EPFL: Course ME-419

Production Management

Chap 3, Demand Management, Forecasting

The main purpose of a company is to serve its market and fulfil the demands of its customers. Therefore, knowing this demand is absolutely essential for running the company. Depending on the type of business, the demand must be known, or at least estimated, a long time before it effectively materialises into firm orders. This chapter presents the methodologies and tools for establishing aggregated demand on the basis of forecasts. In a company, these data serve as a basis for decisions made on the strategic, tactical as well as operational levels. Their importance is therefore tremendous, but unfortunately and as everyone knows, it is extremely difficult to elaborate reliable forecasts, particularly on a long term horizon. This dilemma is partly covered in this chapter and will be discussed further in chap. 4.

3.1 Introduction

This first section presents the importance of demand planning and forecasting for operations management in companies and its influence on the optimisation of the supply chain. It describes the connections between demand management and the other functions in the company like supply management, production planning, etc., and the important relationship between lead times and demand forecasting.

3.1.1 Types of forecast

Demand forecasts are required as inputs for decisions related to various horizons (short, medium, long term). Furthermore, forecasts can be established at different aggregation levels, for example:

- For each single product (detail level);
- For each product type;
- For a larger product group such as a product family or
- For the total turnover of the company.

Both the *forecast horizon* and the *forecast aggregation level* depend on the use of forecasted data. Generally, the following situations occur, as illustrated in the next fig.:

- Short term forecasts at a detail level are used for short term planning and execution;
- Medium term forecasts at an average aggregation level are used for medium term planning (general planning);
- Long-term forecasts at a high aggregation level are used for long term planning (strategic planning).

Forecasts play an important role in:

Short-term operations management:

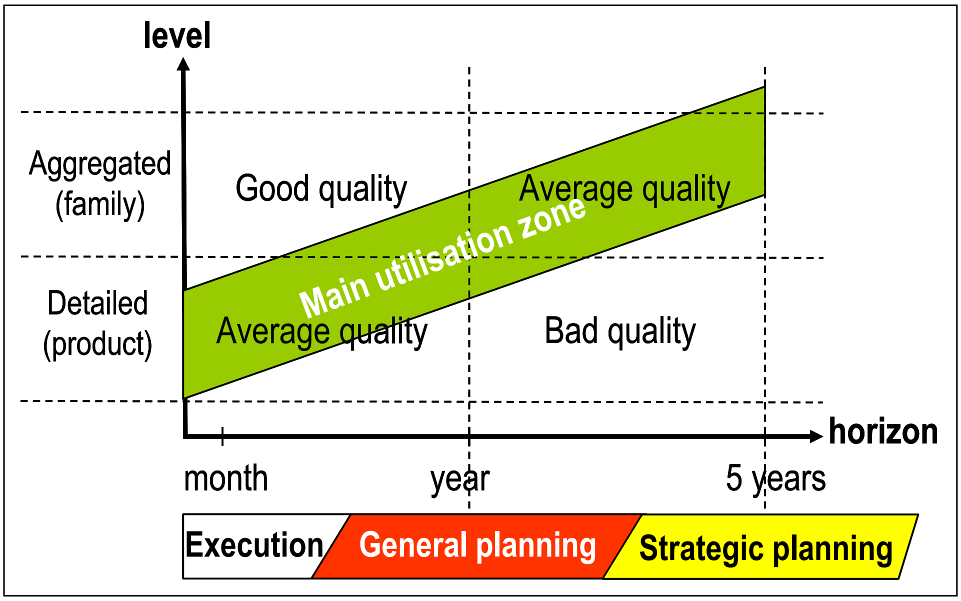
Demand forecasts are essential inputs for operations management since they serve as a basis for assuring the availability of resources (humans and machines) and therefore the achievement of high *service levels* at minimum cost.

Medium term resource acquisition:

Demand forecasts are also useful during medium term planning which involves acquisition of resources and modification of the workforce. Depending on the industry, these types of planning decisions can be made over periods such as a few days up to several months.

Long-term strategy of the company:

Each organisation must make decisions regarding its strategic policies over the long term. These decisions involve technical choices, financial and property investments, marketing actions, development of new products, etc... In general, long term forecasts help managers improve decisions by reducing the risk associated with future demand.

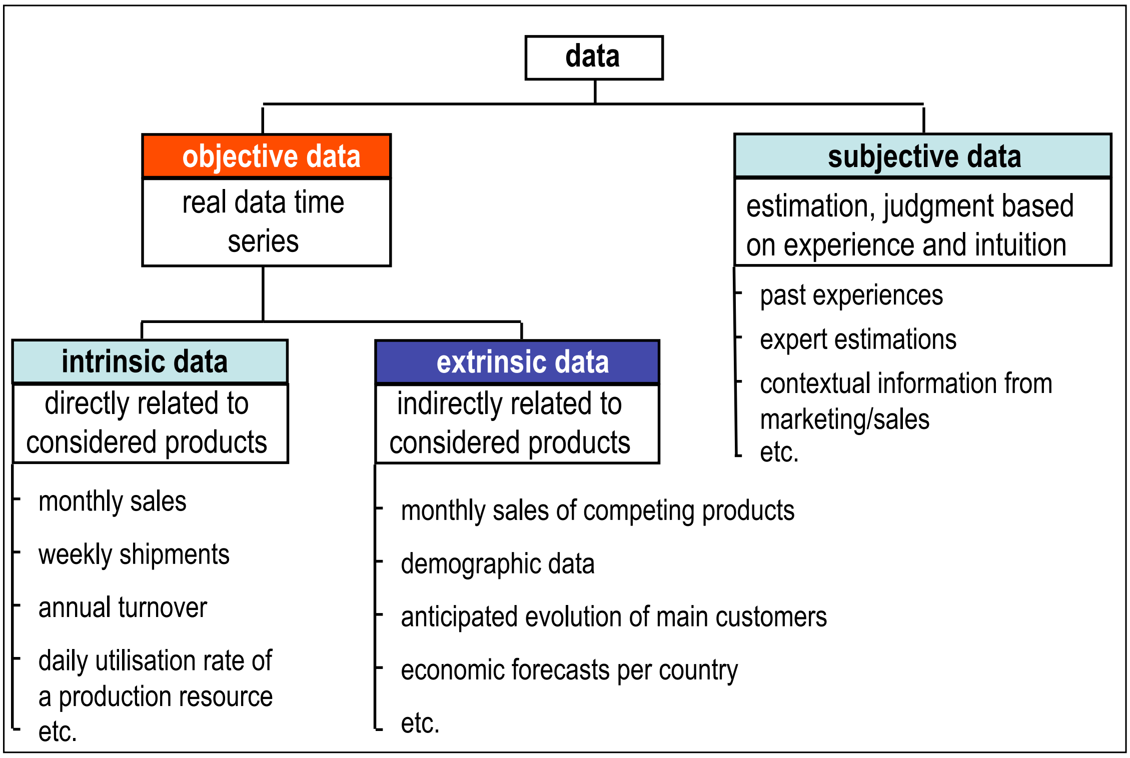


Relation between *forecast horizon* and *forecast aggregation level*

Therefore, the most useful working zone goes from the lower left to the upper right corner of the previous figure. It can be seen that this corresponds to situations where forecasts of medium quality should be expected.

3.1.2 Types of data

Raw data are needed to establish forecasts. These data can come from various sources and can be of different natures as illustrated in the fig. below.



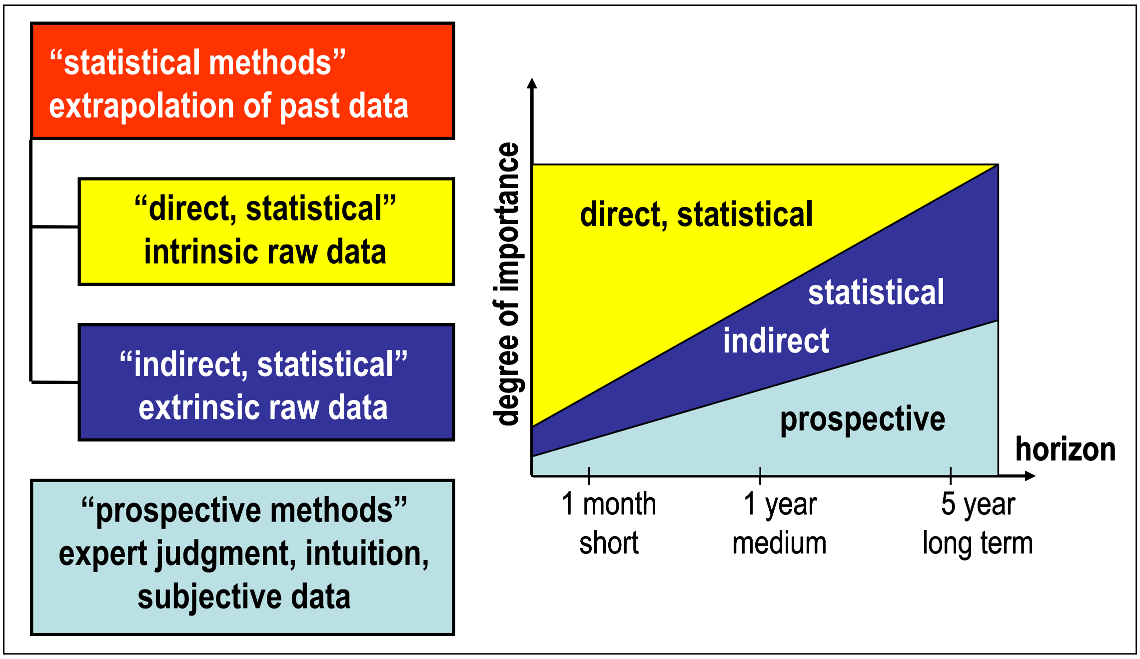
Types of data used in forecasting

Firstly, the data can be a time series extracted from past databases or they can result from personal experience, prospective estimation or intuition, i.e. human judgment.

Secondly, the time series can come from data directly related to the products concerned by the forecast (for example the monthly sales of the considered product family) or from data indirectly related to them (for example demographic data of a market) as shown in the previous figure.

3.1.3 Main categories of forecasting methods

Three main categories of forecasting methods can be defined depending on the nature of the raw data used. If intrinsic time series are used, one would speak of direct, statistical forecasting methods, while indirect statistical methods relate to forecasting based on extrinsic time series. Prospective methods cover the forecasting approaches using human judgment (personal experience, expert judgment, contextual information, etc.). The relative importance of these three broad categories of forecasting methods is illustrated in the next fig. as a function of the forecast time horizon.



Main categories of forecasting methods

3.2 Role and Importance of Demand Management

Demand represents the need for a particular product or component (in the marketplace). Demand management encompasses all activities required to link the market to a company. This function ensures that all sources of demand will be taken into account in production management activities such as in *master production scheduling*. Consequently, demand management consists of demand forecasting, order entries, quoted delivery times, customer order services, after market demand, etc.

Demand management is strongly connected externally to the marketplace or the source of the demand. Internally, it is closely tied with the *manufacturing planning and control (MPC)* system. If the value adding activities of the company cannot be realised within the quoted delivery *lead-time*, then there is a need to forecast the demand at the appropriate level of detail, prior to other *MPC* activities. In other words, if the production *lead-time* is longer than the promised delivery *lead-time*, then the enterprise needs to rely on forecasts to produce in advance, ahead of actual demand. This primarily concerns the most upstream part of the *value adding chain* that becomes therefore quite dependent on forecasts.

In a *Make to Stock (MTS)* situation all value adding activities are run according to forecasts and detailed forecasts are required for production management.

In contrast, in a *Make to Order (MTO)* environment forecasts are less essential for managing production, although they may be necessary for purchasing. Long-term forecasts are nevertheless necessary to define long-term strategy.

In an intermediate situation such as *Assemble to Order (ATO),* detailed forecasts are required to manage the upstream part of the *value adding chain* (in particular procurement, part manufacturing and production of sub-assemblies), while the downstream finish product assembly is run on the basis of customer orders. Long-term forecasts are again necessary for long-term strategy.

The following sections describe the Demand Management function in three companies having different production organisations: *Make to Order (MTO), Assemble to Order (ATO)* and *Make to Stock (MTS).*

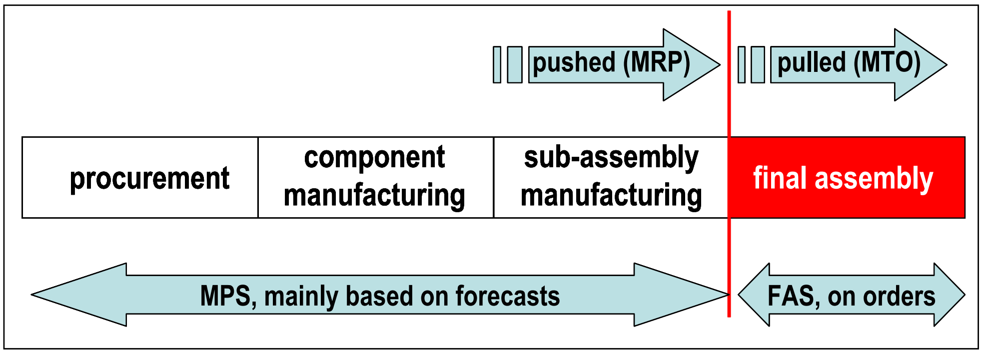
3.2.1 Demand management in a Make to Order company

Nigolava is a company specialised in the design and production of installations for amusement parks, such as roller coasters. Their products are designed and built according to customer specifications and must satisfy high security constraints. Consequently, products are mostly engineered using state-of-the-art manufacturing techniques and high performance materials. Logically, all their products are different and require a year or more to produce. A large part of this lead-time consists of order processing, design engineering and procurement.

The approach Nigolava has chosen to manage customer orders and to schedule production is based on the creation of an imaginary *Bill of Material* for the major *sub-assemblies* of the *finished product*. This *BoM* is established on the basis of the company's experience with similar projects. It includes all *lead-times* required to produce the different *sub-assemblies* (sum of the engineering design, procurement and manufacturing *lead-times*). Once it is established, it is processed using standard *MRP* logic. This procedure is used in mid-term planning that helps to assure that all *components*, raw materials and *sub-assemblies* are available on time. It also supports the establishment of a mid-term capacity requirement profile for each work centre. Nigolava also creates long-term forecasts related to the number of fairs they expect to win and the size of the projects. These long-term forecasts help them plan their resource requirements (machine acquisition, hiring of research and development engineers, etc.) for the following few years.

3.2.2 Demand management in an Assemble to Order company

Stabiviol produces packaging machines. Stabiviol has developed five different families of packaging machines but their wide range of option offers creates a large number of *finished product* combinations. Since the machines produced are quite large and expensive, it is unrealistic to keep an inventory of finished products. Production is articulated around two distinct management modes. The replenishment of raw materials and components, as well as the production of the common sub-assemblies follows a push production system, which is consistent with MRP logic. The final assembly and customisation processes are pulled by the customer, which is referred to as *Make to Order (MTO).* This *Assemble to Order (ATO)* organisation is illustrated in the next fig.



Demand management scheme in an *Assemble to Order (ATO)* environment

This *Assemble to Order (ATO)* environment requires both a *Final Assembly Schedule (FAS)* and a *Master Production Schedule (MPS).* Rough capacity requirements for the final assembly and for component and sub-assembly manufacturing are established during the long-term planning process while respecting maximum capacity. The replenishment of raw materials and the manufacturing of components and sub-assemblies are controlled by dependent needs and are calculated using *MRP* logic, which is therefore directly dependant on the mid-term planning data included in the *MPS*. Both mid-term and long-term planning for product families are based on forecasts since the production *lead-times* of *components* and *sub-assemblies* are greater than the promised delivery *lead-time* to the customer.

3.2.3 Demand management in Make-to-Stock company

Ginivik produces 30 different types of lawnmowers categorized in 3 families according to their size and type of use. Basically, the products are differentiated by their type of body, blade and motorisation. Ginivik also offers different mulching kit options that can be added to the basic models.

Demand is highly seasonal but Ginivik has stabilised the workload in the manufacturing plant by permitting fluctuations in the *finished products inventory* in agreement with the distribution centres. The competitive market requires Ginivik to guarantee high quality products with short delivery *lead-times*. Ginivik's production follows a *Make to Stock* approach, and *just-in-time* policies are used to produce common *sub-assemblies*. Customer orders are fulfilled from the *finish product inventory* held by the company's distribution division. The *Master Production Schedule* is created from forecasts of *finished product families* but a specific model is used to plan the final assembly operations. Instead of establishing individual product forecasts, the model evaluates percentages from the aggregated forecasts for *product families*. This technique is appropriate since over the previous years, Ginivik has observed that its product mix ratio remains constant over time.

