. You do not know how many wheels you need, but it is always 5 times the number of cars.

## 2. Forecasting

Here are listed the main features of forecasting.

### 2.1 Characteristics

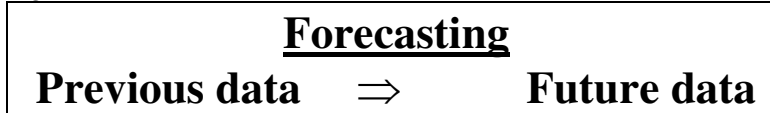
First, a causal relationship is needed. If what happens is purely random and does not depend on anything, you cannot predict what will happen.

On the other hand, if you observed some correlations between some variables, you could use these correlations to make some forecasts.

- **based on causal relationship**

If a student got A grades only during his/her first two years, the probability that he/she fails an exam is smaller than the average.

If the weather gets hot, the ice cream sales will increase.



- **usually wrong**

The sentence: "Forecasting is difficult, especially about the future" is clear.

- **requires more than one number**

The forecast is the value which is looked for, but some idea about its probability distribution is necessary.

- **aggregate is better**

If you have to forecast some demand, you might find that it has the distribution of Gauss (normal) with average  $\mu$  and standard deviation  $\sigma$ . The coefficient of variation of the prediction is thus  $\sigma/\mu$ . If you have to predict the sum of two such demands, you will find a coefficient of variation  $0.71 \sigma/\mu$ , where  $0.71 = \sqrt{2}/2$ . The sum of three such demands has a coefficient of variation  $0.58 \sigma/\mu$ , etc.

- **long horizon  $\Rightarrow$  large errors**

So many things can happen.

- **use different approaches**

You gain confidence in your forecast if you can find it by different ways.

- **use any other known information**

Any additional information must be incorporated. You cannot estimate the future demand accurately if you do not take into account facts which influence the demand. Two examples:

- an advertising or promotion campaign has been launched;
- a new product which replaces the old one is now available.

## 2.2 Type of Forecasts:

Forecasts can be obtained in different ways.

- **qualitative**

These approaches are based on judgments and opinions. Here are four examples.

1. Ask the guy in contact with the customers. Compile the results level by level.
2. Ask a panel of people from a variety of positions. Derive the forecasts and submit again. An example is the Delphi method used by the Rand Corporation in the 1950s.
3. Perform a customer survey (questionnaire or phone calls).
4. Look how similar products were sold. For example, the demand for CDs and for CD players are correlated. Washing machine and dryers are also related.

- **time series analysis**

What we will analyze in details. The idea is that the evolution in the past will continue into the future.

**Time series:**    **stationary**  
                         **trend-based**  
                         **seasonal**

Different time series will be considered: stationary, trend-based and seasonal. They differ by the shape of the line which best fits the observed data.

**Methods :**        **moving average**  
                         **regression**  
                         **exponential smoothing**

The methods which can be used are (linear) regressions, moving averages and exponential smoothings. They differ by the importance they give to the data and by their complexity.

- **causal relationship**

Here one tries to verify whether there is some causal relationship between some variables and the demand. If this is the case and if the variable is known in advance, it can be used to make some forecast. For example, there is a correlation between the number of building permits which are delivered and the demand for wall paper.

- **simulation**

Here a dynamic model which incorporates all the relevant variables is designed and programmed. This model is supposed to incorporate all the internal and external important variables. The model is then used to test different alternatives such as what happens: if the price is reduced by so much or if an advertisement campaign is started?

### Notations

$D_t$	Observed demand in period $t$
$F_{t,t+\tau}$	Forecast made at $t$ for period $t+\tau$
$F_t = F_{t-1,t}$	Forecast made at $t-1$ for period $t$
$e_t = F_t - D_t$	Forecast error in period $t$

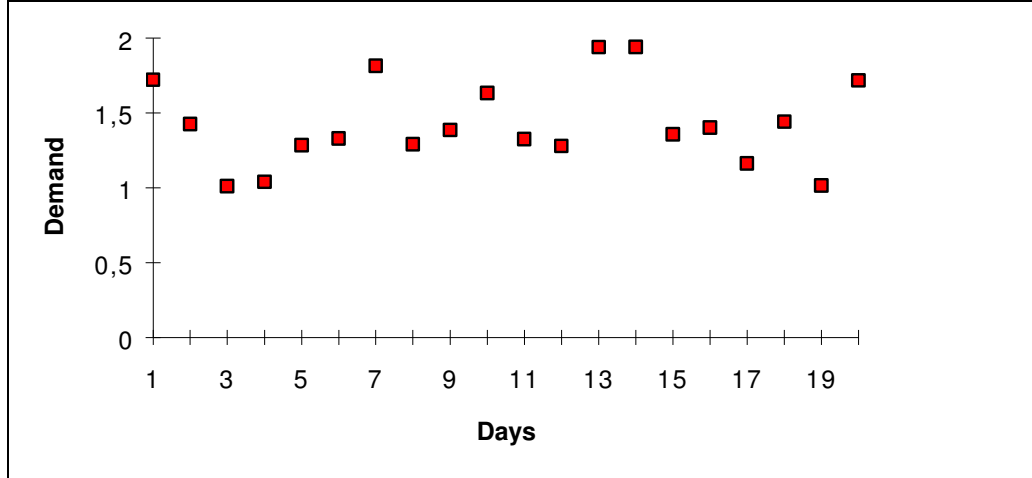
These are the notations that we will use.

### 3. Time series methods

Let us first look at some examples

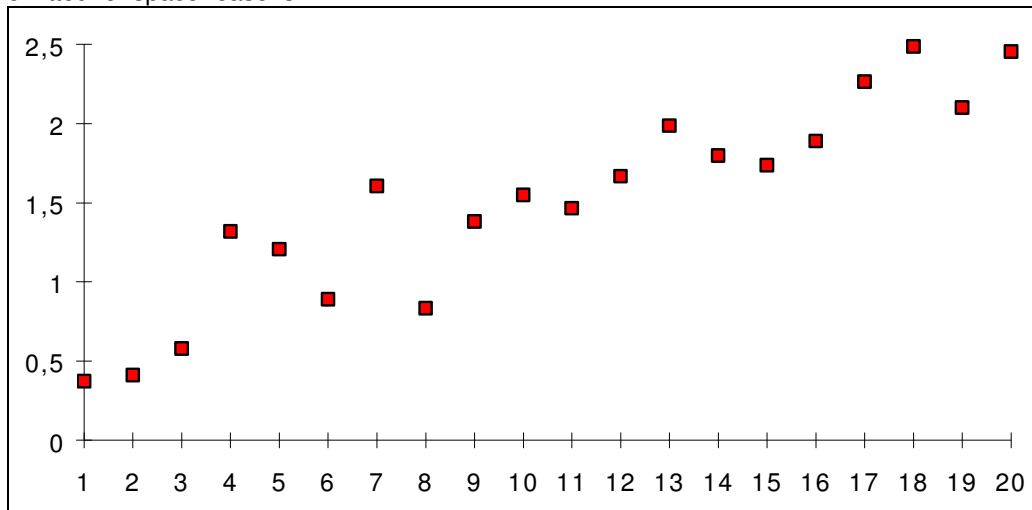
#### 3.1 Examples

Look at the demand in the past, how would you characterize these values ?