3.3 Significance of Forecasting in Supply Chain Management

For a large majority of companies, the delivery *lead times* of *finished products* are greater than what the market will accept, thus making it impossible to completely adopt a *Make to Order* production policy. Consequently, demand forecasting represents a key function for the survival and success of the company. In fact, demand forecasts condition the optimisation of the *value adding chain*, whatever the organisation type and the level of flexibility and reactivity. The knowledge of future demand is extremely useful information, playing an important role for many decisions made at various levels in the *supply chain*.

In the purchasing department, two kinds of information are essential to define the optimal procurement policy: the estimated future needs in *components* and raw materials and the reliability of these forecasts. This information is important in determining the minimal coverage of needs for a specific horizon and the level of the *safety stock*.

During the process of production planning, demand forecasts play an important role in making key decisions such as make or buy, in creating the *Master Production Schedule*, which drives production, or in determining the required resources.

For distribution planning, forecasts are again important to dimension inventories and transportation capacities.

Although this example is non exhaustive, it reveals the importance of forecasts in the global optimisation of the *value adding chain*.

3.3.1 Forecast horizon

Forecasts are essential for making decisions at various management levels. In order to attain its goals, a company tries to anticipate contextual and environmental factors and to integrate them into its forecasts. The more a company wants to reduce uncertainties the more it must put effort into improving forecasts. Furthermore, due to the interconnectivity of departments and sectors in modern organisations, the quality of forecasts will most likely benefit the entire organisation. Forecasts play an important role in the following functions:

Short term operations management:

Demand forecasts are essential inputs for operations management since they serve as anticipation for resource requirement planning, both human and material. They are essential for assuring the availability of resources (human and material) and therefore, the achievement of high service levels at minimum cost.

Medium term resource acquisition:

Demand forecasts are also useful during medium term planning which involves the acquisition of resources and modification of the workforce. Depending on the industry, these types of planning decisions can be made over periods such as a few weeks up to several months.

Long-term strategy of the company:

Each organisation must make decisions regarding its strategic options over the long term. These decisions involve technical choices, financial and property investments, marketing actions, development of new products, etc... Generally, long term forecasts help managers improve decisions by reducing the associated risk.

3.4 Basic Concepts and Definitions

This section introduces several concepts and definitions that are central to forecasting in general. They will be used in further sections dedicated more specifically to demand forecasting.

Basically, direct statistical forecasting (see section 3.1.3) consists in extrapolating past data that all have a time reference; a month, a week, an hour, a second, …Such a collection of data with a time reference is called a time series.

3.4.1 Time Series Definition

A time series is a collection of values observed sequentially over time. For example, the monthly sold quantities, the weekly number of air plane passengers, the yearly profits, etc…

Statistical forecasting consists in estimating how the sequence of past observations in a time series will continue into the future.

3.4.2 Statistical Summaries

The most common descriptive statistics for the analysis of a time series are:

- Mean;
- Median;
- Variance;
- Standard deviation;
- Coefficient of variation.

It is also very useful to compare the observation of a time series at one time period with the observation at another time period. The autocorrelation analysis provides a useful check for seasonality, cycles and other time series patterns.

Let us consider a time series $Y_t = \{Y_1, Y_2, ..., Y_n\}$

The mean is given by $$\overline{Y} = \frac{1}{n}(Y_1 + Y_2 + ... + Y_n)$$

The median is the middle observation when all the observations of the series are arranged in increasing or decreasing order.

In statistics, the median is the value that separates the highest half of the sample from the lowest half. To find the median, arrange all the observations from lowest value to highest value and pick the middle one. If there is an even number of observations, take the mean of the two middle values.

The definition of the median is:

$\tilde{x} = Y_{(N+1)/2}$ if N is odd

with $Y_i \geq Y_{i-1}$

$\tilde{x} = \frac{1}{2}(Y_{N/2} + Y_{1+N/2})$ if N is even

The variance is defined by $$\sigma_Y^2 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}$$

The standard deviation is simply the square root of the variance. Both are measures of the variability of the variable. In probability and statistics, the standard deviation is the most commonly used measure of statistical dispersion. It provides a measure of dispersion that is 1) a non-negative number; and 2) has the same units as the data.

The standard deviation is defined by $$\sigma_Y = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}}$$

Coefficient of variation

The variance and the standard deviation are measures of absolute variability in the data of the time series. The coefficient of variation (CV) is a relative measure of this variability and is defined as the ratio of the standard deviation to the mean.

The coefficient of variation is defined by $CV_Y = \frac{\sigma_Y}{\bar{Y}}$

This coefficient of variation also provides qualitative information on the difficulty in establishing reliable statistical forecasts. The greater the value of CV, the more the data is said to be scattered, which makes it more difficult to obtain reliable forecasts. Generally, a coefficient of variation greater than one suggests a difficult time series.

However, caution is required when analysing a time series with strong trend and/or seasonality as the data is strongly fluctuating; i.e. exhibits a high CV value, but may nevertheless still be easily forecasted with an appropriate model.

In the case of a time series having a seasonal pattern and/or a trend, the coefficient of variation should be calculated only after these patterns have been removed from the time series.

3.4.3 Autocorrelation analysis

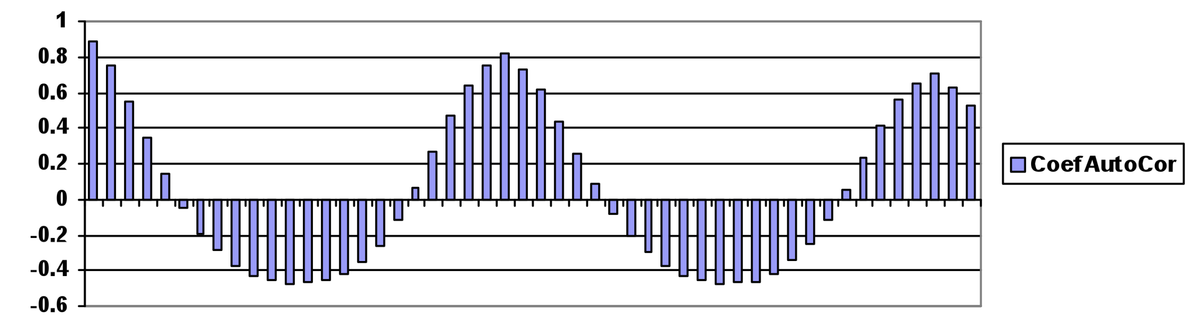
The correlation coefficients are parameters that measure the extent of the linear relationship between two variables. The autocorrelation coefficients are the correlation coefficients between two values of the same data series but separated by a fixed time lag k.

The autocorrelation coefficient at time lag k, for a series of *N* values is defined by:

$$r_k = \frac{\sum_{t=1}^{N-k} (Y_t - \overline{Y})(Y_{t+k} - \overline{Y})}{\sum_{t=1}^{N-k} (Y_t - \overline{Y})^2}$$

The series of autocorrelations for a time series at lags 1, 2 … is called the **autocorrelation function (ACF).** A plot of the ACF against the time lag is known as the correlogram. ACF can be used for several purposes, such as to identify the presence and length of seasonality and to determine whether the data of a given time series is stationary, or has a constant mean.

The next figure represents the ACF of domestic energy consumption data. These data show a strong seasonal pattern.



ACF of the domestic energy consumption time series

It is noticeable that the autocorrelation coefficients at lags 24 and 48 are higher than for the other lags. This is due to the seasonal pattern in the data: the peaks and the troughs tend to be 24 hours apart.

The ACF is also useful to check if there is a remaining pattern in the errors after a forecasting model has been applied. Generally we consider the model to be adequate when all the autocorrelations are inferior to $\pm \frac{2}{\sqrt{n}}$

where n is the number of observations taken into account in the ACF. This test only indicates that there is no remaining pattern in the forecast errors but it gives no indication concerning the accuracy of the forecasts.

3.4.4 Abnormal values

Before conducting any statistical analysis of a time series, it is necessary to check if it presents any unusual or accidental observations. These abnormal values are often observed in the gross market segment. For instance, an advertising campaign could cause an abnormal increase in demand whereas a strike in transportation companies could cause an abnormal decrease in demand. Outliers can also result from data entry errors and are often found in industrial databases.

Different methods may be used to detect such abnormal values, but none of them are completely satisfactory.

One possible method consists in verifying if the time series values belong to a certain confidence interval given by $IC = \overline{Y} \pm 1.96\sigma_Y$

The coefficient 1.96 corresponds to a confidence level of 95% for a normal distribution.

Every observation of the time series, which is outside of this interval, is considered as abnormal and is replaced for example, by the average of its neighbour values.

If the time series presents a strong trend or a seasonal pattern, this test is not adequate and should not be adopted. In this case, a confidence interval is evaluated for each year and each period. An observation is considered as abnormal if it is outside both of the “yearly” and of the “periodic” confidence interval. An example with a quarterly time series is given in the next table where IC1 and IC2 are respectively, the lower and upper limits of the confidence interval.

	1st quarter	2nd quarter	3rd quarter	4th quarter	**Mean**	**Standard deviation**	IC_1	IC_2
Year 1	166	196	83	180	156.3	43.6	70.8	241.7
Year 2	169	203	97	196	166.3	41.9	84.0	248.5
Year 3	195	205	99	215	178.5	46.4	87.5	269.5
Year 4	230	243	121	263	214.3	55.1	106.2	322.3
Year 5	296	307	123	266	248.0	73.7	103.5	392.5
Mean	211.2	230.8	104.6	224.0				
Standard deviation	48.2	41.5	15.3	34.9				
IC_1	116.7	149.5	74.7	155.6				
IC_2	305.7	312.1	134.5	292.4				

3.5 Mathematical forecasting methods

This section presents the basic mathematical description of two forecasting methods:

- Moving average;
- Exponential smoothing.

Several other approaches, such as causal models or the Box-Jenkins technique, have been proposed, but they will not be treated here.

3.5.1 Moving Average

The moving average techniques may be used for two purposes:

- For smoothing values in a time series;
- For forecasting.

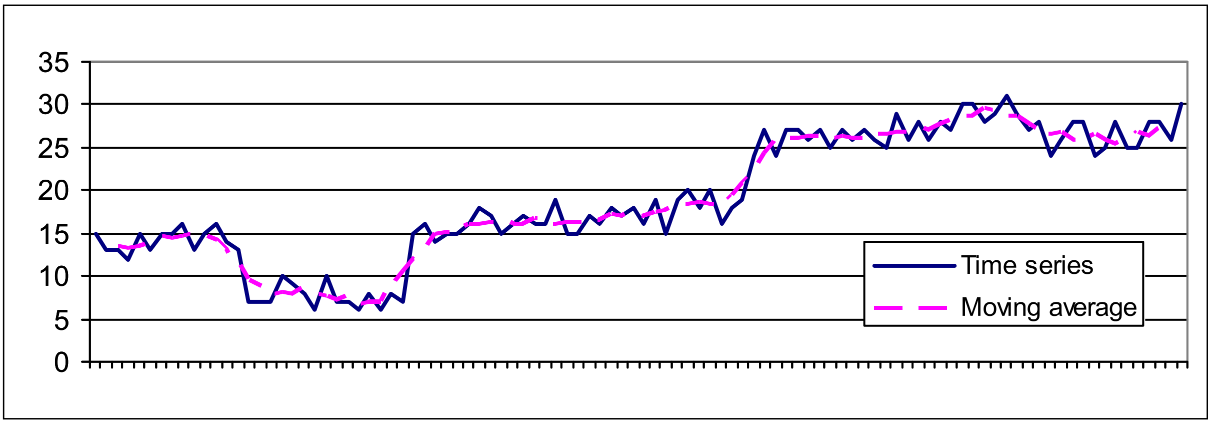
Smoothing time series values consists in reducing the random fluctuation amongst a group of observations. The moving average provides a simple method for doing this. The idea is that taking the mean value of observations within a certain neighbourhood will provide a reasonable estimate of that neighbourhood while eliminating randomness in the data.

There are different ways of calculating a moving average. Here, we consider the "centred moving average". Any original value Y_t of the time series is replaced by the average $\overleftrightarrow{Y}_t$ of its symmetrical neighbour data. Let *p* be the number of neighbour data taken into account in the "centred moving average" smoothing.

Then, if *p* is odd, $\overleftrightarrow{Y}_t = \frac{1}{p}\sum_{i=t-k}^{t+k} Y_i$ with $k = \frac{p-1}{2}$

If p is even, $\overleftrightarrow{Y}_t = \frac{1}{p}\left[\sum_{i=t-k}^{t+k} Y_i + \frac{1}{2}(Y_{t-k-1} + Y_{t+k+1})\right]$ with $k = \frac{p}{2} - 1$

The next figure illustrates the effect of the smoothing on a time series. It represents the plots of the original and of the smoothed values obtained using a "centred moving average" with *p*=5.



Smoothing effect on a time series obtained using "centred moving average" with *p*=5

Forecasting a time series using moving averages is very similar to the previous smoothing procedure. However, in this case, the moving average is not centred. Let us consider *p* as the number of observation taken into account and F_{t+1} the forecast value for period t+1.

Then, $F_{t+1} = \frac{1}{p}\sum_{i=0}^{p-1} Y_{t-i}$

In this forecasting model, we assume that the next value of the time series will be close to the *p* previous ones. Therefore, this model is well adapted for the constant or quasi-constant time series of the form $Y_t = a + e_t$ where a is constant is and e_t is a random variable with a zero mean.

3.5.2 Exponential Smoothing

In establishing forecasts with a moving average of length *p*, the *p* -last observations are considered to have a similar weight in the forecasting process. In some cases, it is more appropriate to assign different weights to the previous observations, the most recent being more heavily weighted than the oldest since it will usually provide the best guide as to the future. This can be done by attributing weighting coefficients to the *p* data considered, which requires p weighting coefficients. The exponential smoothing basically does the same thing, but uses only one single coefficient (the smoothing constant).

The forecast value for the next period with the single exponential smoothing model is:

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

where $\alpha \in [0;1]$ is the smoothing constant.

Exponential smoothing advantageously requires only two records for forecasting F_{t+1}: the previous forecast F_t and the last known observation Y_t. Of course, an initial value for F is required in the initialising phase of the single exponential smoothing. This point will be discussed in the demand forecasting section.

The development of the previous equation clearly shows that it is a weighted moving average:

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

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

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

$$F_{t-1} = \alpha Y_{t-2} + (1-\alpha)F_{t-2}$$

$$F_{t+1} = \alpha Y_t + (1-\alpha)[\alpha Y_{t-1} + (1-\alpha)(\alpha Y_{t-2} + (1-\alpha)F_{t-2})]$$

$$F_{t+1} = \alpha[Y_t + (1-\alpha)Y_{t-1} + (1-\alpha)^2 Y_{t-2}] + (1-\alpha)^3 F_{t-2}$$

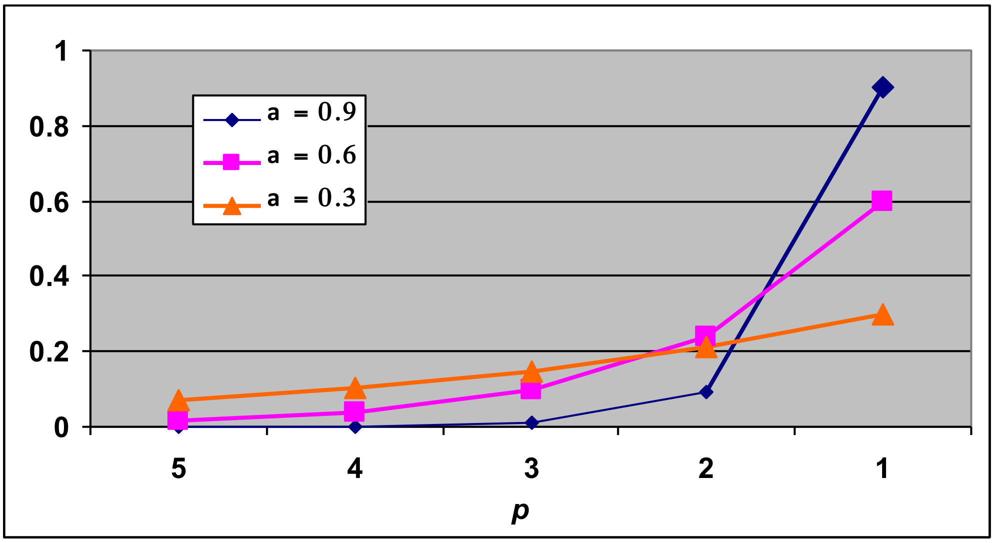
$$F_{t+1} = \alpha \sum_{p=0}^{\infty} (1-\alpha)^p Y_{t-p}$$

Obviously the value of alpha will influence the importance of the smoothing effect.

- For $\alpha \cong 0$ the model is robust and not too dependant on an abnormal value of the last observation. It will however react slowly to a change in the level of the time series.
- Conversely, for $\alpha \cong 1$the model is reactive to changes in the data and less robust.

The role of the smoothing constant is illustrated by the next figure which shows the relative contribution of past values in the calculation of the forecast for various alpha values.

Notice that in the previous equation the forecast horizon is just for one period, t+1. Forecasting over longer horizons will be dealt with in the demand forecasting section.



Contribution of past values in the calculation of the forecast for various α values

3.5.3 Forecasting error determination

Reliability plays an important role in evaluating forecasting methods. *Forecast reliability* measures how well the results of the forecasting process correspond to the actual data.

While *forecast reliability* is a significant factor in evaluating forecasts, it is difficult to define. This problem stems from the absence of a universally accepted measure of forecast error. Different methods usually lead to different measures of error. Since *forecast reliability* is an important concept in forecasting, it is equally important to understand the strengths and weaknesses of the forecast error calculation methods for a given context.

If Y_t is the actual value at period t and F_t is the forecasted value for this period, the forecast error E_t for this period is the difference between the actual and the forecasted values:

$$E_t = Y_t - F_t$$

Since a forecast error of e units has a completely different significance if the actual value Y was 10e units versus 100e units, a relative error measure, called percentage error PE, is defined:

$$PE_t = \frac{Y_t - F_t}{Y_t} = \frac{E_t}{Y_t}$$

If there are actual values and forecasts over a horizon $\mathcal{H}$ of h periods, then any error term number $n \in [1;h]$ can be considered in computing a measure of the *forecast reliability*. Accordingly, the following standard statistical estimators of the *forecast reliability* can be defined:

- Mean error: $ME_t = \frac{1}{n}\sum_{i=t-n+1}^{t} E_i$
- Mean percentage error: $MPE_t = \frac{1}{n}\sum_{i=t-n+1}^{t} PE_i$
- Mean absolute error: $MAE_t = \frac{1}{n}\sum_{i=t-n+1}^{t} |E_i|$
- Mean absolute percentage error: $MAPE_t = \frac{1}{n}\sum_{i=t-n+1}^{t} |PE_i|$
- Mean square error: $MSE_t = \frac{1}{n}\sum_{i=t-n+1}^{t} (E_i)^2$
- Root mean square error: $RMSE_t = \sqrt{\frac{1}{n}\sum_{i=t-n+1}^{t} (E_i)^2}$

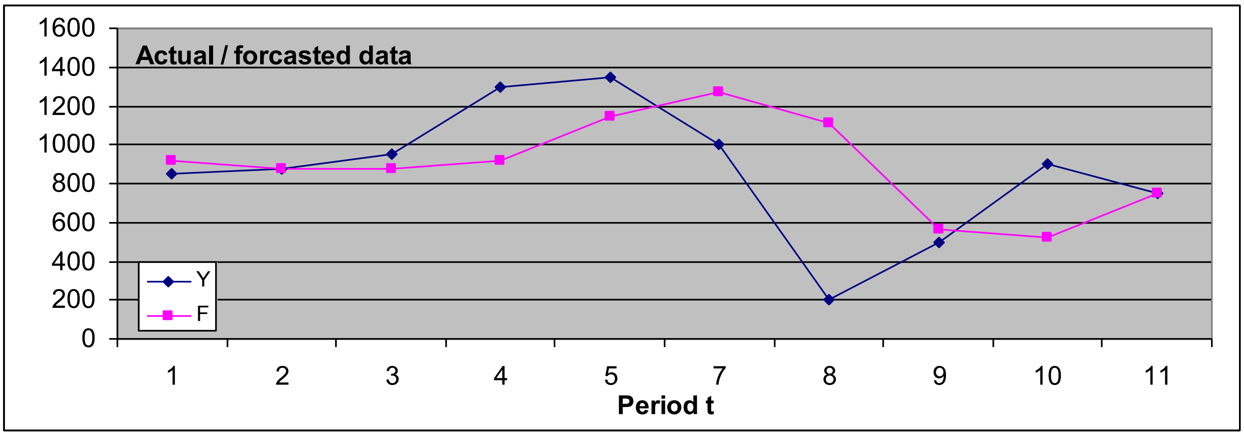
Note that in the previous equations the mean values are not centred. The reason is that in forecasting one is interested in the reliability of the n last forecasts immediately proceeding the current period.

Instead of using an average calculation with a fixed number of n terms, the exponential smoothing can be used to compute smoothed average values. It leads to the definition of the following statistical estimator of the *forecast reliability*, with $\alpha \in [0;1]$:

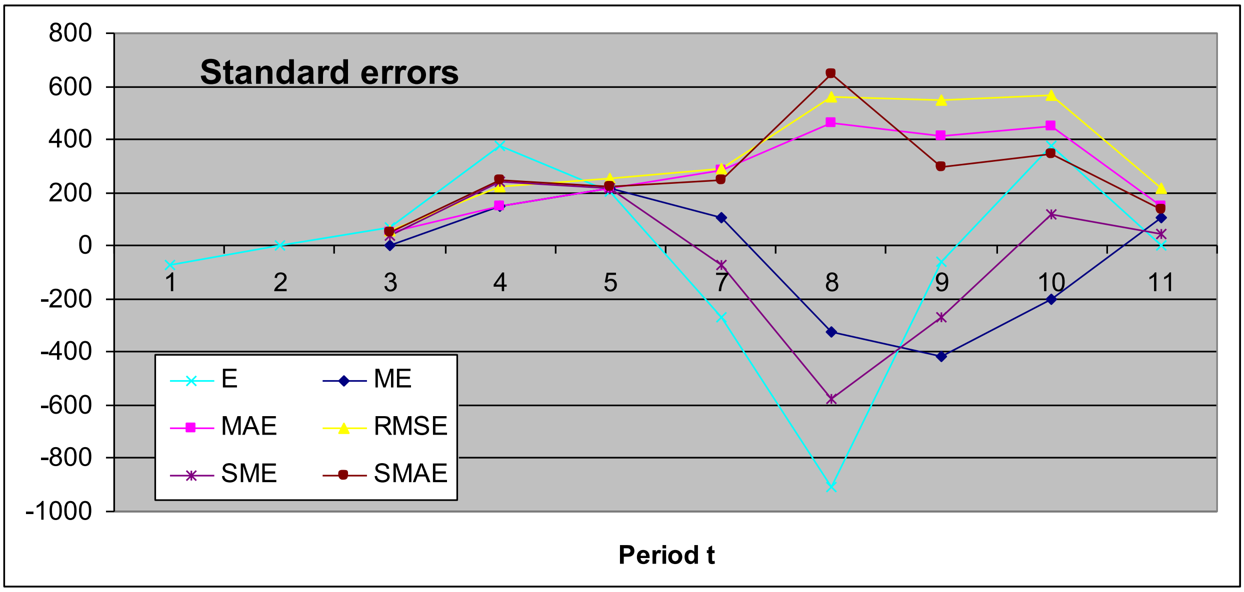
- Smoothed mean error: $SME_t = \alpha E_t + (1-\alpha) SME_{t-1}$
- Smoothed mean absolute error: $SMAE_t = \alpha |E_t| + (1-\alpha) SMAE_{t-1}$
- Smoothed mean percentage error: $SMPE_t = \alpha PE_t + (1-\alpha) SMPE_{t-1}$
- Smoothed mean absolute percentage error: $SMAPE_t = \alpha |PE_t| + (1-\alpha) SMAPE_{t-1}$

The next table and figures provide a simple example of the estimation of forecast reliability over a horizon $\mathcal{H}$ of 11 periods.

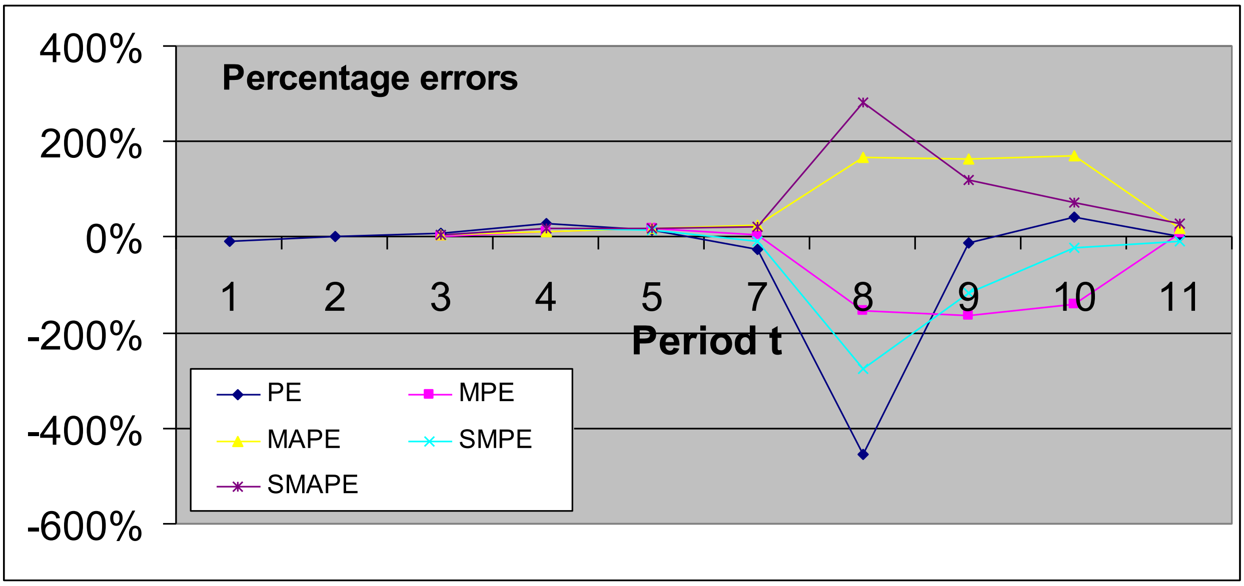
			α= 0.6			*n* = 3				
Period t	1	2	3	4	5	7	8	9	10	11
Actual value Y	850	880	950	1300	1350	1000	200	500	900	750
Forecasted value F	920	878	879	922	1149	1269	1108	563	525	750
Error E	-70	2	71	378	201	-269	-908	-63	375	0
Percentage error PE	-8%	0%	7%	29%	15%	-27%	-454%	-13%	42%	0%
Mean error ME			1	150	217	103	-325	-413	-199	104
Mean perventage error MPE			0%	12%	17%	6%	-155%	-164%	-142%	10%
Mean absolute error MAE			48	150	217	283	460	413	449	146
Mean absolute percentage error MAPE			5%	12%	17%	24%	165%	164%	169%	18%
Mean square error MSE			3306	49381	62891	85424	312408	300225	322834	48141
Root mean square error RMSE			57	222	251	292	559	548	568	219
Smoothed mean error SME			36	241	217	-75	-575	-268	118	47
Smoothed mean absolute error SMAE			50	247	220	249	644	296	343	137
Smoothed mean percentage error SMPE			4%	19%	17%	-10%	-276%	-118%	-22%	-9%
Smoothed mean absolute percentage error SMAPE			5%	20%	17%	23%	281%	120%	73%	29%
Tracking signal TS			0.02	1.00	1.00	0.37	-0.71	-1.00	-0.44	0.71
Smoothed tracking signal STS			0.73	0.98	0.99	-0.30	-0.89	-0.91	0.34	0.34



Forecast example: actual, Y versus forecasted, F data



Forecast reliability estimation: standard errors



Forecast reliability estimation: percentage errors

The selection of the appropriate forecast error calculation requires some comments:

- It is obvious that the result of the error calculation depends on *n*, the number of observations used in the data set or α, the smoothing constant.
- Standard errors, such as mean error ME, SME or root mean square error MSE, typically provide an error measure in the same units as the data. However, the real significance of the error is difficult to evaluate.
- The use of a percentage error facilitates the understanding of its significance. However, since the percentage error is defined as a ratio, problems arise in the computation of values that are zero or close to zero.
- If the sign of the error is significant (for example non-symmetrical loss function), a non-absolute error may be preferred. However, if the error sign changes regularly during the observed horizon, the calculated forecast error may be close to zero due to the compensation of positive and negative values, although the error per period may be very large (but of alternating sign).
- The use of absolute error, like the mean absolute error (MAE), makes it possible to take into account both positive and negative errors (they may cancel each other out in the mean error ME). However, the shortcoming of these calculations is that they assume a symmetrical loss function, where the organisational cost of over-forecasting is assumed to be the same as under-forecasting and they are summed together. With these measures, the total magnitude of error is provided but not the true bias or direction of the error.
- With the mean square error (MSE), errors are weighted based on their magnitude: larger errors are given more weight than smaller ones. This can be quite beneficial in situations where the loss function increases with the square of the error. The disadvantage of MSE is that it is inherently difficult to understand.
- Using the root mean square error (RMSE), which is simply the square root of MSE, may be preferred as the error is provided in the same units as the data. Like the MSE, the RMSE penalizes errors according to their magnitude. Nevertheless, remember that both MSE and RMSE are not unit-free and therefore comparisons across series are difficult.
- Outliers can distort the evaluation of forecast error. The problem can arise due to a mistake in recording the data, promotional events or just random occurrences. In order to avoid the problem, outliers can be filtered out before forecasting, particularly

if they are due to recording mistakes. Another solution is to use the median instead of the mean, which will remove higher and lower values in favour of the middle values.

3.5.4 Tracking signal

Mathematical forecasting consists in extracting some mathematical modelled representation of the data in past time series and to use this model to compute future data. The very important and implicit hypothesis here is that the model of the past will still hold for the future. In other words, a continuity of the main historical pattern is assumed; i.e. the fundamental behaviour of the system must stay the same. Consequently, it is important to have a means of checking whether this very critical assumption is valid.

Generally, it consists of computing a monitoring signal and of defining critical limits between which the model is considered to remain valid. When the signal crosses a limit, the model validity must be checked and the model possibly revised. The tracking signal provides a means for monitoring the validity of a forecast model.

A good forecasting model should produce errors that are independent and have a mean of zero. If the mean error shows a continuously positive or negative value, then the forecast is biased and the model must be re-evaluated. The tracking signal (*TS)* is an indicator of forecast bias and it is defined as the ratio of the mean forecast error to the absolute mean forecast error. For a forecast horizon $\mathcal{H}$ of *h* periods:

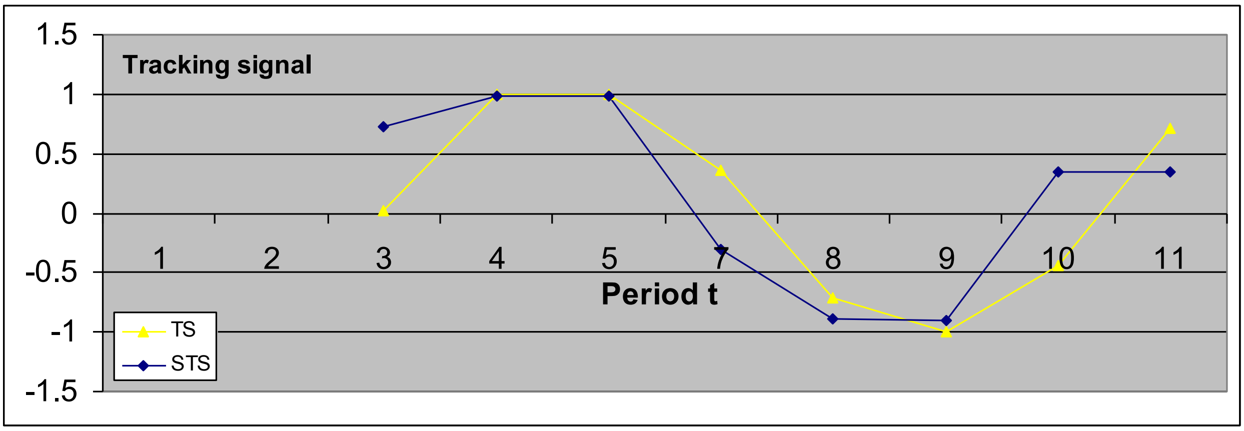
$$TS_t = \frac{ME_t}{MAE_t} = \frac{\sum_{i=t-n+1}^{t} E_i}{\sum_{i=t-n+1}^{t} |E_i|} \quad \text{with } n \in [1;h]$$

Exponential smoothing can be applied for the evaluation of a smoothed tracking signal (*STS)*, in order to give more weight to the most recent demand data:

$$STS_t = \frac{\delta E_t + (1-\delta) SME_{t-1}}{\delta |E_t| + (1-\delta) SMAE_{t-1}}$$

The tracking signal varies between –1 and +1. It is an indicator of forecast bias that is consistent for all observations. A tracking signal far from zero indicates that the forecasting model should be reviewed.

The next figure represents the evolution of the tracking signal for the previous forecast example over a horizon $\mathcal{H}$ of 11 periods.