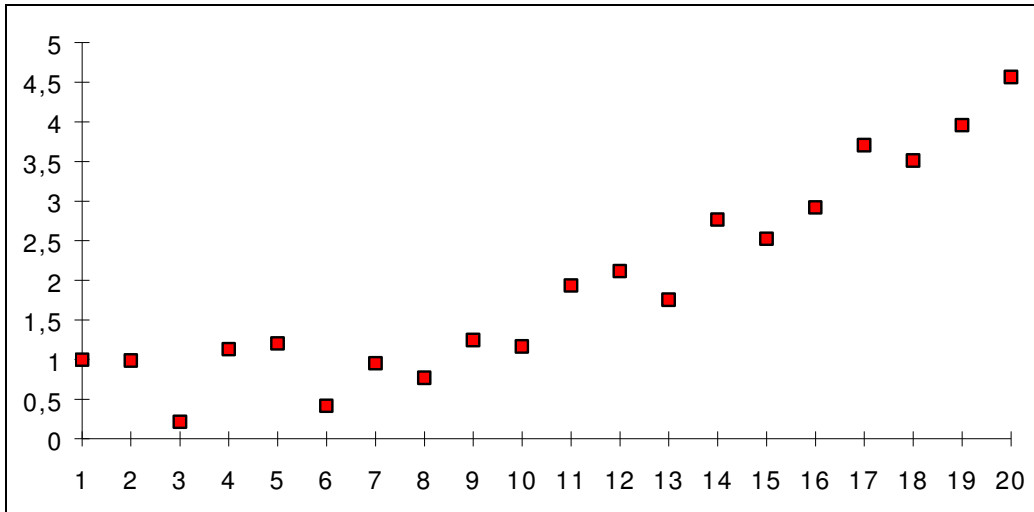
All the numbers seem distributed between 1 and 2, that is around 1,5 with some variation. We could say that the demand at time  $t$ ,  $D(t)$ , should be given by a constant (here, 1,5) plus some error term varying in  $[-0.5, +0.5]$ .

The model in this case is  $D(t) = a + \varepsilon(t)$  with  $\varepsilon(t)$  uniformly distributed in  $[-0.5, +0.5]$  and  $a$  a constant. In the chart below, as in all the next ones, the chart axis (days, demand) will be omitted for space reasons.



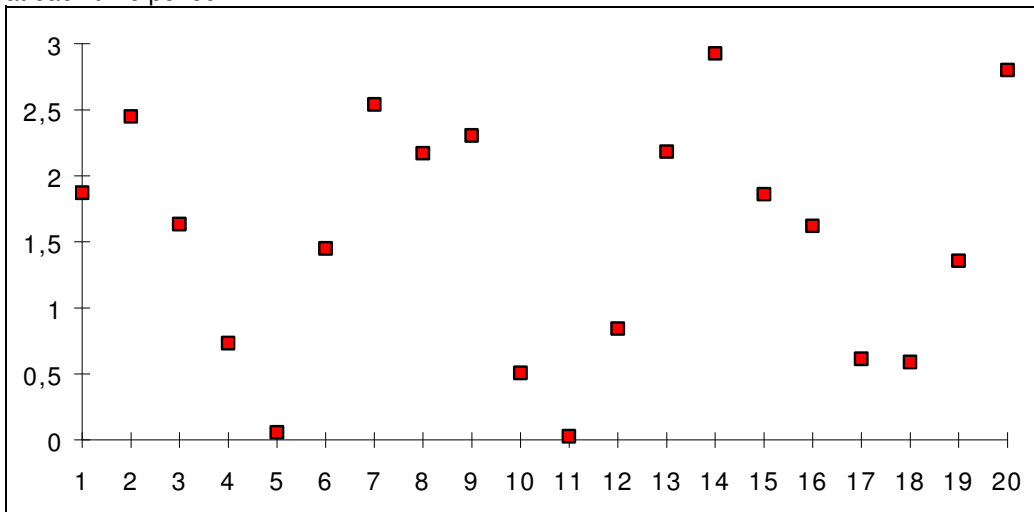
Here the demand seems to indicate a trend. It increases like a straight line with some variation around it. The model in this case is  $D(t) = a + bt + \varepsilon(t)$  where  $(a + bt)$  is the equation of the line and  $\varepsilon(t)$  is the error term.

# Examples



Here the demand seems to indicate a quadratic growth. The model in this case is  $D(t) = a + bt + ct^2 + \varepsilon(t)$  where  $\varepsilon(t)$  is the error term.

An exponential model  $D(t) = a e^{bt} + \varepsilon(t)$  could be correct too. In this case, we find the relation (neglecting the error term)  $D(t+1) = e^b D(t)$ , that is the demand is multiplied by some constant at each time period.



The seasonal model is characterized by some periodic term. For this example, the model  $D(t) = a c(t \bmod T) + \varepsilon(t)$  correspond to a stationary time series with a seasonal coefficient  $c(t \bmod T)$ . The coefficient is seasonal since  $c(t \bmod T) = c(t+T \bmod T) = c(t+2T \bmod T)$ . For example, if  $T$  is one week and  $t$  is a precise day, then there is a seasonal coefficient for each day. For example,  $c(\text{Monday})=2.0$  means that the demand on Mondays is always twice the average. The seasonal coefficients are chosen so that they sum up to the length of the period  $T$ .

## 3.2 Principles

Forecasts based on time series analysis are based on a 3-step procedure.

### 1. Select the model

**Question:**

**Is it reasonable to assume a constant, an increasing, a decreasing or a seasonal demand ?**

**Answer:**

**A mathematical model (a curve):**

$$D(t) = (a + bt + dt^2) c(t \bmod T) + \varepsilon(t)$$

The answer to this question depends of the product and of the market. It is reasonable to assume a constant demand pattern if several conditions are satisfied: old classical product, saturated market, no new competitor or competitor's move, etc.

Here below we will review the stationary, the trend-based and the seasonal models.

### 2. Parametrize the model

**Question:**

**How to find values for the parameters:  $a, b, \dots$  ?**

**Answer:**

**Use a method which tunes the parameters in order to fit the curve with the observations**

Here below different methods will be reviewed. They differ by their complexity (simple or not) and by their sensitivity to the data. We will also give some hints for the tuning of the parameters of these methods.

At this point, we have a mathematical curve which fits the past data (quite well).

### 3. Make a forecast and estimate confidence

**Forecast = model extrapolated in the future**

**Question:**

**How confident are we in this forecast ?**

**Answer:**

**If the errors made with this model in the past are small, the error we will make will be small too.**

We will describe a technique for building a confidence interval around the forecast.



## 4 Stationary time series

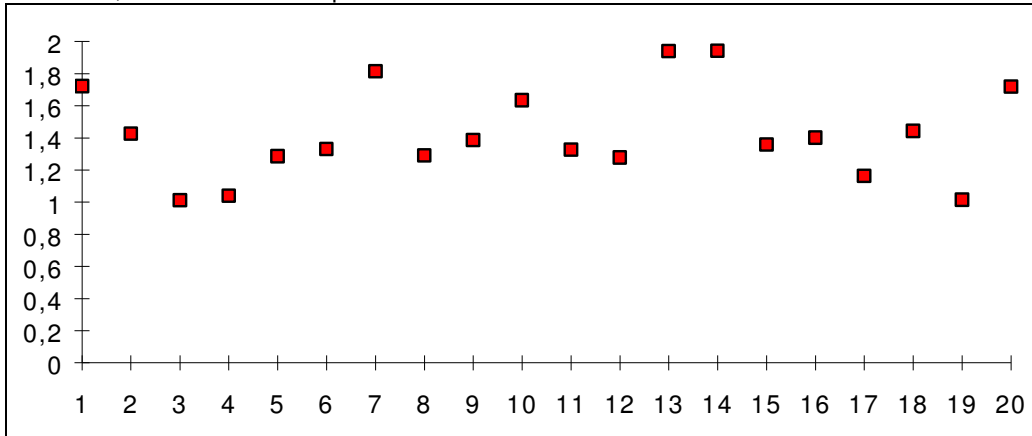
Here, we assume that the underlying model is stationary.