Tracking signal, *TS* and smoothed tracking signal, *STS*

3.6 Demand forecasting models

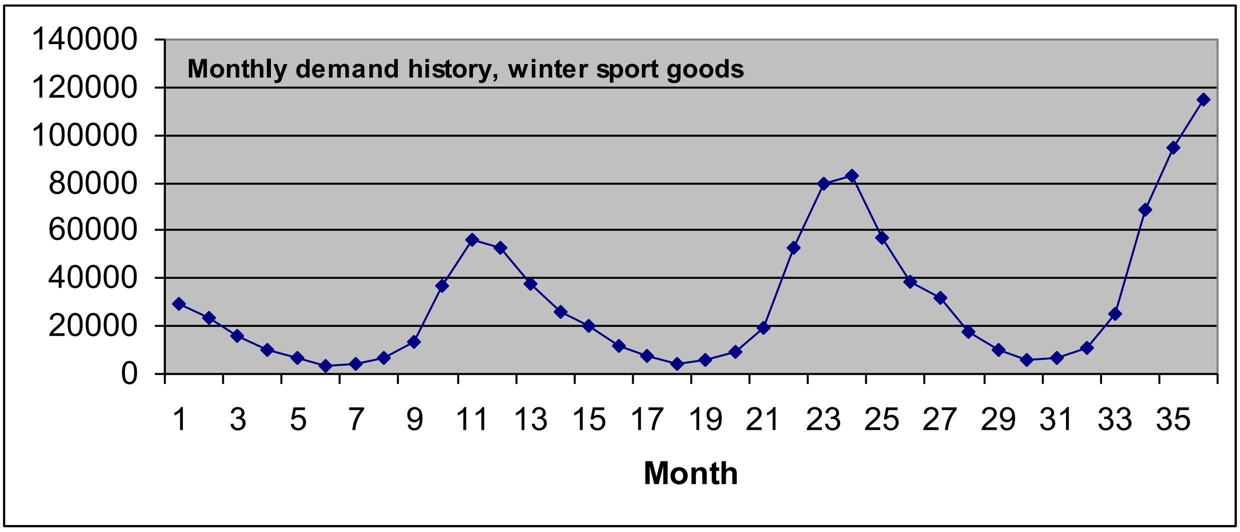
Demand forecasting consists in finding an estimate of future demand that is as reliable as possible. As discussed at the beginning of this chapter, such anticipation is necessary for decision-making at the strategic and tactical levels, as well as for operating the manufacturing company, particularly in the upstream part of the *value adding chain*.

This section is restricted to presenting demand forecasting models based on exponential smoothing.

3.6.1 Decomposition principle

In real cases, demand varies cyclically due to seasonal effects and/or to economic cycles. Furthermore, due to the usual product life cycles, the demand for a given product or product family can exhibit a growing trend, a stabilisation or a decreasing trend.

The next figure provides an example of the historical evolution of the demand, over a horizon of 36 months, for a product family of a winter sport goods producer. It can easily be observed that the demand varies cyclically with an annual cycle (seasonality) and that it grows every year.



Example of monthly demand history for a product family

The forecasting of such a demand requires the identification and description of its constitutive elements, in the case of the previous example:

- A base level that describes the average volume of the demand;
- A seasonal modulation that describes how the base level is cyclically modified;
- A trend component that describes the increase of the demand with time.

The objective of decomposing a time series is to isolate and model its constituting components. One possible means to do this is to decompose any demand time series into five components:

- The base level $\mathcal{B}$ that provides the average volume of the demand;
- The trend $\mathcal{T}$ that expresses a long term change in the mean level, where the long term period is a function of the model;
- The seasonal effect $\mathcal{S}$ that corresponds to variations repeating at regular intervals of *c* periods;

- A cyclic pattern, $\mathcal{C}$, apart from the seasonal variation, that may have various causes and is often found for example, in economic data such as in business cycles in the electronic or automotive industries;
- Random fluctuations $\mathcal{R}$ that are all phenomena not accounted for in the previous components and representing the unexplained stochastic variability of the demand.

The decomposition is based on the assumption that the time series $\mathcal{Y}$ can be completely described by:

Data = Pattern + Random fluctuations

$$\mathcal{Y} = \mathcal{F}(\mathcal{B}, \mathcal{T}, \mathcal{S}, \mathcal{C}) + \mathcal{R}$$

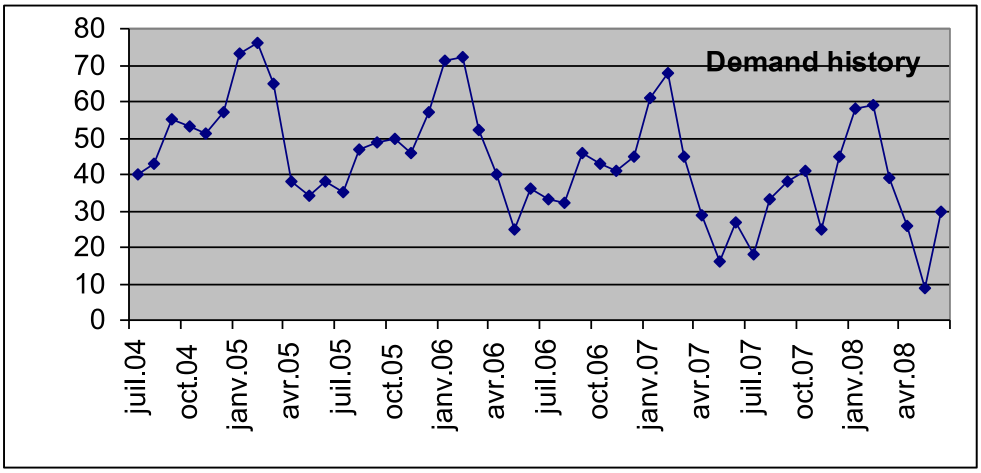
In the following description, for reason of simplicity, the cyclic component $\mathcal{C}$ will not be considered, as it can be handled similarly to the seasonal component $\mathcal{S}$. Consequently, the decomposition will be limited to:

$$\mathcal{Y} = \mathcal{F}(\mathcal{B}, \mathcal{T}, \mathcal{S}) + \mathcal{R}$$

The exact mathematical formulation of the function $\mathcal{F}$ depends on the relation between the magnitude of the trend, seasonal components and the base level of the time series. If no relation exists, the components of the time series can be considered as being independent and the time series is modelled with an additive form:

$$\mathcal{Y} = \mathcal{B} + \mathcal{T} + \mathcal{S} + \mathcal{R}$$

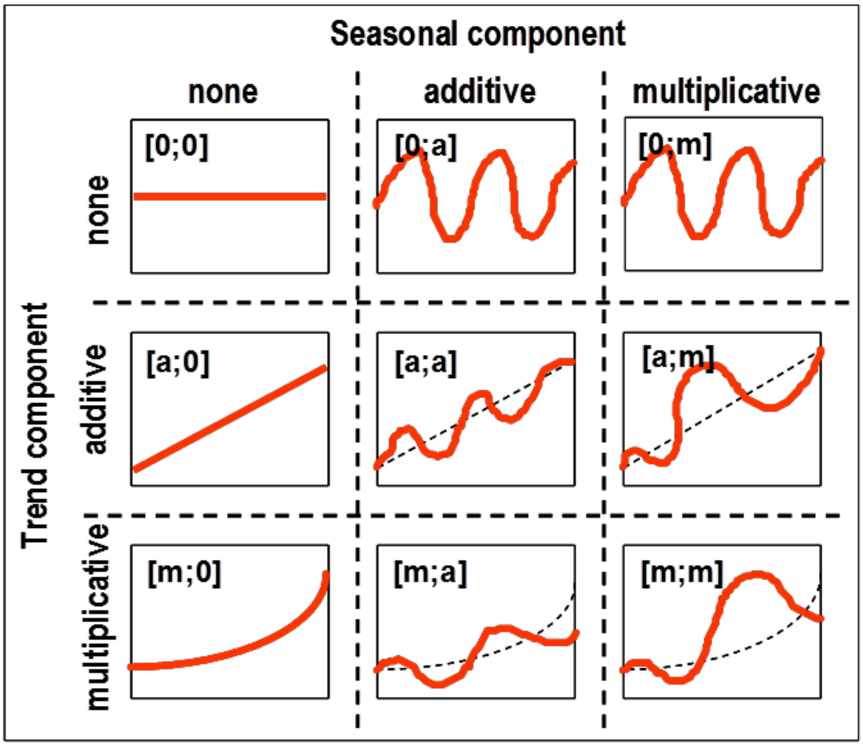
The next figure provides an example of a demand with an additive form.



Example of a demand with an additive form

In contrast, if a component increases or decreases proportionally to the variation of the base level, then a multiplicative model is appropriate.

Both the seasonal, as well as the trend component, can be of an additive or multiplicative nature. Consequently, nine combinations are mainly possible as schematically illustrated in the next figure; the combinations are noted by the permutations of the two components $[\mathcal{T}:\mathcal{S}]$ with the three possible states *0,a,m* for none, additive and multiplicative relations respectively.



Combination of additive/multiplicative seasonal/trend components

Multiplicative decomposition is more common with economic series because most seasonal economic series have seasonal variations, which increase with the level of the series.

The most commonly found decompositions in practice are:

[a;0] $\mathcal{Y} = \mathcal{B} + \mathcal{T} + \mathcal{R}$

[0;m] $\mathcal{Y} = \mathcal{B} \times \mathcal{S} + \mathcal{R}$

[a;m] Y= (B + T) S+ R

The demand of the winter goods presented above is of [a:m] type

Note that the random component $\mathcal{R}$ can be described as multiplicative also.

3.6.2 Demand forecasting models

Based on the decomposition principle, various demand forecasting models have been proposed. In this section, three well known models will be presented which handle the three most common decompositions cases presented above.

When neither a trend, nor a seasonal component is present, **[0;0]** case, single exponential smoothing is used to eliminate the random component. The model is then as described under 3.5.2:

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

And for any period > t+1,

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

[a;0] case, additional trend

For the **[a;0]** case (additive trend, no seasonal component), the Holt model is used. This model takes into account the presence of an additive trend component. It uses two smoothing constants α and β to estimate the base level of the time series (B_t) and its per-period trend (T_t) for a defined period *t*. The forecasts are established with three equations:

$$B_t = \alpha Y_t + (1-\alpha)(B_{t-1} + T_{t-1})$$

$$T_t = \beta(B_t - B_{t-1}) + (1-\beta)T_{t-1}$$

$$F_{t+h} = B_t + hT_t$$

In this model, two initial values are required: the base level *B* and the trend *T*. One alternative is to set for the initial values:

$B_t = Y_t$ and

$T_t = Y_t - Y_{t-1}$ or $T_t = \dfrac{Y_t - Y_{t-n}}{n}$

[0;m] case, multiplicative seasonality

Similarly to the Holt model for the additive trend case, the [0;m] case is modelled by using two smoothing constants α and γ to estimate the base level of the time series (B_t) and its per-period seasonal component (S_t) for a defined period *t*. The forecasts are established with three equations:

$$B_t = \alpha \frac{Y_t}{S_{t-c}} + (1-\alpha)B_{t-1}$$

$$S_t = \gamma \frac{Y_t}{B_t} + (1-\gamma)S_{t-c}$$

$$F_{t+h} = B_t \times S_{t+h-c}$$

with c = seasonal cycle periodicity (often 12 months)

In this model, two initial values are required: the base level *B* and the seasonal component *S*. One alternative is to set for the initial values:

$B_t = Y_t$ and

For any $S_j = \dfrac{Y_j}{\frac{1}{c}\sum_{i=j-c+1}^{j} Y_i}$

[a;m] case, additive trend & multiplicative seasonality

The Holt model for additive trend has been extended to include a multiplicative seasonality. The Holt and Winter's model is used to forecast time series with combined trend and seasonal components. The forecast requires three smoothing equations, integrating three smoothing constants: α for calculating the base level of the time series (B_t), β for the trend component (T_t) and γ for the seasonal one (S_t).

$$B_t = \alpha \frac{Y_t}{S_{t-c}} + (1-\alpha)(B_{t-1} + T_{t-1})$$

$$T_t = \beta(B_t - B_{t-1}) + (1-\beta)T_{t-1}$$

$$S_t = \gamma \frac{Y_t}{B_t} + (1-\gamma)S_{t-c}$$

$$F_{t+h} = (B_t + hT_t) \times S_{t+h-c}$$

The model initialisation requires three values: the base level *B,* the trend *T* and the seasonal component *S*. One alternative is to set for the initial values:

$$T_t = \frac{1}{nc}\left[(Y_t - Y_{t-c}) + (Y_{t-1} - Y_{t-1-c}) + ... + (Y_{t-n} - Y_{t-n-c})\right]$$

$$B_t = \frac{1}{c}\sum_{i=t-c+1}^{t} Y_i + \frac{c-1}{2}T_t$$

$$S_{t-j} = \frac{Y_{t-j}}{B_{t-j}} \text{ and } B_{t-j} = \frac{1}{c}\sum_{i=t-c+1}^{t} Y_i + (c/2 - j)T_t$$

[a;a] case, additive trend & additive seasonality

Although less common in practice, the [a;a] case can also be modelled using the Holt and Winter's model. Again, three smoothing constant α for the base level (B_t), β for the trend component (T_t) and γ for the seasonal one (S_t) are required.

$$B_t = \alpha(Y_t - S_{t-c}) + (1-\alpha)(B_{t-1} + T_{t-1})$$

$$T_t = \beta(B_t - B_{t-1}) + (1-\beta)T_{t-1}$$

$$S_t = \gamma(Y_t - B_t) + (1-\gamma)S_{t-c}$$

$$F_{t+h} = (B_t + hT_t) + S_{t+h-c}$$

The initial values for B_t and T_t are identical to those for the multiplicative method but the seasonal components are initialised with:

$$S_{t-j} = Y_{t-j} - B_{t-j} \text{ with } B_{t-j} = \frac{1}{c}\sum_{i=t-c+1}^{t} Y_i + (c/2 - j)T_t$$

3.7 Demand forecasting methodology

Forecasting demand in a real situation requires some systematic procedure in order to produce reliable results in an efficient way. This section proposes a methodology to support demand forecasting on the basis of raw historical data, as they would be obtained in practice by extraction of an Enterprise Resource Planning system or any enterprise data warehouse.

Important warning:

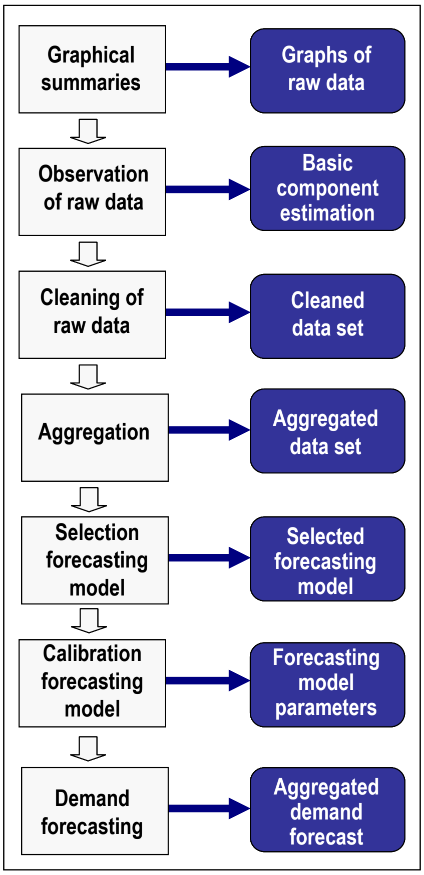
- The proposed methodology should in no circumstances be used blindly as a kitchen recipe;
- It must be used as a support for the elaboration of an appropriate methodology related to a specific case;
- Each specific situation requires a particular adaptation of the proposed methodology.

3.7.1 General demand forecasting methodology description

The process starts with raw data extracted from a computer system. It ends with the delivery of demand forecasts for a given *forecast horizon* and demand aggregation level. The main steps are:

1. Graphical summaries and observation of the raw data
2. Cleaning the raw data
3. Aggregation of the cleaned raw data
4. Selection of the forecasting model
5. Calibration of the forecasting model
6. Demand forecasting

This procedure is schematically represented in the flowchart of the next figure.