Assume :  $D_t = a + \varepsilon_t$

where :  $a$  is an unknown constant;

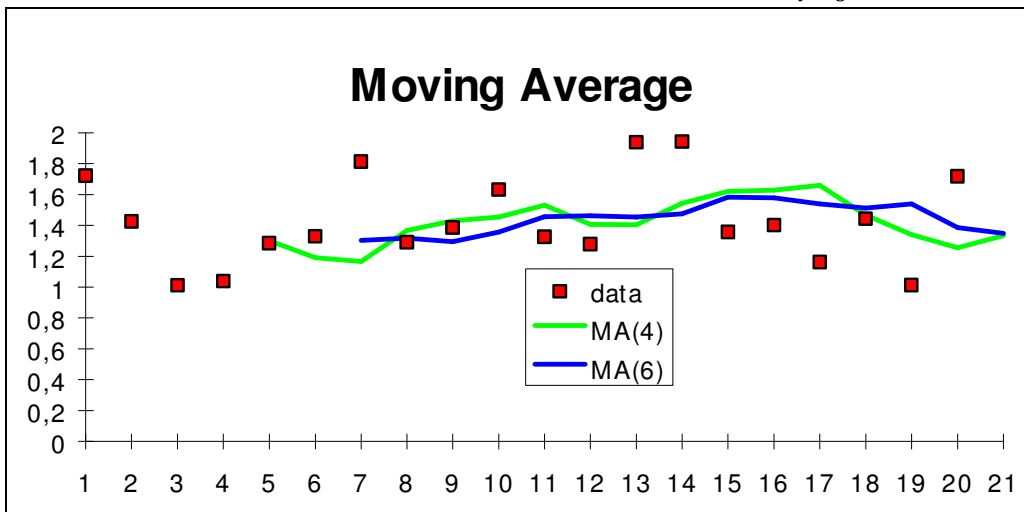
$\varepsilon_t$  is a random var:  $E[\varepsilon_t]=0$ ,  $\text{StDev}[\varepsilon_t]=\sigma$

The random variable  $D(t)$  has thus a mean equal to  $a$  and a standard deviation equal to  $\sigma$ . This is an assumption. In reality we do not know whether the demand is distributed like this. We just think here, it is constant. The problem then is to determine the constant  $a$ .



The moving average assumes that the constant is given by the average of the last  $N$  values. Indeed, this average is a random variable with mean  $a$  and standard deviation  $\sigma/\sqrt{N}$ . Larger  $N$  will thus lead to more stable (and thus also less dynamic) estimation.

4.1 Moving Average: 
$$F_t = \frac{1}{N} \sum_{i=0}^{N-1} D_{t-1-i}$$



# Stationary time series

The moving average is moving because we look each time at the last N values. All these values have the same importance. For example,  $D(t-N)$  and  $D(t-1)$  have equal importance in the computation of the moving average. We could find this inappropriate and want to give higher weights to more recent data. This leads to the weighted moving average:

## 4.2 Weighted Average: $$F_t = \sum_{i=0}^{N-1} w_{t-1-i} D_{t-1-i}$$

The weights should of course sum up to one. Here is an example.

...	w(t-4)	w(t-3)	w(t-2)	w(t-1)
0.0	0.1	0.2	0.3	0.4

Here are two other examples to introduce the exponential smoothing. Assume we look at all the past data and give them decreasing weights as in this table.

w(t-k)	...	w(t-4)	w(t-3)	w(t-2)	w(t-1)
$(1/2)^k$	...	1/16	1/8	1/4	1/2
$9 \cdot 10^{-k}$		0.0009	0.009	0.09	0.9

## 4.3 Exp. Smoothing: $$F_t = \alpha \sum_{i=0}^{\infty} (1-\alpha)^i D_{t-1-i}$$

The two above examples correspond to  $\alpha$  equal to 0.5 and to 0.9 respectively. Expressing  $F(t)$  in terms of  $F(t-1)$  leads to the following recursive expression:

$$F_t = \alpha D_{t-1} + (1-\alpha) F_{t-1} = F_{t-1} + \alpha (D_{t-1} - F_{t-1})$$

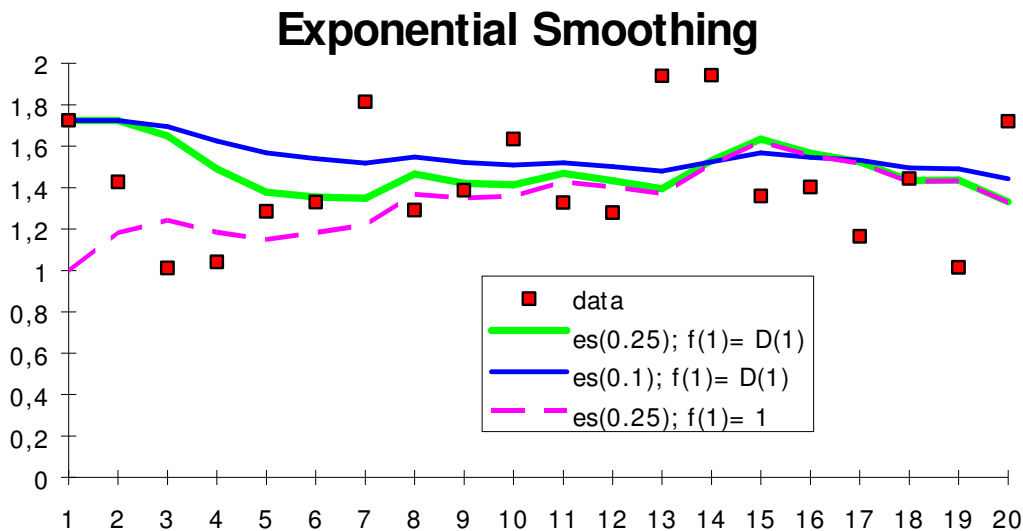
This can be seen on the table of weights too. Assume you computed at time (t-2), the forecast  $F(t-1)$  using the last row of weights. The weight was 0.9 for  $D(t-2)$ , 0.09 for  $D(t-3)$ , etc. Then, at time (t-1), you got  $D(t-1)$ , the real demand. You then want to compute your next forecast  $F(t)$ . This is similar to adding a column at the right of the table. You will multiply all the previous data by 1/10 so that, for example, the weight for  $D(t-2)$  changes from 0.9 to 0.09 and you will give the weight 0.9 to  $D(t-1)$ . This is your next forecast  $F(t)$ .

On the previous chart, you see the curves corresponding to the forecasts obtained using the exponential smoothing method with a factor  $\alpha$  equal to 0.25 and to 0.1.

It is very important to understand the equation  $F(t) = \alpha D(t-1) + (1-\alpha) F(t-1)$  as follows.

A forecast  $F(t-1)$  has been made at time (t-2). At time (t-1), some observation  $D(t-1)$  has been performed. You decide then that your next forecast  $F(t)$  will be between what you had forecast,  $F(t-1)$ , and what you observed,  $D(t-1)$ . If  $\alpha$  equals 0, then you never modify your forecast. If  $\alpha$  equals 1, then you set your forecast equal to what you just observed  $D(t-1)$ . For  $0 < \alpha < 1$ , your forecast is a compromise between  $F(t-1)$  and  $D(t-1)$ .

## 4.4 Comparison between MA and ES



### Similarities

1. assume stationary process
2. single method parameter

$N$  for the MA and  $\alpha$  for the exponential smoothing.

3. lag behind trend

If there is a general trend, both approach will follow but with some delay.

### Differences

By their definition, the two approaches do not incorporate the same amount of data.

1. number of data taken into account

Although ES takes all the data into account, its recursive formulation allows the new forecast to be determined on the basis of the last forecast and last observation only.

2. computation and memory

### Variables

for the MA: choose  $N$

A large value for  $N$  will characterize a stable forecast. This is a quality if the process is stable, it isn't if the process undergoes some changes.

for the ES: choose  $\alpha$

A small value for  $\alpha$  will characterize a stable forecast (we give very little importance to the new observation  $D(t)$ ). Again, this is appropriate if the process is stable and it isn't if the process undergoes some changes.