Flowchart demand forecasting methodology

3.7.2 Graphical summaries and observation of the raw data

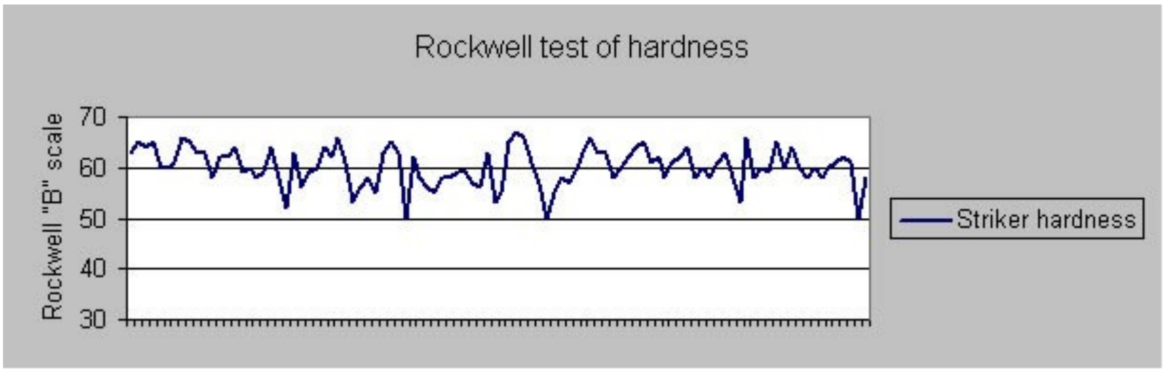
The foremost step when analysing time series is to plot data in graphs. Often, the basic characteristics of the data such as patterns or any unusual information may be quickly observed in the graph. Furthermore, the graphical observation can help decide if deeper statistical analyses are required for the identification of the time series behaviours, and it may also suggest possible explanations for some of the variations in the data.

As a time series is a collection of sequential observations and therefore is not a continuous function, it should be represented in a chart graph adopted for discrete functions. Nevertheless, it is much more practical to represent a time series in a time plot because it immediately reveals any trends over time, any seasonal behaviour and other systematic features of the data.

The four following types of data patterns can be observed from time series:

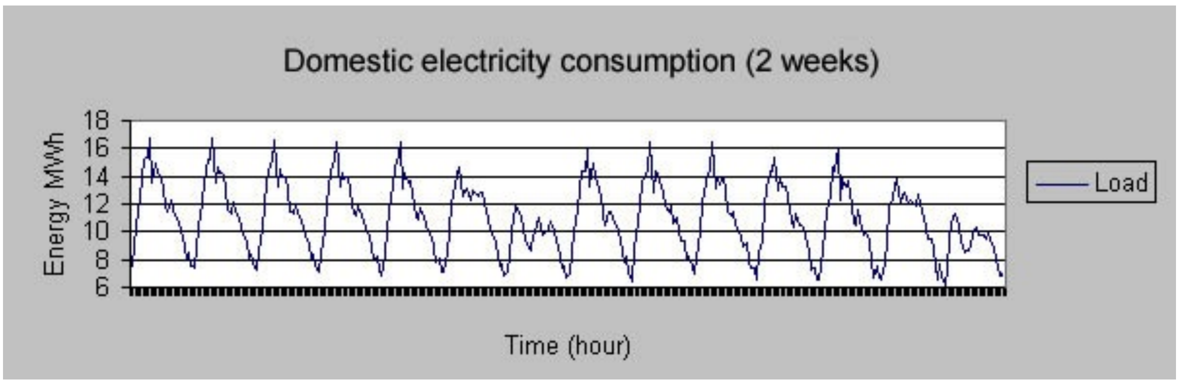
- Constant or quasi constant (horizontal) ;
- Seasonal;
- Cyclical;
- Trend.

A time series is said to be constant or quasi-constant when its values fluctuate around a constant mean. Such a series is said to be stationary in its mean. In the next figure results of a Rockwell test on 100 strikers consecutively produced are reported. This time series is constant since values fluctuate around a mean of about 60 units.



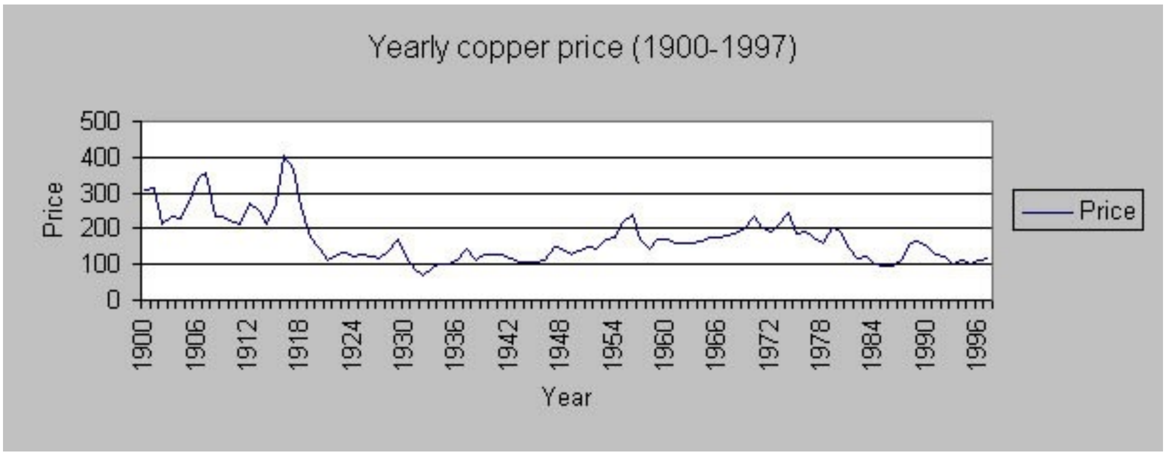
Constant series: Rockwell hardness test

A seasonal pattern is present in a time series when the values of the data vary according to a fixed time interval (a few hours, one week, one quarter, one year, etc…). For example, the domestic electricity consumption always exhibits seasonal fluctuations since the needs vary regularly in function of the time of the day, the day of the week and the season (see next figure).



Seasonal series: electricity consumption

A cyclical pattern exists when the data exhibit rises and falls over a rather long interval that is not of a fixed number of periods, contrary to a seasonal pattern. This type of pattern is typically observed with economic series where the cycles are usually due to economic fluctuations, such as those associated with business cycles. In the following time series of yearly copper prices, different cycles corresponding to the periods of price increase and decrease can be observed.

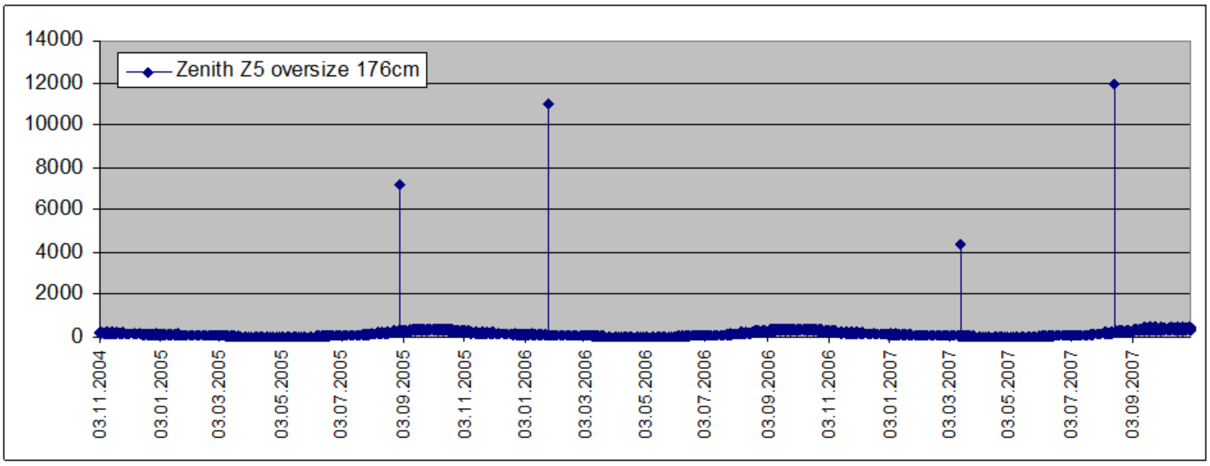


Cyclic series: copper price

A time series presents a trend when there is a significantly long increase or decrease in the data. This type of pattern is very common and may be found in many situations such as with a turnover of a company, gross national product, etc... In the following figure, the annual industrial production in the U.S. from 1860 till 1970 clearly shows an exponential trend.

Series with trend: US industrial production

The following figure provides an example of the plot of raw sales data for one finished product of a winter good producer (the WinTech company) over a 36 month time horizon.



Historic daily sales data for one finished product of WinTech

Three observations can be made:

- The series is continuous; sales values are present every month and probably every week and day;
- Four unusual points (outliers) are observed. It can be expected that they are due to recording errors;
- A weak seasonal effect can be observed in the form of yearly increases around September-October.

The next step is obviously to clean the raw data and remove the outliers.

3.7.3 Cleaning of the raw data

To clean the raw data, the confidence interval approach presented under 3.4.4 is used.

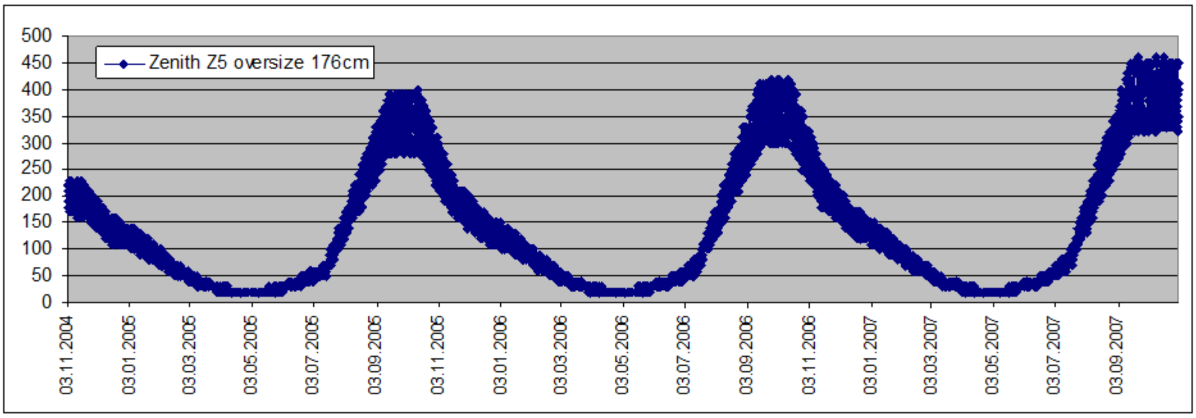
Although the series seems to present seasonal variations, the graphical observation indicates that a first cleaning over the full horizon would be appropriate as the outliers are far apart from the main curve.

The values Y_i satisfying the following criterion

$Y_i \geq \overline{Y} + 1.96\sigma_Y$ are eliminated and replaced by

$$\breve{Y}_i = \frac{Y_{i+1} + Y_{i-1}}{2}$$

The cleaned series is then plotted again and its general appearance is observed. The next figure shows the same data set for one finished product of WinTech after cleaning; i.e. replacing the four outliers.



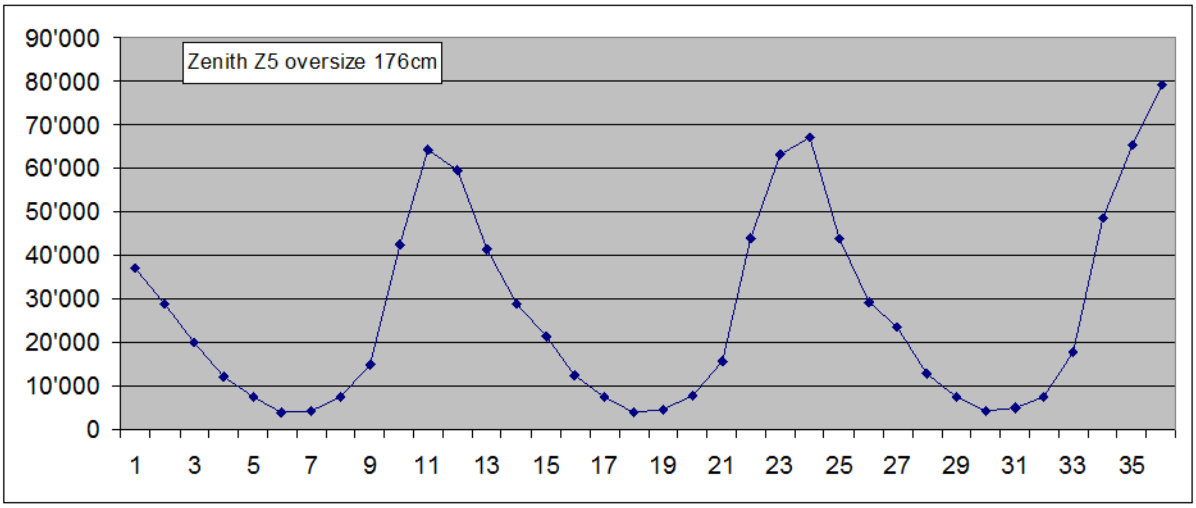
Cleaned data of the historic daily sales of WinTech for one finished product

The data look now much more regular and the seasonal effect is totally visible. Its annual periodicity is confirmed which is not surprising for a winter goods producer.

3.7.4 Aggregation of the cleaned raw data

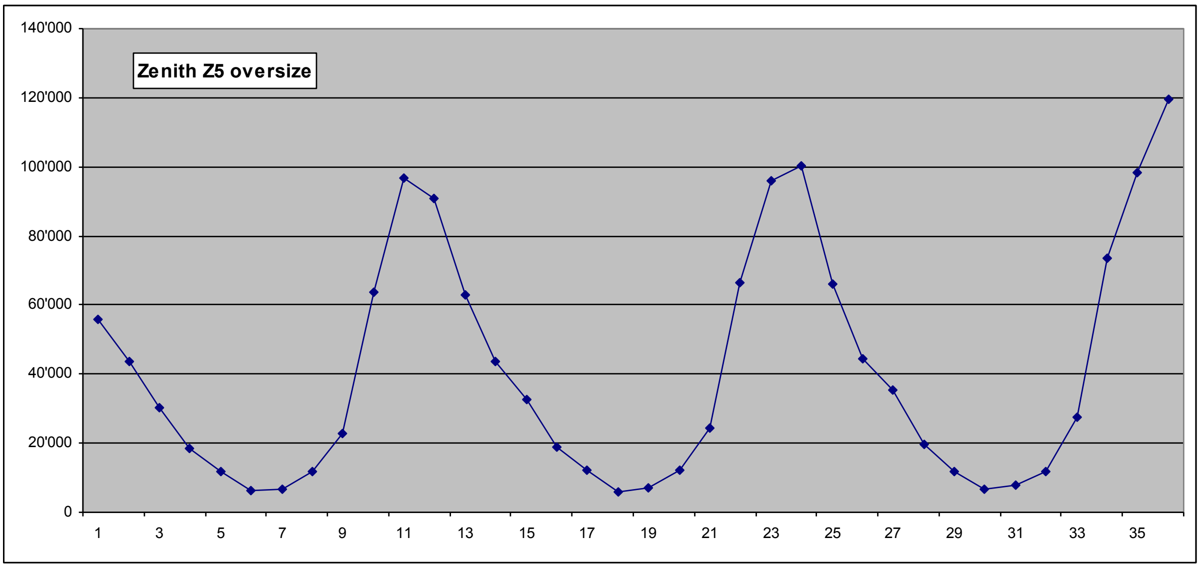
The data considered so far are single customer deliveries that are ranked according to a time scale. In a company forecasts are required per *planning period*. If the purpose of demand management is the elaboration of the *Production Plan*, which is an aggregated plan, the *planning period* is usually a month. The next step then consists in aggregating the raw data per month in order to generate a new monthly time series.

The next figure presents the monthly aggregated time series of the historical sales for the same finished product of the WinTech company.



Monthly aggregated sales data of WinTech for one finished product

If the purpose of the forecasting process is to elaborate the *Production Plan*, an additional aggregation for the *product family* (Zenith Z5 oversize) may be required. This aggregation consists simply in adding all the monthly sales data of each finished product belonging to the same product family. The next figure shows the results of this family aggregation.