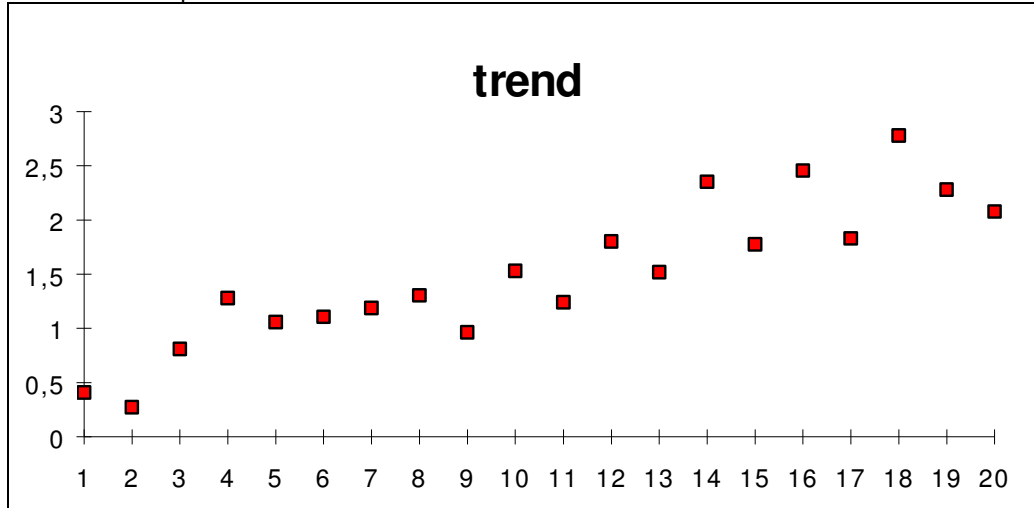
## 5 Trend-based time series

Here, we assume that the underlying model shows a linear trend.

Assume :  $D_t = a + bt + \varepsilon_t$

where :  $a, b$  are unknown;  $E[\varepsilon_t]=0$ ,  $\text{StDev}[\varepsilon_t]=\sigma$

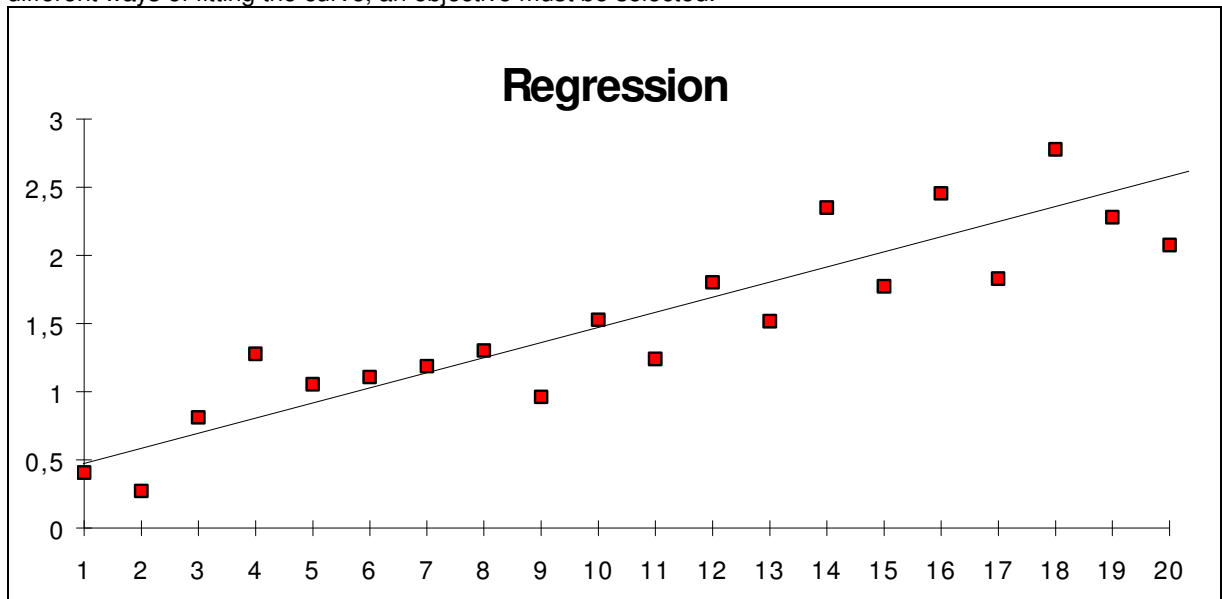
This is again an assumption. In reality we do not know whether the demand roughly follows a line or not. The problem here is to determine the constants  $a$  and  $b$ .



You see that the data roughly follows a straight line. You could ask yourself what would happen if you use the MA or the ES to estimate the forecast? (You will lag behind.)

Linear Regression :  $Y(t) = a + bt$

The regression analysis aims at fitting a straight line in the set of points. Since there are different ways of fitting the curve, an objective must be selected.



## 5.1 Linear Regression

The classical approach consists in minimizing the sum of the squared deviations.

Objective: 
$$\min_{a,b} \sum_{t=1}^n [D_t - (a + bt)]^2$$

In order to reach this objective, the model parameters  $a$  and  $b$  must be found.

$\Rightarrow a = \dots \quad b = \dots$

We follow therefore the general procedure by which we compute the derivatives with respect to the variables (here  $a$  and  $b$ ). We then set these derivatives to zero and solve the equations to find  $a$  and  $b$ . Theoretically, we should still verify that the extremum we obtain is really a minimum and not a maximum. The sign of the second derivative would tell us.

In order to solve :

$$\min_{a,b} f(a,b)$$

You must set the :

$$\frac{\partial f(a,b)}{\partial a} = 0, \quad \frac{\partial f(a,b)}{\partial b} = 0$$

Here is the equation we obtain when deriving with respect to  $a$ .

$$\begin{aligned} \sum_{t=1}^n 2[D_t - (a + bt)](-1) &= 0 \Rightarrow \sum_{t=1}^n [D_t - a - bt] = 0 \\ \Rightarrow \sum_{t=1}^n D_t - an - b \sum_{t=1}^n t &= 0 \Rightarrow a = -b \frac{n+1}{2} + \frac{1}{n} \sum_{t=1}^n D_t \end{aligned}$$

Here is the equation we obtain when deriving with respect to  $b$ .

$$\begin{aligned} \sum_{t=1}^n 2[D_t - (a + bt)](-t) &= 0 \Rightarrow \sum_{t=1}^n [-tD_t + at + bt^2] = 0 \\ \Rightarrow \sum_{t=1}^n tD_t &= a \frac{n(n+1)}{2} + b \frac{n(n+1)(2n+1)}{6} \\ \Rightarrow b &= \frac{6}{n(n+1)(2n+1)} \sum_{t=1}^n tD_t - \frac{3}{2n+1} a \end{aligned}$$

Numerically, we obtain the following equations:  $a = -10.500b + 1.501$ ;

$$b = 0.133 - 0.0732a$$

which can be solved to give  $a = 0.43$  and  $b = 0.10$ . Looking at the chart on the previous page, do you agree with these values ?

Note that similarly to the moving average method, the linear regression analysis implicitly gives the same weights to all the data.

## 5.2 Double Exp. Smoothing (Holt)

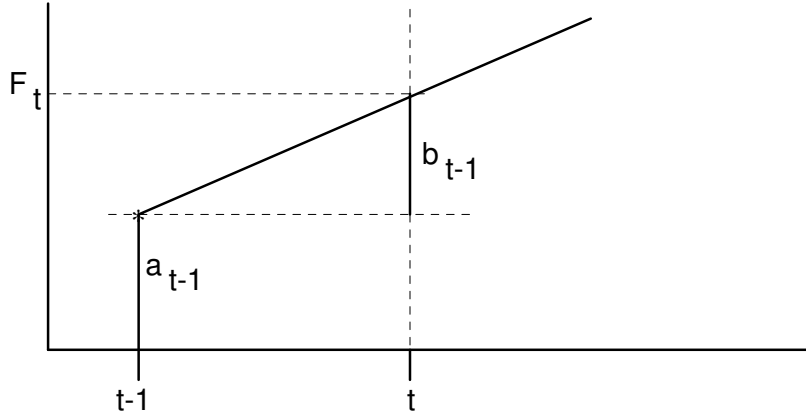
Again, by using an exponential smoothing, decreasing weights will be given to the data.

The goal here is also to determine the model parameters  $a$  and  $b$  where  $a$  is the intercept at time 0 and  $b$  the slope. Here however, we will use the parameters  $a(t-1)$  and  $b(t-1)$  where  $a(t-1)$  is the intercept at time  $(t-1)$  and  $b(t-1)$  is the estimated slope for the interval  $[t-1, t]$ . Note (but incidentally) that both model parameters are computed at time  $(t-1)$ .

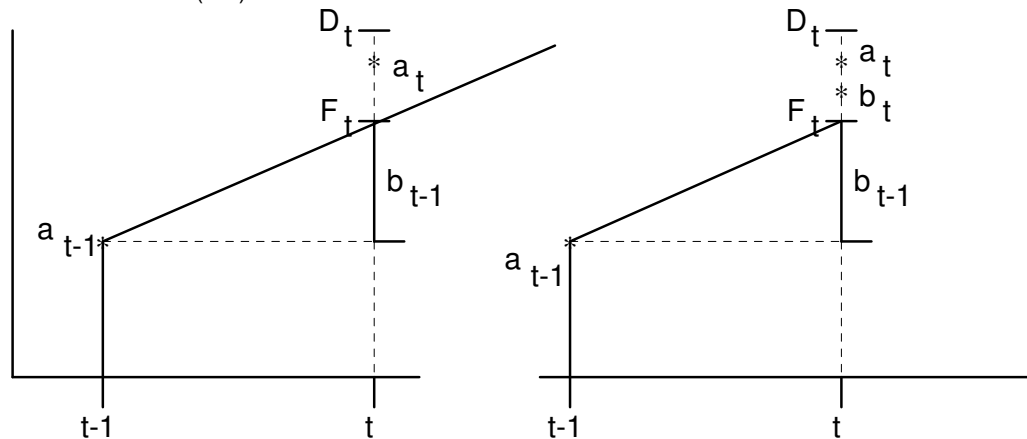
Determine :  $a_{t-1}$  and  $b_{t-1}$

Such that :  $F_t = a_{t-1} + b_{t-1}$

Since  $a(t-1)$  is the intercept at time  $(t-1)$ , the slope has to be multiplied by 1 in order to obtain the forecast for time  $t$ .



Here we show how a new forecast is derived on the basis of the previous forecast and of the observed data. On the basis of  $a(t-1)$  and of  $b(t-1)$ , the forecast  $F(t)$  has been computed at time  $(t-1)$ . Now we are at time  $t$ . We observed  $D(t)$  and we must compute  $a(t)$  and  $b(t)$  so that a new forecast  $F(t+1)$  can be determined.



We choose to set  $a(t)$  between the previous forecast  $F(t)$  and the observed data  $D(t)$ . That is,  $a(t)$  is a compromise between  $D(t)$  and  $F(t) = a(t-1) + b(t-1)$ .

For the slope, we choose a value between the previous slope,  $b(t-1)$ , and the slope of the line passing by  $a(t-1)$  and  $a(t)$ . That is,  $b(t)$  is a compromise between  $b(t-1)$  and  $(a(t) - a(t-1))$ . Both compromises are ruled by smoothing factors:  $\alpha$  and  $\beta$ .

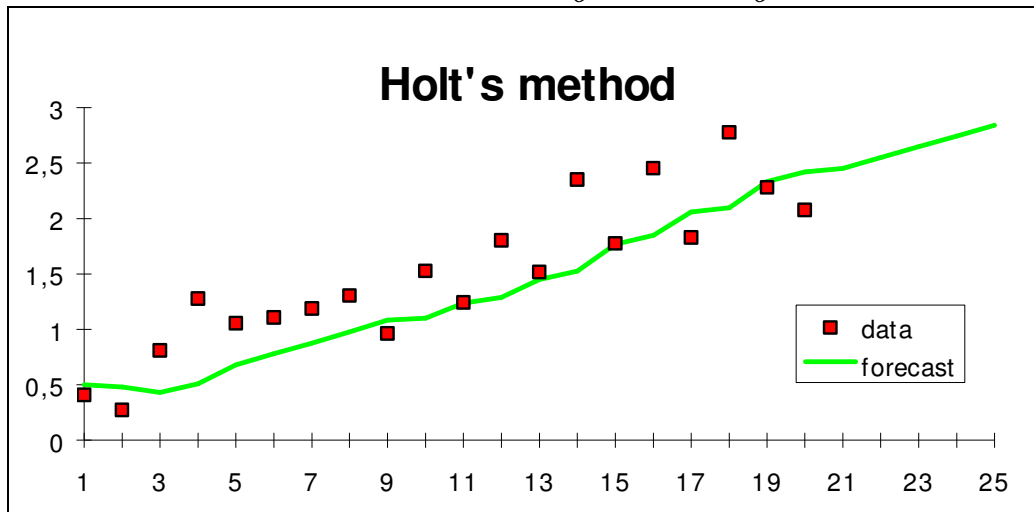
$$a_t = \alpha D_t + (1 - \alpha) (a_{t-1} + b_{t-1})$$

$$b_t = \beta (a_t - a_{t-1}) + (1 - \beta) b_{t-1}$$

# Double Exponential Smoothing (Holt)

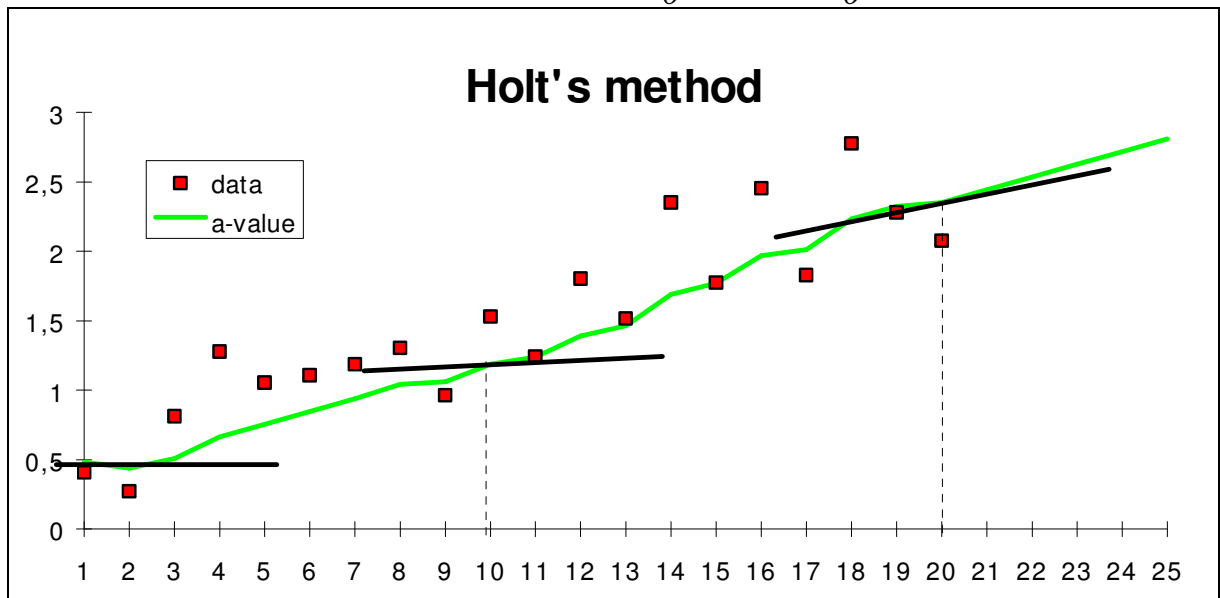
Let us look at the forecast we obtain with different smoothing factors and different initial conditions.

Forecast curve obtained with  $\alpha = 0.2$   $\beta = 0.1$   
 $a_0 = 0.5$   $b_0 = 0.0$



With these values, we obtained  $b(0)=0$ ;  $b(10)=0.050$ ;  $b(20)=0.091$  while the model used to generate the data had a slope equal to 0.1. Here below you see the values of the model parameters  $a$  and  $b$  as the forecasts evolved.