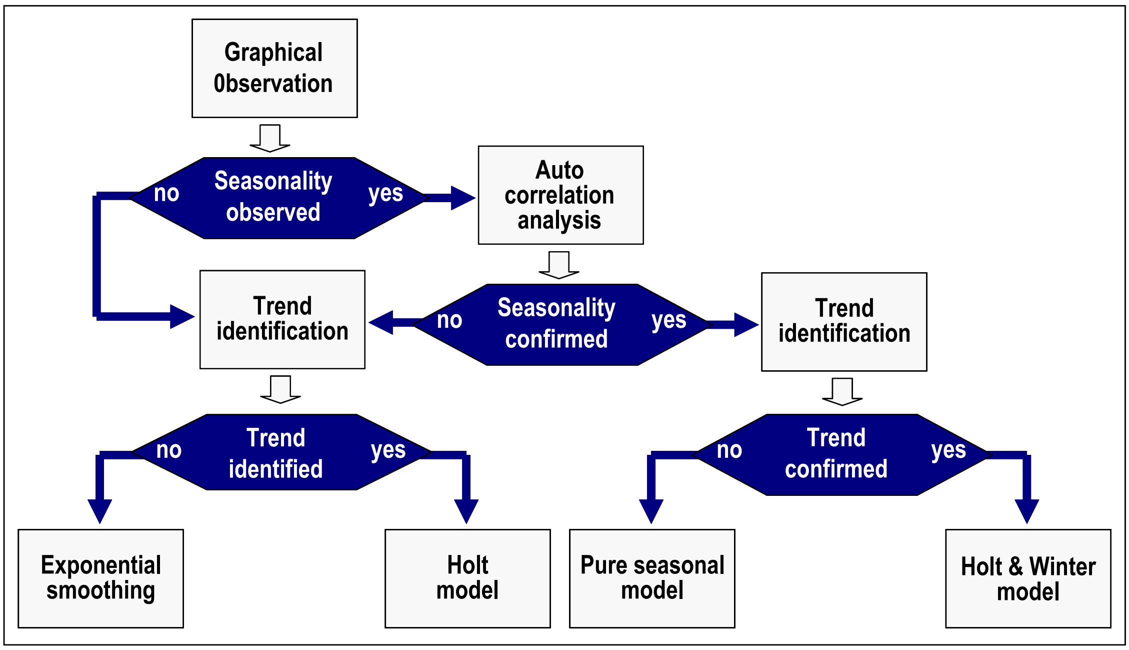
Monthly aggregated sales data of WinTech for one product family

A simple observation suggests the existence of a strong seasonal effect and a slight positive trend.

3.7.5 Determination of the forecasting model

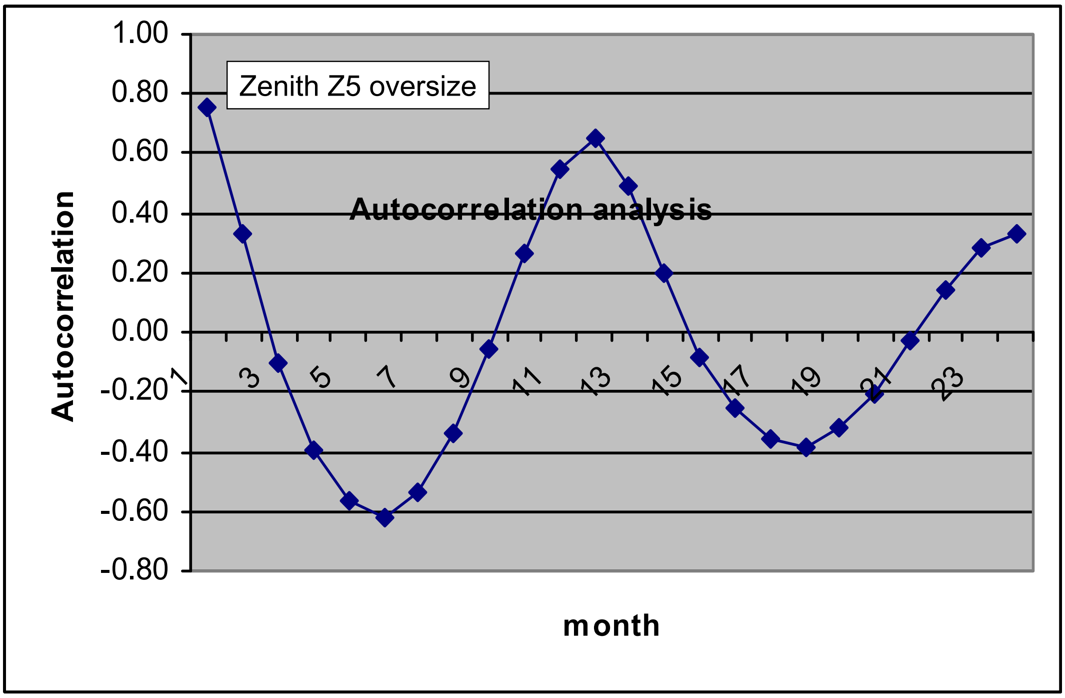
This section is restricted to the choice of an appropriate model for forecasting on the basis of exponential smoothing and the decomposition principle.

As seen in 3.6.1, a time series may be decomposed into seasonality, trend and randomness (possibly a cyclic effect). There is no single best procedure to analyse a time series and identify its components as the optimal procedure may depend on the relative importance of the trend versus seasonal components. The first step is to identify and eliminate the most significant component as estimated from the graphical observation. The following flowchart describes the procedure for the case of a dominating seasonal component. It should simply be reversed for the case of a dominating trend component.



Decision flowchart for forecasting model selection

In the WinTech case, the seasonal effect is obvious and its periodicity is also straightforward; i.e. 12 months. If any doubt were to appear, a verification using autocorrelation analysis could be made. The result of the autocorrelation for the aggregated data (Zenith Z5 oversize) is given in the next figure.



Autocorrelogram for the Zenith Z5 aggregated data

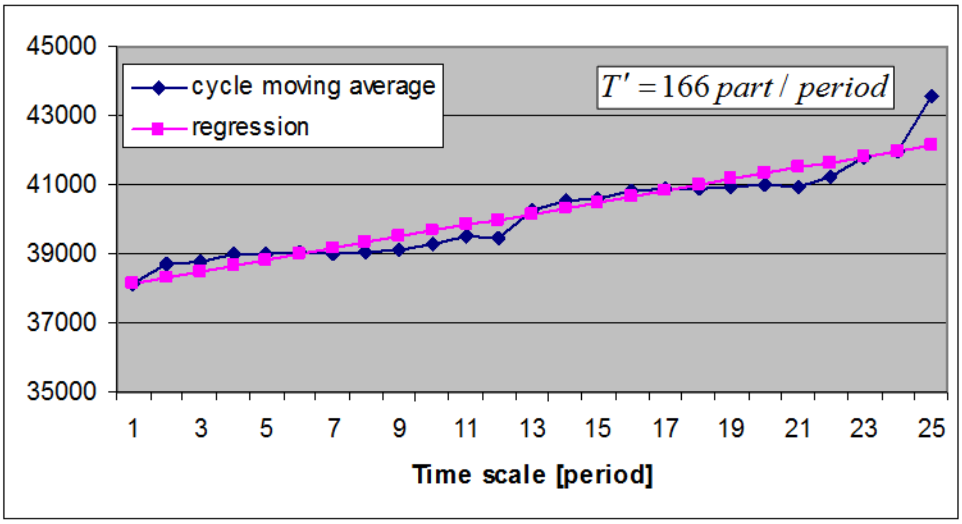
It clearly confirms the existence of a 12 month seasonal effect.

According to the decision flowchart, the next step consists in the identification of a significant trend effect. In the case of an additive trend, this can be done be calculating the cycle moving average values $\overline{Y}_t$ observing their graphical plot and calculating the trend coefficient by linear regression.

$$\overline{Y}_t = \frac{1}{c} \sum_{i=t-(c-1)/2}^{i=t+(c-1)/2} Y_i \qquad \text{for c = odd}$$

$$\overline{Y}_t = \frac{1}{c} \left[\sum_{i=t-(c-2)/2}^{i=t+(c-2)/2} Y_i + \frac{1}{2}\left(Y_{t-c/2} + Y_{t+c/2}\right) \right] \qquad \text{for c = even}$$

In the next figure, the $\overline{Y}_t$ values and the linear regression line are represented for the previous WinTech Zenith Z5 family. It can be concluded that an additive trend model is reasonable. The linear regression calculation provides an initial value of the additive trend coefficient $T' = 166\, part / period$



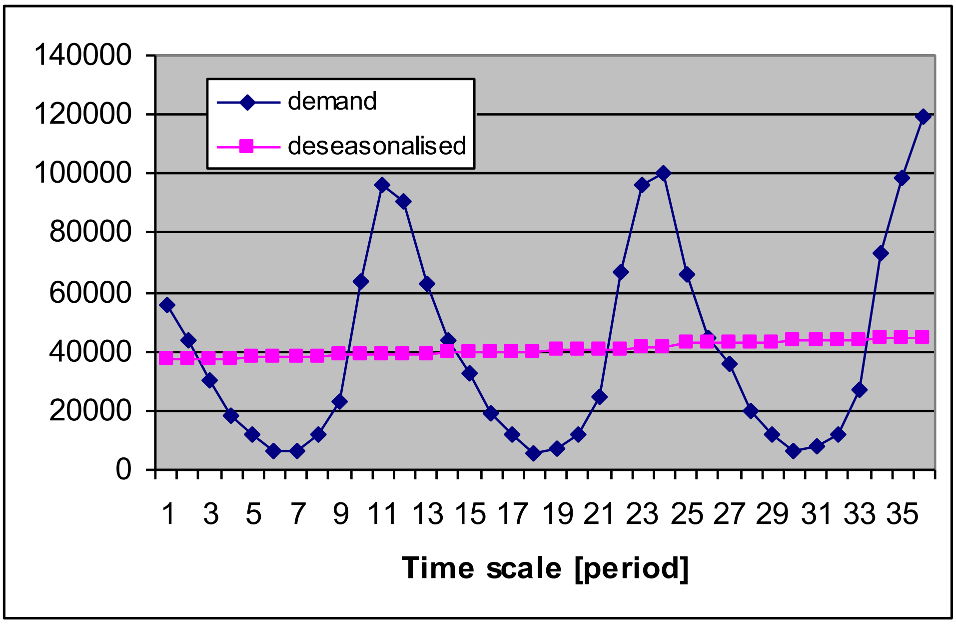
Additive trend identification for WinTech Z5 family

The next step consists in determining the values of the seasonal components. This can be obtained by eliminating the seasonal effect and creating a so-called deseasonalised time series $\tilde{Y}_t$ according to:

$$\tilde{Y}_t = \overline{Y}_c + T'\left[t - t^\circ - \frac{c-1}{2}\right]$$

with cycle = $\left[t^\circ; t^\circ + (c-1)\right]$ and cycle average $\overline{Y}_c = \frac{1}{c}\sum_{t=t^\circ}^{t^\circ+(c-1)} Y_t$

The next figure shows the deseasonalised time series $\tilde{Y}_t$ for the WinTech Zenith Z5 family.



Deseasonalised time series for WinTech Z5 family

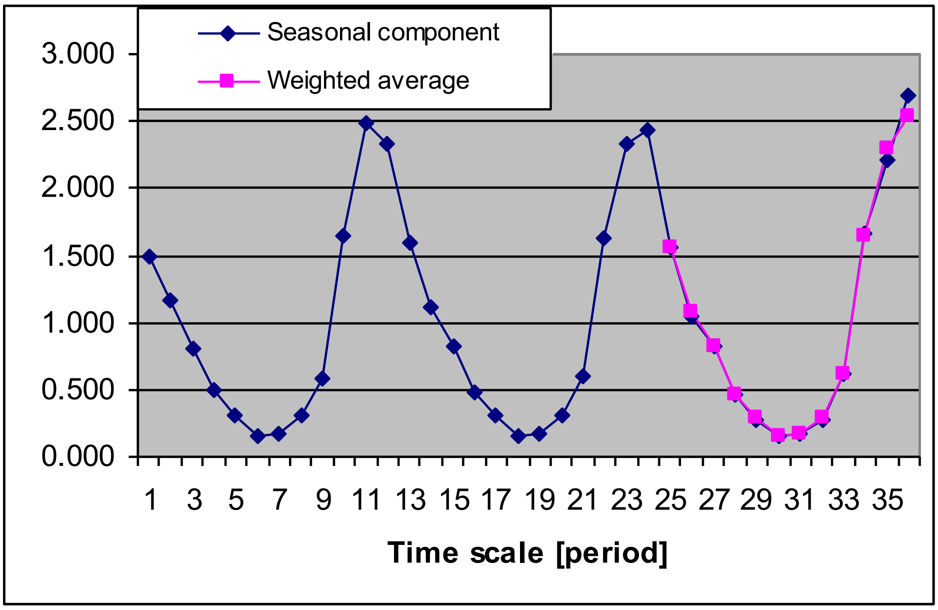
The initial seasonal components S'_t can then be determined by the ratio of the actual demand Y_t over the deseasonalised one $\tilde{Y}_t$:

$$S'_t = \frac{Y_t}{\tilde{Y}_t}$$

In the case of the WinTech Zenith Z5 family, three cycles are available so that three sets of seasonal components can be calculated. Weighted averages can then be chosen as initial values for the seasonal components. The next table provides the values of the three sets *S(1), S(2), S(3)* and the chosen initial values *S'* obtained by a weighted average for the WinTech Zenith Z5 family. These data are also provided graphically in the next figure.

	$\boxed{S_t(1)}$	$\boxed{S_t(2)}$	$\boxed{S_t(3)}$	$\boxed{S'_t}$
weight	0.15	0.35	0.5	
t				
1	1.50	1.60	1.55	**1.56**
2	1.16	1.11	1.04	**1.08**
3	0.80	0.83	0.83	**0.82**
4	0.49	0.48	0.46	**0.47**
5	0.31	0.30	0.27	**0.29**
6	0.16	0.15	0.15	**0.15**
7	0.17	0.18	0.18	**0.18**
8	0.30	0.30	0.27	**0.29**
9	0.59	0.60	0.62	**0.61**
10	1.64	1.63	1.66	**1.65**
11	2.48	2.34	2.22	**2.30**
12	2.32	2.44	2.69	**2.54**

Seasonal components for WinTech Z5 family



Seasonal components for WinTech Z5 family

With the initial trend and seasonal components being calculated, only the initial base remains to be determined to obtain the final forecasting model. This latter will be validated by using the first one or two available data cycles to initiate the base and the last cycle to validate the model as illustrated by the next figure. The last cycle used for the model validation will be called the "validation cycle''.

cycle 1	**cycle 2**	**cycle 3 'validation cycle"**
used to initiate trend and seasonal components		
	used to initiate the base	
		used to validate the model

Initiation and validation process

It could be argued that using the third cycle for initialisation as well as for validation is incorrect. Although this is formally true, it can be counter argued that using more data for the determination of the initial trend and seasonal components can only lead to an improvement of the model reliability.

At this point, the procedure calls for the five steps described in the next figure:

1) Initialisation for validation cycle
2) Initial forecasting of validation cycle
3) Model validation on validation cycle
4) Determination of the smoothing constants for further forecasting (after the initial one)
5) Initialisation for future horizon (after validation cycle)

Procedure for model validation and calibration

3.7.6 Initialisation for, and initial forecasting of the validation cycle

Recall that for the **[a;m]** case, the forecasting model is:

$$F_{t+h} = (B_t + hT_t) \times S_{t+h-c}$$

The model initialisation for the validation cycle therefore requires the determination of the initial value for the base B'_v , which is the value of the base of the last period of the previous cycle. For simplicity, it is assumed here that the validation cycle is the third cycle of the available historical data and consequently that the previous cycle is cycle 2. The B'_v can be computed from the average value of cycle 2 $\overline{Y}_2$ and the initial trend T'

$$B'_v = \overline{Y}_2 + \frac{c-1}{2}T'$$

In the case of the WinTech Zenith Z5.family, this gives:

$$B_v = B'_3 = B_{24} = \overline{Y}_2 + \frac{c-1}{2}T' = 40243 + 5.5 \times 166 = 41156$$

The forecast can then be fully initialised according to the **[a;m]** model:

$F'_{t+h} = (B' + hT') \times S'_{t+h}$ with $h \in \mathrm{N}$

It is very important to understand that this provides the initial forecast F' over any horizon $\mathcal{H}$. Thus, the model is used to forecast $\mathcal{H}$ values at once and not to forecast one single value after the other $\mathcal{H}$ times in a row.

For the case of the WinTech Zenith Z5 family, the initial forecasting model for cycle 3 (validation cycle) is:

$$F'_{t+h} = (41156 + h \times 166) \times S'_{t+h}$$

3.7.7 Model validation on validation cycle

After determining the initial model for the validation cycle, the next step consists in its validation. As indicated above, the historical data of the third cycle will be used for this purpose. The model is initialised for the first period of the third cycle and the forecasting horizon is set to one cycle (the third one). The validation is estimated by computing the mean absolute percentage forecasting error $MAPE'$ over the validation cycle.