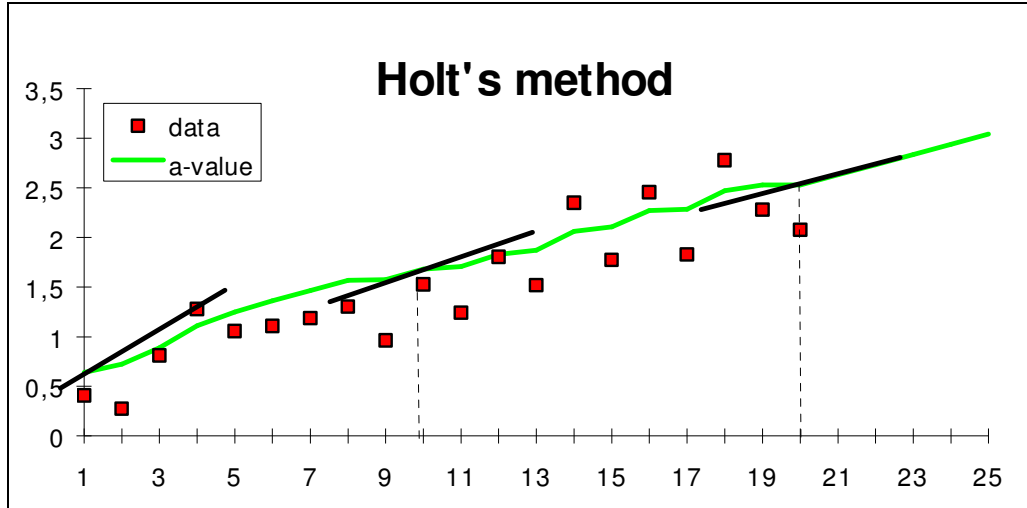
Curves for  $a$  and  $b$  obtained with  $\alpha = 0.2$   $\beta = 0.1$   
 $a_0 = 0.5$   $b_0 = 0.0$



# Double Exponential Smoothing (Holt)

Here you see the same curves with a different initialization.

Curves for  $a$  and  $b$  obtained with  $\alpha = 0.2$   $\beta = 0.1$   
 $a_0 = 0.5$   $b_0 = 0.2$



The following values were observed:  $b(0)=0.2$ ;  $b(10)=0.141$  and  $b(20)=0.102$  for a model using a slope equal to 0.1. To avoid this initialization problem, a solution consists in using a linear regression on the first data to initialize the model parameters  $a$  and  $b$ .

## Making forecasts :

Assume the forecasts are along the line

For example, with the LR, the forecast  $F(t,t+x)$  is given by the curve  $Y(t+x)$ . With the double exponential smoothing, the forecast is  $F(t,t+x) = a(t) + b(t)x$ .

## 5.3 Comparison ( Lin. Regression - 2ES)

similarity :      adequate for series with trends

differences :

- LR:
- lots of work
  - equal weight for all the data
  - (a weighted version is possible)

Double ES (Holt):

- easy to compute
- decreasing (possibly dynamic) weights
- difficult to initialize



## 6. Seasonal Time Series

Here, we assume that the underlying model is seasonal.

Assume :  $D_t = (a + bt) c_{t \bmod T} + \varepsilon_t$

where :  $a, b$  and  $c_0, \dots, c_{T-1}$  are unknown;

$$c_0 + \dots + c_{T-1} = T$$

$$E[\varepsilon_t] = 0, \text{ StDev}[\varepsilon_t] = \sigma$$

Note first, that the previous model can be derived from this more general one. By choosing  $T=1$  and  $c_0 = 1$ , we get the trend-based series  $(a + bt) + \varepsilon(t)$ . By further setting  $b=0$ , we obtain the stationary model:  $a + \varepsilon(t)$ .

The seasonal model is characterized by the fact that each season has its own seasonal coefficient. For example, if  $T$  is one week and  $t$  is a precise day, then there is a seasonal coefficient for each day. For example,  $c(\text{Monday})=1.2$  means that the demand on Mondays is always 20 percent larger than the forecast  $(a+bt)$  determined without seasonal variation. Here is an example of a set of seasonal coefficients for a whole week ( $T=6$ ).

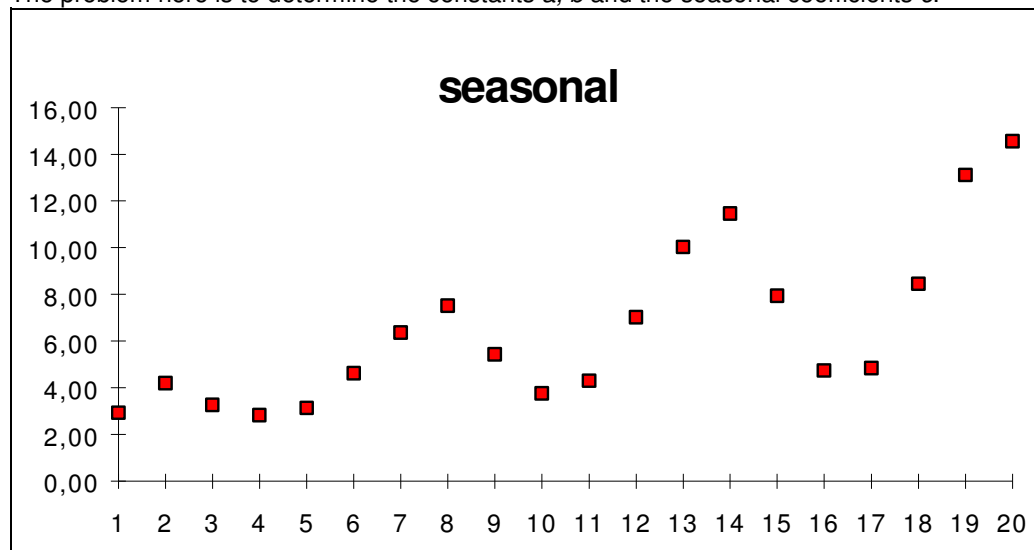
Example:

Mon.	Tues.	Wed.	Thu.	Fri.	Sat.
$c_0$	$c_1$	$c_2$	$c_3$	$c_4$	$c_5$
1.2	1.0	0.6	0.9	1.3	1.0

The seasonal coefficients are chosen so that they sum up to the length of the period  $T$ . By doing so, the values without seasonal coefficients,  $a + bt$ , correspond to the average. In the above example, Monday and Friday are 20 and 30 percent above the average (also called the deseasonalized values) and Wednesday and Thursday are 40 and 10 percent below this average.

Note that we assume a multiplicative model, that is, the seasonal coefficient multiplies the normal forecast  $(a+bt)$ . In an additive model, the seasonal coefficient would just be added to the normal forecast, that is,  $D(t) = (a+bt) + c(t \bmod T)$ . In an additive model, the seasonal coefficients should sum up to zero.

The problem here is to determine the constants  $a, b$  and the seasonal coefficients  $c$ .



## 6.1 CMA method

Two different methods will be described. The first method is called the CMA method. CMA stands for centered moving average. In reality this method provides the seasonal coefficients only. Afterwards, the trends (the parameters  $a$  and  $b$ ) should be determined by one of the previous methods (linear regression or double exponential smoothing). The second method is called triple exponential smoothing or method of Winters.

Here are the successive steps of the CMA method.

### CMA Method: 5 steps

1. examine data and determine cycle length (T)
2. determine the seasonal coefficients
3. extract seasonal coefficients from data (deseasonalize)
4. determine the trends (by another method)
5. make (deseasonalized) forecasts
6. introduce seasonal coefficients

First the cycle length must be determined. This is the model parameter  $T$ . Note that a cycle has, by definition of a seasonal model, a fixed length. Most often this cycle results from the nature of the product which is sold. On a chart, the only way to recognize a cycle consists in checking whether the data follows the same behavior. Anyway, if you choose a wrong value for the cycle length, you will realize it in step 2.

#### 1. Determine cycle length: $T = 6$

This means that  $T$  different seasonal coefficients are needed:  $c_0, \dots, c_{T-1}$ . The step 2 aims at evaluating these coefficients.

#### 2. determine the seasonal coefficients

Now, let us organize the data in a table whose length is the cycle length.

Periods (t)				Values ( $D_t$ )			
1	7	13	19	2.93	6.36	10.03	13.11
2	8	14	20	4.19	7.51	11.46	14.56
3	9	15		3.27	5.43	7.93	
4	10	16		2.82	3.76	4.73	
5	11	17		3.13	4.30	4.84	
6	12	18		4.62	7.02	8.45	

Compute then the centered moving average (CMA) for each time  $t$ .

$$CMA(6)_t = \frac{1}{6} \left( \frac{D_{t-3}}{2} + D_{t-2} + D_{t-1} + D_t + D_{t+1} + D_{t+2} + \frac{D_{t+3}}{2} \right)$$

The goal is here double. By summing over a set of values, we first hope to get rid of the errors  $\varepsilon(t)$ . This principle was already used in the MA method. Second, by summing over exactly a cycle (that is,  $T$  values), we sum approximately equal  $(a+bt)$  values with each a different seasonal coefficient. But this sum is by assumption equal to a constant ( $T$ ). Dividing by  $T$  leads thus to the average value at time  $t$ ; that is,  $(a + b t)$ .

# Seasonal times series : CMA

Here is the kind of expression we used.

$$CMA(T)_t \approx \frac{1}{T} \sum_{cycle} (a + bt) c_{t \bmod T} + \varepsilon_t \approx (a + bt)$$

The CMA values can thus be seen as rough estimates of the real demand without seasonal variation. Note that these moving averages must be centered to reduce bias.

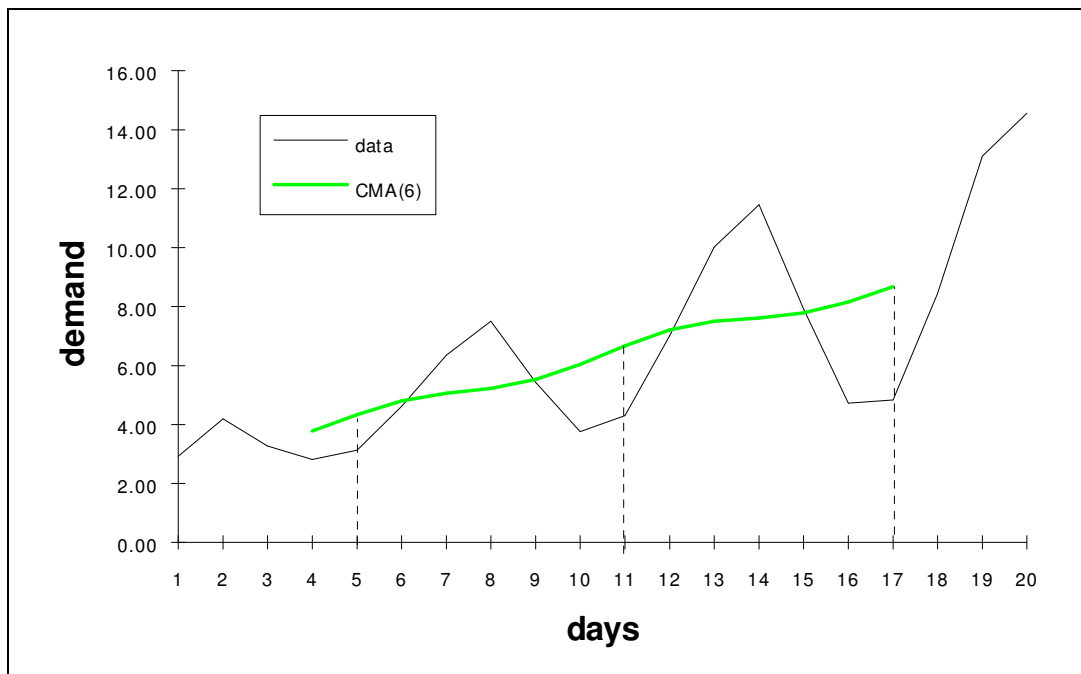
Practically, the centered mean average for the first Thursday is obtained by adding to the demand of this day, the demand of the previous and next days up to a cycle length.

**Values  $D_t$**

2.93	6.36	10.03	13.11
4.19	7.51	11.46	14.56
3.27	5.43	7.93	
2.82	3.76	4.73	
3.13	4.30	4.84	
4.62	7.02	8.45	

**$CMA(6)_t$**

*	5.06	7.50	*
*	5.23	7.62	*
*	5.53	7.79	
3.78	6.04	8.16	
4.34	6.67	8.68	
4.80	7.21	*	



Since,  $D_t \approx (a + b t) * c_t$  and  $CMA_t \approx (a + b t)$ , the ratio between the observation and CMA gives therefore an idea of the seasonal coefficient  $c_t$ . In fact, we can compute this ratio for each period. In our example, we will obtain values for  $c_4, c_{10}, c_{16}$ . However, all these values correspond to the season number 4 (Thursday, for example). The question is then, which value should we choose for the seasonal coefficient corresponding to Thursday. The answer is: the normalized average. We take the average over all the values we have:  $1/3 (c_4 + c_{10} + c_{16})$ . These are the Avg(t) values below.

## Seasonal times series : CMA

---

$D_t/CMA(6)_t$			$\Rightarrow$	$Avg_t$	$\Rightarrow$	$c_t$	
	1.26	1.34		1.30		1.29	<b>c<sub>1</sub></b>
	1.44	1.50		1.47		1.46	<b>c<sub>2</sub></b>
	0.98	1.02		1.00		0.99	<b>c<sub>3</sub></b>
0.75	0.62	0.58		0.65		0.65	<b>c<sub>4</sub></b>
0.72	0.64	0.56		0.64		0.64	<b>c<sub>5</sub></b>
0.96	0.97			0.97		0.97	<b>c<sub>0</sub></b>
				<b>6.03</b>		6.00	

We still normalize these average (here to T=6) since we want the seasonal coefficient to sum up to the cycle length. To understand the need for this last normalization, assume you have a cycle of length 2 and you found (avg(0)=1.8 and avg(1)=0.6).

Here are the formulas you use to compute these seasonal coefficients.

$$Avg_t = \frac{1}{k} \sum_{i=0}^{k-1} D_{t+i*T} / CMA(T)_{t+i*T}$$

$$c_t = Avg_t * \left( T / \sum_{i=0}^{T-1} Avg_i \right)$$

Now we have estimated all the seasonal coefficients. The next step consists in extracting these coefficients from the observed data in order to get deseasonalized observations. Each observation corresponds to a season and is thus divided by the corresponding seasonal coefficient.

# Seasonal times series : CMA

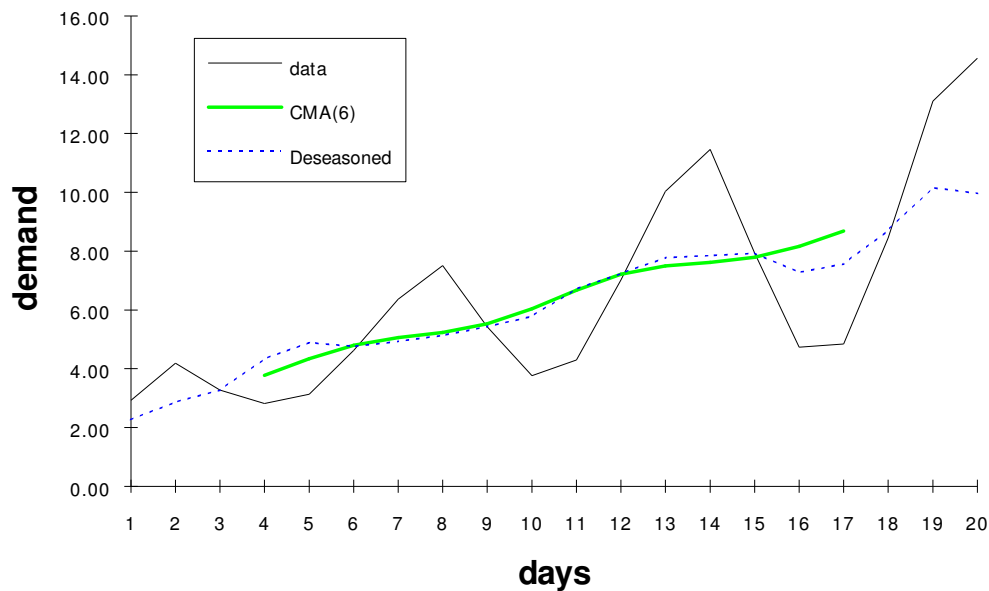
Now we proceed with the final steps of the method.