In the case of the WinTech Zenith Z5 family, the third cycle corresponds to months 25 to 36. Thus the initial forecasted values F'_t must be compared to the actual historical values Y_t using the following relations:

$$F'_{24+h} = (41156 + h \times 166) \times S'_h$$

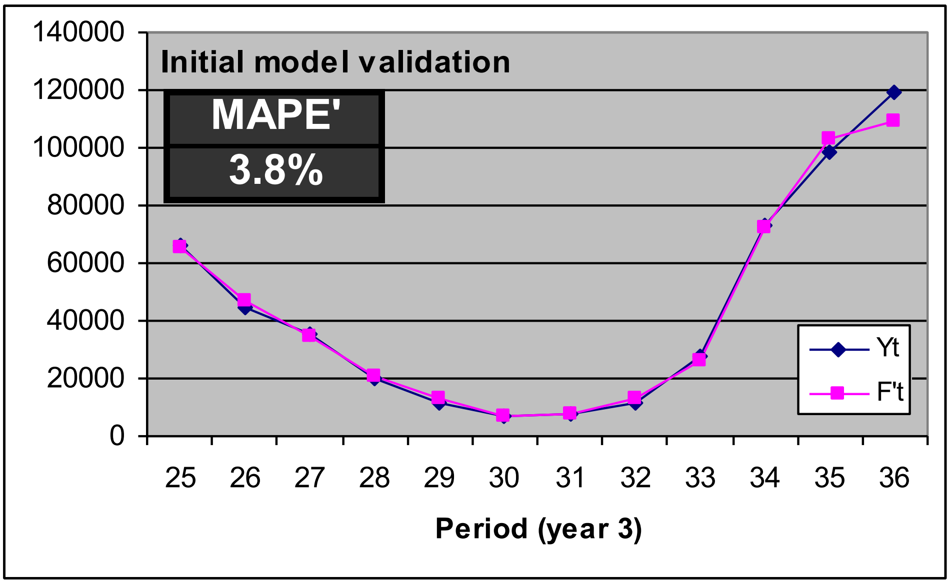
with $h \in [1;12]$

$$MAPE' = \frac{1}{12}\sum_{h=1}^{12}\left|\frac{Y_{24+h} - F'_{24+h}}{Y_{24+h}}\right|$$

The results for the WinTech Zenith Z5 family are shown in the next table and figure.

t	Y(t)	F'(t)	APE' (t)	MAPE'
25	66260	65631	0.9%	3.8%
26	44510	46908	5.4%	
27	35500	34902	1.7%	
28	19840	20399	2.8%	
29	11855	12715	7.3%	
30	6710	6675	0.5%	
31	7750	7629	1.6%	
32	11920	12697	6.5%	
33	27440	26408	3.8%	
34	73360	72035	1.8%	
35	98290	103150	4.9%	
36	119500	109545	8.3%	

Initial model validation



Initial model validation

These results demonstrate that the model is reliable enough.

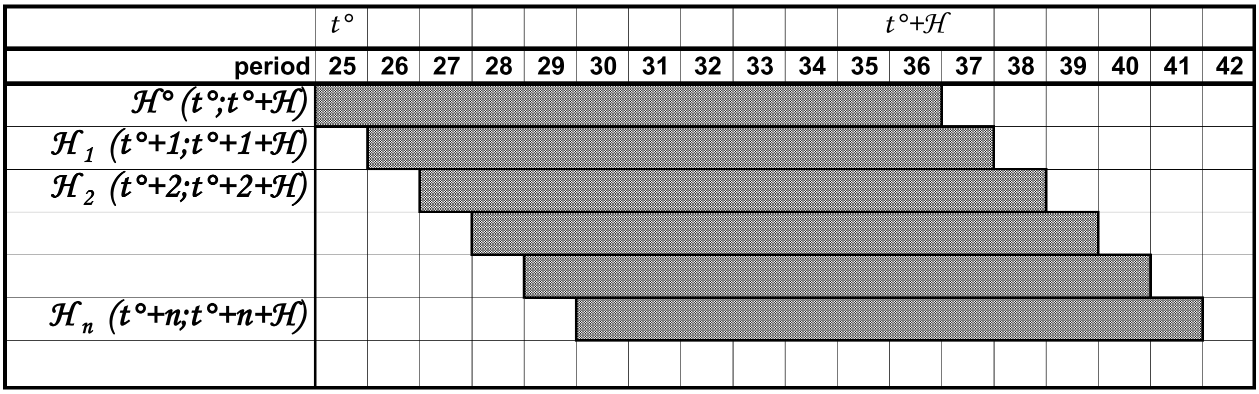
3.7.8 Determination of the smoothing constants for further forecasting

The initial forecast model developed so far allows computing a first (initial) forecast over a chosen horizon $\mathcal{H}^\circ\ (t^\circ;t^\circ+\mathcal{H})$.

Later forecasts will be successively shifted by one period and, assuming that the forecast horizon remains unchanged (which is not necessary the case), these forecast horizons will be:

$\mathcal{H}1\ (t^\circ+1;t^\circ+1+\mathcal{H})\ ;\ \mathcal{H}2\ (t^\circ+2;t^\circ+2+\mathcal{H})\ ;\ \ldots;\ \mathcal{H}n\ (t^\circ+n;t^\circ+n+\mathcal{H})$

This situation is illustrated by the next figure.



Shifting of the forecasting horizons

For all forecasts following the initial one, the model parameters, B_t, T_t and S_{t+h-c} will have to be adapted using exponential smoothing, according to the following equations for the **[a;m]** case:

$$B_t = \alpha \frac{Y_t}{S_{t-c}} + (1-\alpha)(B_{t-1} + T_{t-1})$$

$$T_t = \beta(B_t - B_{t-1}) + (1-\beta)T_{t-1}$$

$$S_t = \gamma \frac{Y_t}{B_t} + (1-\gamma)S_{t-c}$$

The values of the three smoothing constants have of course an influence on forecast reliability. Therefore, it is important to determine an optimal set of the smoothing constants α,β,γ.

This can be done by using the previous equations for the calculation of forecasts over a horizon of the historical data and finding the best smoothing constant set by minimising the forecast error.

To do so, let $\mathcal{H}_v$ $(t_v;t_v+(c-1))$ be the forecast horizon of the validation cycle of length c. The procedure calls for the following steps:

1) Initiate the forecast over a horizon $\mathcal{H}$ at a period t_v-n, with $n \in [3;c/2]$

2) Compute the initial forecast for the horizon $\mathcal{H}°$ $(t_v-n;t_v-n+(c-1))$ as indicated in 3.7.6 using the relation:

$$F'_{t+h} = (B' + hT') \times S'_{t+h-c} \text{ with } h \in [1;c]$$

3) Choose a set of values α_j, β_k with α, $\beta \in [0;1]$

4) Compute a new forecast over the horizon $\mathcal{H}_1$ $(t_v-n+1;t_v-n+1+(c-1))$ by adjusting the initial model parameters using the following relations:

$$B_t = \alpha_j \frac{Y_t}{S_{t-c}} + (1-\alpha_j)(B_{t-1} + T_{t-1})$$

$$T_t = \beta_k (B_t - B_{t-1}) + (1-\beta_k) T_{t-1}$$

$$F_{t+h} = (B_t + hT_t) \times S'_{t+h-c}$$

5) Compute the forecast error $MAPE_1(j,k)$

6) Repeat steps 4 to 5 with n other horizons $\mathcal{H}_q\ (t_v\text{-}n\text{+}q;t_v\text{-}n\text{+}q\text{+}(c\text{-}1))$ with q=[2;n]

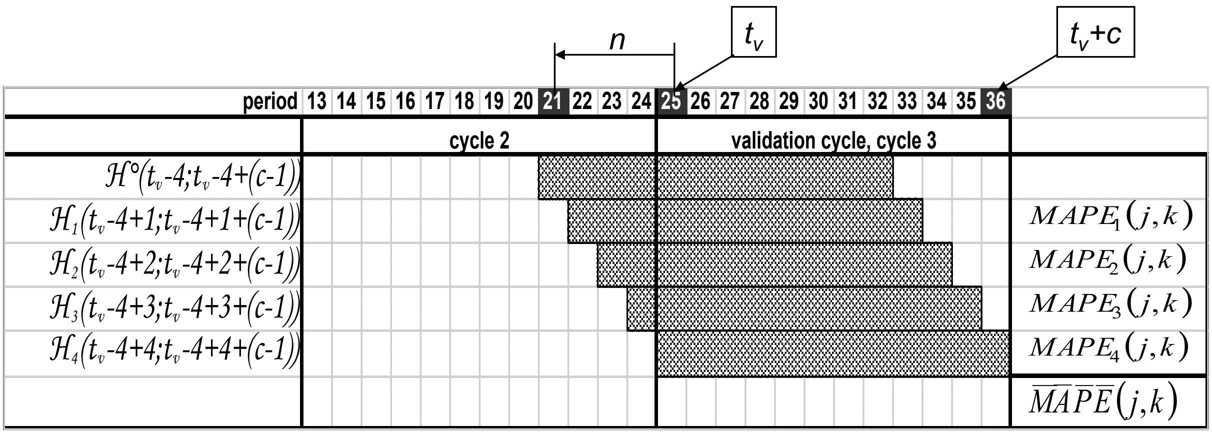
7) Compute the average MAPE of the n forecasted horizons $\overline{MAPE}(j,k)$

8) Repeat steps 3 to 7 with several other sets of values α_j, β_k in order to cover the variable space $\alpha_j, \beta_k \in [0;1]$

9) Plot the function $\overline{MAPE}(j,k) = f(\alpha_j, \beta_k)$

10) Select the set $(\alpha,\beta)_0$ that minimises $\overline{MAPE}(j,k)$

The procedure is schematically illustrated in the next figure for the specific case of *n*=4 and *c*=12

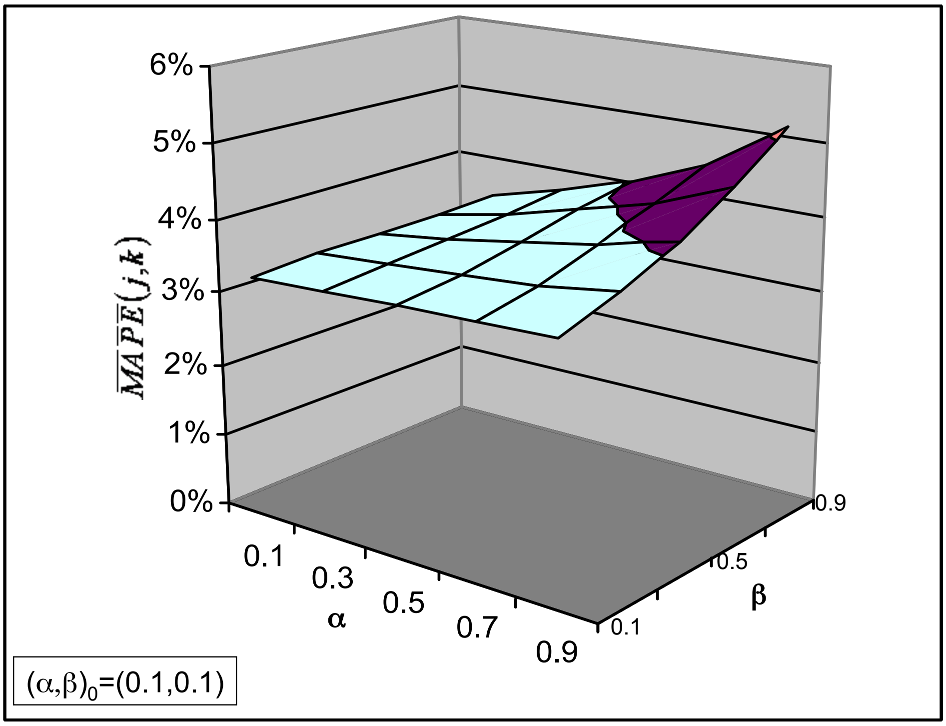


Optimal smoothing constant determination; for *n*=4 & *c*=12

Note that the smoothing constant γ for the seasonal component is not taken into account in the previous procedure. This is justified by the fact that in many actual cases, the seasonal cycle is one year. This means that an exponential smoothing requires at least 2 previous years to be significant. However, in reality, the environment changes so much in 3 years that an optimisation of γ does not make sense. To combat this, it is recommended using a fixed high value (for example γ=0.9) in order to give a predominant weigh to the most recent value of the seasonal components.

However, if the periodicity of the seasonal effect is much shorter (say one week) a similar procedure for the optimisation of γ can be used.

In the case of the WinTech Zenith Z5 family, the application of the previous procedure leads to the function $\overline{MA}\overline{P}\overline{E}(j,k) = f(\alpha_j, \beta_k)$ given in the next figure and to an optimal set $(\alpha,\beta)_0 = (0.1, 0.1)$ corresponding to a $\overline{MA}\overline{P}\overline{E}(j,k) = 3.33\%$



Optimal smoothing constant determination for the WinTech Zenith Z5.family

With the determination of the optimal values $(\alpha,\beta)_0$, the forecasting model is fully elaborated.

3.7.9 Final forecasting model

To summarise, the final forecasting model is composed of two sub-models:

- The initialisation model for the very first forecast over a horizon $\mathcal{H}$;
- The running model for all later forecasts.

Initialisation model

$$F'_{t+h} = (B' + hT') \times S'_{t+h-c}$$

Running model

$$B_t = \alpha \frac{Y_t}{S_{t-c}} + (1-\alpha)(B_{t-1} + T_{t-1})$$

$$T_t = \beta(B_t - B_{t-1}) + (1-\beta)T_{t-1}$$

With $(\alpha,\beta)=(\alpha,\beta)_0$

$$F_{t+h}=(B_t+hT_t)\times S'_{t+h-c}$$

After one full cycle, S'_t should be replaced by:

$$S_t=\gamma\frac{Y_t}{B_t}+(1-\gamma)S_{t-c}$$

With $\gamma=0.9$ if the seasonal cycle is one year or more.

The results obtained for the case of WinTech Zenith Z5 family are shown in the following table and figure.

Period	**Y_t**	**F'_t**	**F_t**	B'_t	B_t	T'_t	T_t	**Period**	S'_t	S_t
25	**66260**									
26	**44510**									
27	**35500**									
28	**19840**									
29	**11855**									
30	**6710**									
31	**7750**									
32	**11920**									
33	**27440**									
34	**73360**									
35	**98290**									
36	**119500**			44492		166				
37	70000	**69659**			44680		168	**25**	1.56	
38		**48483**	**48510**					**26**	1.08	
39		**36994**	**37015**					**27**	0.82	
40		**21231**	**21244**					**28**	0.47	
41		**13116**	**13124**					**29**	0.29	
42		**6969**	**6975**					**30**	0.15	
43		**8089**	**8095**					**31**	0.18	
44		**13144**	**13154**					**32**	0.29	
45		**28094**	**28118**					**33**	0.61	
46		**76043**	**76111**					**34**	1.65	
47		**106519**	**106620**					**35**	2.30	
48		**118294**	**118411**					**36**	2.54	
49			**73133**					**37**		1.57

Computation examples:

1.- Initial forecast computation with the Initial model

General equation: $F'_{t+h}=(B'+hT')\times S'_{t+h-c}$

Computation example for *F'₃₇*: *t=36, h=1* →

$$F'_{37}=(B'+T')\times S'_{25}=(44492+166)\times 1.56=69659$$

Computation example for *F'₄₇*: *t=36, h=11* →

$$F'_{47}=(B'+11\times T')\times S'_{35}=(44492+11\times 166)\times 2.30=106519$$

2.- Further forecast computation with the Running model

General equation: $F_{t+h}=(B_t+hT_t)\times S'_{t+h-c}$

2.1. Running model computation using exponential smoothing (α=0.1; β=0.1; γ=0.9)

$$B_t=\alpha\frac{Y_t}{S_{t-c}}+(1-\alpha)(B_{t-1}+T_{t-1})$$

For t=37, α=01:

$$B_{37} = \alpha \frac{Y_{37}}{S_{25}} + (1-\alpha)(B_{36} + T_{36}) = 0.1 \frac{70000}{1.56} + 0.9(44492 + 166) = 44680$$

$$T_t = \beta(B_t - B_{t-1}) + (1-\beta)T_{t-1}$$

For t=37, β=01:

$$T_{37} = \beta(B_{37} - B_{36}) + (1-\beta)T_{36} = 0.1(44680 - 44492) + 0.9 \times 166 = 168$$

2.2. Forecast computation using the Running model

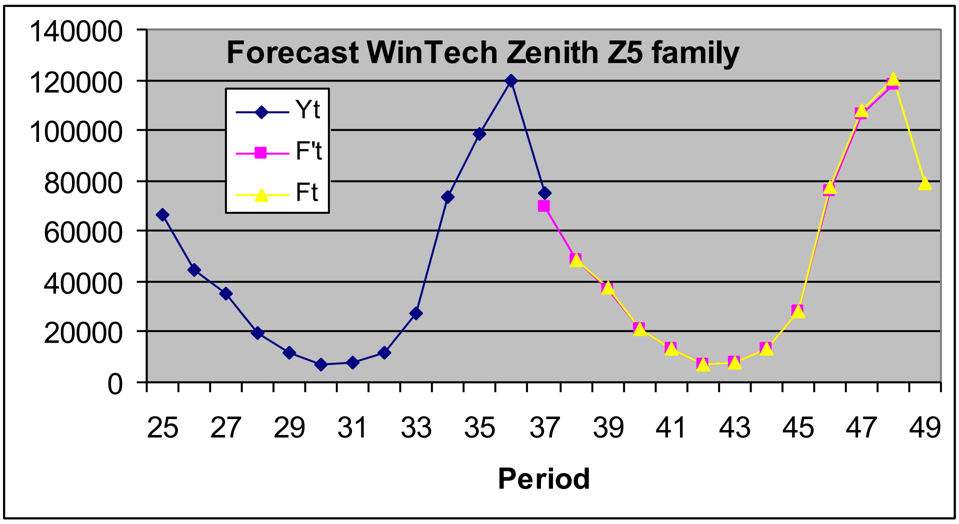
General equation: $F_{t+h} = (B_t + hT_t) \times S'_{t+h-c}$

Computation example for F_{38}: *t=37, h=1* →

$$F_{38} = (B_{37} + T_{37}) \times S'_{26} = (44680 + 168) \times 1.08 = 48510$$

Computation example for F_{48}: *t=37, h=11* →

$$F_{48} = (B_{37} + 11 \times T_{37}) \times S'_{36} = (44680 + 11 \times 168) \times 2.54 = 118411$$



Initial, F'_t, and first running, F_t, forecast for the WinTech Zenith Z5 family

It can be observed that the actual value Y_t for period 37 is higher than the value F'_t of the initial forecast. This leads to a slight modification of the values obtained through the first running forecast, F_t, due to the adaptation of the model resulting from the exponential smoothing.