## 3. Deseasonalize the data

Since we think that  $D_t \approx (a + b t) * c_t$ , by dividing by  $c_t$ , we should obtain the term  $(a + b t)$ , that is, a line of deseasonalized observations. Here are the values we obtain.

**Deseasonalized Values  $D_t/c_t$**

2.27	4.93	7.78	10.16
2.87	5.14	7.85	9.97
3.27	5.43	7.93	
4.34	5.78	7.28	
4.89	6.72	7.56	
4.76	7.24	8.71	



## 4. Compute the trends: $a$ and $b$

The trend can be computed by a linear regression or by a double exponential smoothing.

## 5. Make forecast: $DF_{21} = 10.4$

We determine then a new deseasonalized forecast  $(a+bt)$  and then, the seasonal coefficient is introduced.

## 6. introduce seasonal coefficient:

$$F_{21} = 10.4 * c_3 = 10.3$$

## 6.2 Winters' Method

The method here consists in a triple exponential smoothing by which all the model parameters ( $a$ ,  $b$  and the  $c$ 's) are updated when a new observation is obtained. We also assume that the cycle length has been previously determined. The model parameters " $a$ " and " $b$ " are updated as with the double exponential smoothing, except that the observation is first deseasonalized. The third equation updates the corresponding seasonal coefficient. There again, a compromise is made between the previous value of the seasonal coefficient and the observed coefficient.

$$1) a_t = \alpha(D_t / c_{t \bmod T}) + (1 - \alpha)(a_{t-1} + b_{t-1})$$

as in Holt's method, except  $D_t \rightarrow (D_t / c_{t \bmod T})$

$$2) b_t = \beta(a_t - a_{t-1}) + (1 - \beta) b_{t-1}$$

as in Holt's method

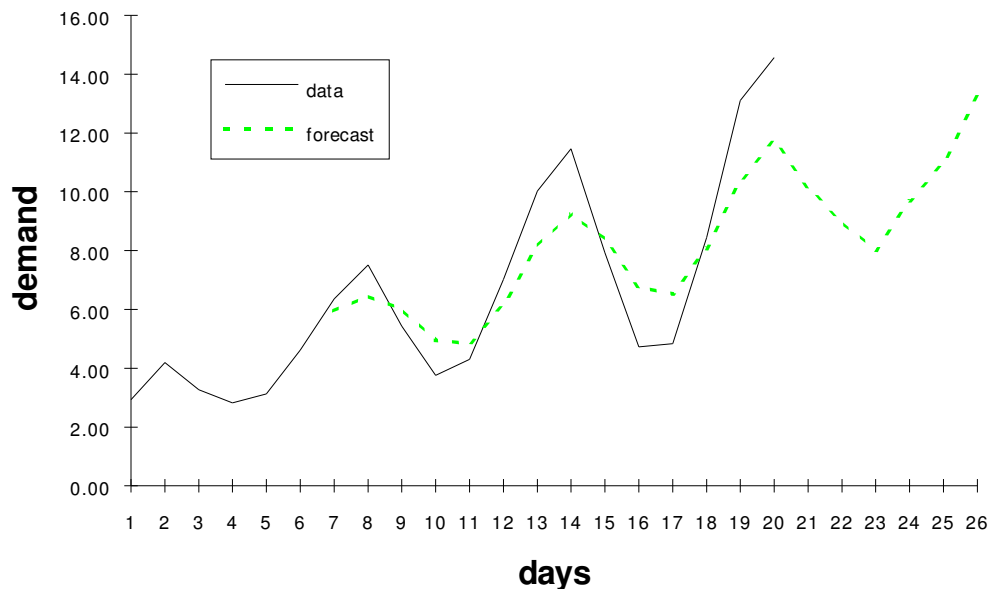
$$3) c_{t \bmod T} = \gamma(D_t / a_t) + (1 - \gamma)c_{t \bmod T}$$

The forecast is determined as usually.

$$\text{Forecast : } F_{t,t+\tau} = (a_t + \tau b_t) c_{t+\tau \bmod T}$$

First, the trends is used to determine the deseasonalized forecast ( $a + b t$ ). Then, the adequate seasonal coefficient is introduced.

Here below is a plot of the forecast made daily using the Winter's method. We can observe that the trend is adequate but that the seasonal coefficients are not yet adequate.



### Comparison: CMA - 3ES

The comments on the comparison LR - 2ES apply here similarly and are left as an exercise.

## 7. Evaluation of Forecast

Here we try to determine whether the forecast is good, that is accurate and not biased. There are two usual ways of measuring the accuracy of the forecasts.

### ⇒ Mean Absolute Deviation

Here we compute the average absolute deviation.

$$MAD = \frac{1}{n} \sum_{t=1}^n |e_t|$$

The following result is extremely useful since it allows to determine the variation of the error process from the measure of the MAD.

if  $e_t$  is  $N(0, \sigma)$ , then  $\sigma \approx 1.25 \text{ MAD}$

For example, if we measure an MAD value of 8 units, we could derive that the error process has a standard deviation  $\sigma$  equal to 10 units. The probability that the demand exceeds the forecast by 20 units is then 0.023.

### Example

$$F_t = 120 \quad \text{units}$$

$$MAD = 8 \quad \text{units}$$

$$\Rightarrow \sigma = 10 \quad \text{units}$$

We can thus argue that the forecast is correct (equal to 120) up to an error term which is distributed as a normal with mean 0 and standard deviation  $\sigma$ . We can then derive the probability that the demand falls within some confidence interval.

$$\text{Prob}[F_t - 2 \sigma < D_t < F_t + 2 \sigma] = 0.95$$

$$\Rightarrow 95 \text{ percent confidence interval} = [100 - 140]$$

### Valid method if the errors and the previsions are made in the same way

This means that if the errors are measured by comparing the real demand and the forecasts made two days before, then we can build a confidence interval for a forecast made today for the demand in two days.

Note that the MAD is sometimes computed by using an exponential smoothing process.

### ⇒ Tracking Signal

If your forecasting method is correct you should overestimate and underestimate the demand rather regularly. If your method is biased, then you will repetitively under (or over-) estimate the demand. The tracking signal aims at checking whether there is some bias or not.

$$TSE = \sum_t e_t / MAD$$

It simply consists in summing the error over time. Theoretically, if there is no bias, this sum should remain close to zero. The division by the MAD aims at measuring the distance from the mean (here, 0) in terms of MAD. We can plot the tracking signal over time for checking purpose. We could also decide the limits it should not exceed. Typical values are a few standard deviations.

## 8. Other Methods

### Causal Relationship:

#### Multiple Regression Analysis

The idea is that your forecast could be determined by several variables (x, y and z). If we know them in advance we could make some forecasts. The coefficients (X, Y and Z) are determined by a regression on the past data.

$$F_t = \mu + Xx_t + Yy_t + Zz_t$$

#### Focus Forecasting (B. Smith)

The principle here is very intuitive. Ask the salesman how he derives his forecast (a rule). Check whether it would have worked on the past data. If it would, then use it for the subsequent forecasts.

⇒ Try a rule on the past data.

If it worked well, then it will still work ...

## 9. Which method to use ?

depends on: • observation

If you have few observations, it will be difficult to define a detailed model. A linear regression for computing the trends requires at least 10 observations to be valid. If you have some seasonal variation, 3 or 4 cycles of observations are necessary for your model to start being effective.

depends on: • time horizon

The time horizon is also important. Do you want to make a forecast for tomorrow, next month, next year or the next 5 years ? How does the demand process vary with respect to your time horizon ? Linear regressions (and causal relationship methods) seem for example more robust for long term forecast. In any case, the observations on which you base your forecast is a direct indicator of the horizon range for which your forecast is valid.

depends on: • the MAD observed

If sufficient data are available, you could use the 3 fourths of your observations to set the parameters of your model and then, use the last fourth of observations to compare the forecasts of your model with what really happened. Comparing the errors (the MAD) made by different methods (and/or by different models) provides an immediate selection criterion.

depends on: • complexity

Finally, the amount of energy to spend for the forecasting process plays also a role too.

But do not forget the characteristics of forecasts (page 2)