3.8 Implementation particularities

The implementation of a forecasting procedure in real cases requires taking several aspects into account. This section describes and discusses some of them.

3.8.1 Contextual information and anticipated events

As previously indicated, all the forecasting mathematical models rest on the hypothesis that the established patterns and/or relationships in the historical data will not change during the forecasting horizon. Nevertheless, changes in the data can and most likely will occur and they must be detected as early as possible to avoid large and usually costly forecasting errors.

For example, let's consider the telecom market crisis in the early 2000's. In 1999, companies analyzing this market sector expected significant growth in demand and consequently, considerable development for the following years. Many of these companies decided to expand or modify their plants in order to increase capacity to satisfy the expected increase in demand. Contrary to expectations, demand for products in the telecom industry was much lower due to several factors such as decreased demand in the computer and cell phone industries. The low demand levels hurt several companies that had anticipated higher demand and had invested in unneeded resources.

This example demonstrates that forecasts can induce significant error even if they are based on sophisticated mathematical models. The key to anticipating demand is to take into account all information that can affect the demand pattern. The idea can be reflected by the saying:

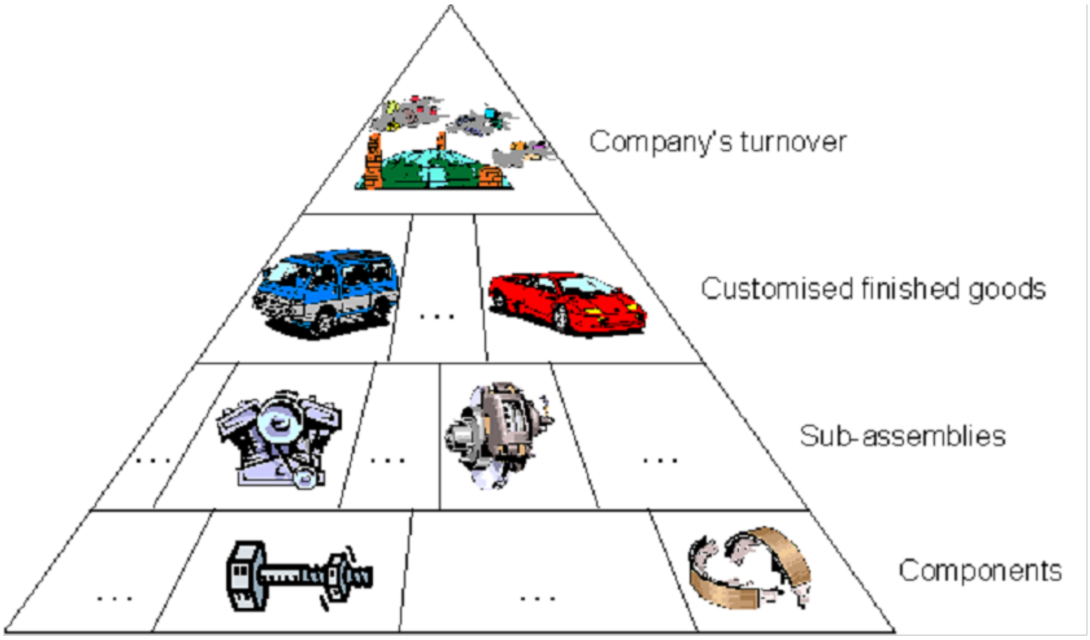
"You cannot drive a car by only looking in the rear-view mirror".

This can be done by integrating mathematical forecast with judgemental ones. This very interesting approach is beyond the scope of this course; however, the reader can consult the following reference for further information on this topic: Séverine. Meunier Martins, Approche stratégique pour la prévision de données de gestion: méthode ciblée intégrant des éléments subjectifs et mathématiques; thèse N° 2749 (2003).

3.8.2 Aggregation and product structure levels

Forecasts are used for a variety of business decisions: purchasing management, production planning, preparation of budgets, etc. Therefore, the question arises: what kind of forecasts does the company have to prepare? Is it the estimation of the needs for specific components for the next months, the sales forecasts for the product XYZ, the estimated turnover for the German market for the next year, etc?

Forecasts may be established at different levels, forming a pyramid in which each stage represents a specific aggregation level as illustrated by the following figure.



Pyramid of the forecast levels

Forecasts can be established for each different level of the pyramid in an independent way (for example the marketing manager will make forecasts concerning the development of a product family, purchasers will estimate the need for some components, etc.) or derived from the forecasts established by a neighbour aggregation level. Making forecasts from a lower level to an upper one is not difficult: it simply consists of summing the forecasts. On the other hand, the ventilation process that uses forecasts from a higher level in order to establish forecasts for a lower one is more difficult and different possibilities exist to estimate the weight of the different forecast components:

$$F_{n,t} = \omega_{n,t} F_t^{ag}$$

$F_{n,t}$ Forecast for the element *n* for the period *t* derived from the forecast of the higher aggregation level F_t^{ag}

$\omega_{n,t}$ Weighting coefficient for element *n* and for period *t*

F_t^{ag} Forecast for period *t* established at a higher aggregation level, containing element *n*

Obviously, the determination of $\omega_{n,t}$ will play a major role. Several approaches can be used for this purpose such as:

$\omega_{n,t} = \omega_{n,t-c}$; weight of the same element, in the same period, one season before. In this approach the seasonality is taken into account, but there is a risk that random effects could induce a bias into the forecasts.

$\omega_{n,t} = \frac{1}{c}\left(\omega_{n,t-1} + \omega_{n,t-2} + \ldots + \omega_{n,t-c}\right)$ average of the previous periodic contributions of the element to the considered aggregated level, taken during the last cycle. This approach avoids the risk of bias of the previous one, but does not take into account a possible recent trend:

$\omega_{n,t} = \alpha \frac{F_{n,t-1}}{F_{t-1}^{ag}} + (1-\alpha)\omega_{n,t-1}$ smoothed average of the previous periodic contributions of the element to the considered aggregated level. In this approach the most recent contributions are given more weight in the average process, thanks to an exponential smoothing constant α with a coefficient equal to or larger than 0.3.

It is a well-known phenomenon that aggregated forecasts (for example product family forecasts) are more accurate than detailed forecasts. The more aggregated the data are, the more reliable and easier the forecasts will be. Nevertheless, if the forecasts are established at a high aggregation level (for example for the product families) and if the behaviour of the lower level (for example products) time series are heterogeneous, the different trends at the lower level will not appear and the resulting detailed forecasts will be biased.

3.8.3 Improving forecasts with firm orders

Forecasts are generally considered as one type of demand input and firm orders as another, independent one. However, if the forecasts are correctly elaborated, they should of course be closely related to firm orders. It is therefore logical to consider that available information about firm orders could be used to improve forecast reliability.

Most companies receive firm orders from customers with a delivery delay greater than the company's planning time period (i.e. several periods in advance). In this situation, at any point in time, part of the demand for a time period in the future is known with certainty (firm orders).

In the following paragraphs, different approaches to integrate firm orders into forecasts are presented.

F_t = Forecast for period *t*

Y_t = actual total demand for period *t*

$\tilde{Y}_t^k$ = partial demand for period *t* (firm orders) known at *t-k* (*k* periods before *t*)

Approach 1: Forecasts = (Inflating factor) X (Known partial demand)

$$F_t = \Delta_t^k \times \tilde{Y}_t^k$$

Δ_t^k is an estimation of the inflating factor. It may be determined with an exponential smoothing model:

$\Delta_t^k = \alpha \dfrac{F_{t'}}{\tilde{Y}_{t'}^k} + (1-\alpha)\Delta_{t'}^k$ where *t'* is the closest period (before *t*) at which the partial demand was observed.

Approach 2: Forecasts = (Additive element) + (Known partial demand)

$$F_t = \Omega_t^k + \tilde{Y}_t^k$$

Ω_t^k is an estimation of the additive element. It may be determined with an exponential smoothing model:

$\Omega_t^k = \alpha\left(F_{t'} - \tilde{Y}_{t'}^k\right) + (1-\alpha)\Omega_{t'}^k$ where *t'* is the closest period (before *t*) at which the partial demand was observed.

Approach 3: Forecasts = (Inflating factor) X (Known partial demand) + (Additive element) + (Known partial demand)

This approach is a weighted combination of the two previous ones.

$$F_t = a_t^k(\Delta_t^k \times \tilde{Y}_t^k) + b_t^k(\Omega_t^k + \tilde{Y}_t^k)$$

$$F_t = b_t^k \Omega_t^k + (a_t^k \Delta_t^k + b_t^k)\tilde{Y}_t^k$$

$$F_t = \Phi_t^k + \Psi_t^k \tilde{Y}_t^k$$

The weighting coefficient Φ_t^k and Ψ_t^k are determined with a least square estimation based on past values.

3.8.4 Typological analysis of products and components

In today's competitive markets, it is not uncommon to observe companies of even medium size offering customers hundreds or thousands of different products. This multiple offer considerably increases the number of forecasts to establish and manage. Faced with the heavy burden of dealing with a large number of forecasts, companies must classify the items for which forecasts are required (sub-assemblies, products, families, etc.) in order to identify the most critical ones and to prioritise them in the forecasting process. Companies will then be able to apply different forecast policies according to their classification. For example, a subjective revision of mathematical forecasts could be applied to the most critical items whereas the less critical ones could be managed by an automated process. This strategic allocation of resources is essential as they are limited and should therefore be invested where they are most beneficial for the company.

The definition of priorities consists in identifying the critical items for the company and those requiring particular attention in the forecasting process. Two classes of items can be defined:

1. The high priority class that groups the items for which the establishment of reliable forecasts is essential to the company (for operational, tactical or strategic reasons);
2. The low priority class containing the items for which the quality of the forecasts has a lower influence on company performances. Meunier et al. have proposed a structured methodology formalising this approach. It is briefly described below.

For high priority items, to improve the reliability of the forecasts, the forecaster can review the mathematical forecasts and adjust them based on his/her experience, his/her knowledge of contextual information, etc...,. For low priority items, forecasts can be purely mathematical. The integration of the forecaster's judgement can be decided depending upon:

- The importance of the item for which forecasts are established;
- The difficulty to reliably establish forecasts (judgemental or mathematical ones);
- The influence of their accuracy on the logistic performances of the company.

Accordingly, different types of criteria can be taken into account in the classification process.

The criteria related to the strategic importance of the products for the company are the contribution of the product to total turnover and its position within the life cycle. The contribution of the products to total turnover helps identify the products making the greatest contribution to financial volume. The position of the products within the life cycle is a subjective criterion, complementary to the previous one. It makes it possible to identify the products with a strong growth potential or in contrast approaching the end of life. This information does not clearly appear when evaluating the contribution to total turnover.

The criteria related to the difficulty in establishing reliable judgemental or mathematical forecasts are respectively, the availability of useful contextual information and the erratic character of the time series related to the historical demand of the products. Several studies show that the availability of contextual information is a crucial condition for the forecaster to establish reliable forecasts. When using mathematical models, the more the data of the time series is erratic, the more difficult it is to establish reliable and accurate forecasts. The first criterion (availability of contextual information) is subjective whereas the second one (randomness of time series) can be evaluated more objectively with statistical tests.

The criteria related to the correlation of the accuracy of forecasts with the logistic performance of the company are the lead times of the components of the products and their commonality, i.e. the ratio between the number of products including the component and the total number of products. Although many other criteria could be chosen concerning the correlation of the accuracy of forecasts with the logistic performances, Meunier et al. justify this choice by the fact that:

- Products with components having long lead times need reliable long term forecasts;
- Products containing specific components require more accurate forecasts than the others.

ABC classification is the most popular and used technique for the classification of products in inventory management and could be applied to identify the most critical items in the forecasting process. The enormous popularity of this approach is explained by its simplicity but nevertheless, it suffers from a major drawback of considering only one criterion (usually the contribution to turnover is taken into account), which could limit its efficiency in certain situations.

A derived approach to the ABC classification can also be imagined. It consists of using a matrix to integrate two criteria into the classification process, for example the annual turnover and the difficulty in establishing forecasts. Nine classes of items are determined (AA, AB, …, CC), requiring nine different policies. This procedure becomes more difficult to manage for more than two criteria.

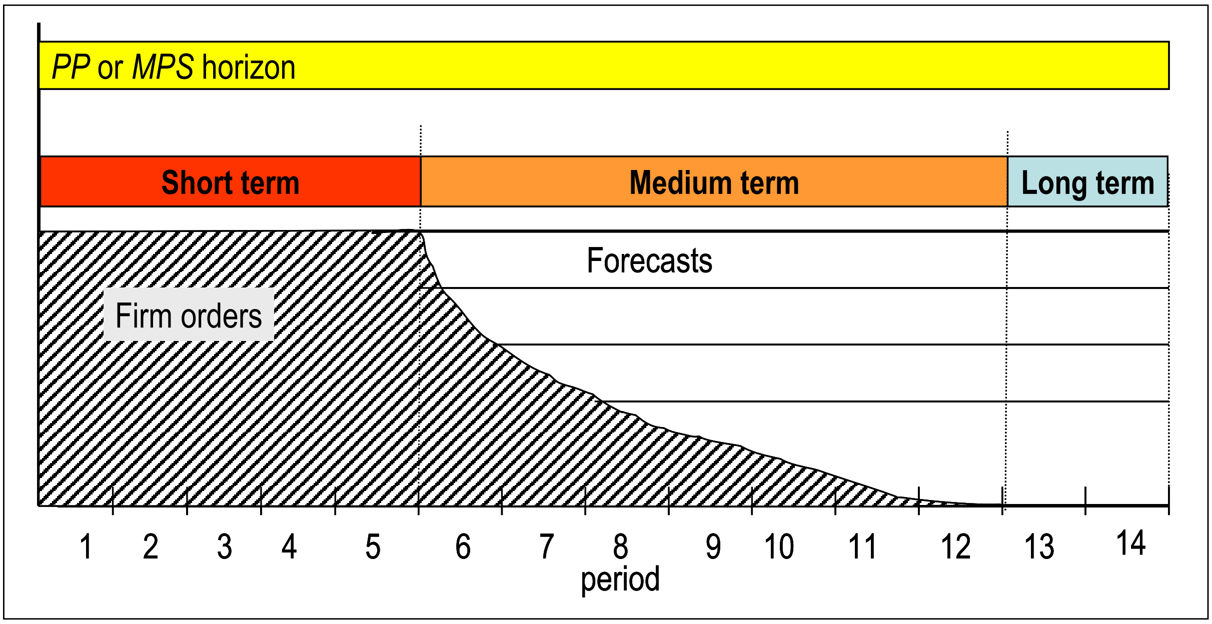
With traditional classification methods, each item is assigned to a precise class. The membership to a class is binary: an item does or does not belong to a class. In reality, there is generally no absolute distinction between classes; they often have overlapping characteristics. Fuzzy classification is appropriate for items characterised by data that can not be clearly assigned to discrete classes. Furthermore, fuzzy classification makes it possible to consider vaguely defined linguistic criteria such as the availability of contextual information and the position of the product within the life cycle.

Many other classification methods exist and the interested readers are kindly asked to refer to the existing literature concerning this topic.

3.8.5 Combining forecast and order book for demand determination

The final goal of demand management and forecasting is to build a reliable *Production Plan* or *Master Production Schedule* that will pilot the whole *Value Adding Chain* of the firm. The procedure for the construction of these plans is the subject of Chap. 4. However, a few general remarks are necessary here.

The *Production Plan* usually covers a fairly long planning horizon; typically 18 months. For the latest part of the planning horizon, long term, no firm orders are available. Consequently, demand is solely based on forecasts. For the immediate part of the planning horizon, short term, firm orders are certainly available in the *Order Book*. Thus, as time goes by, parts of the forecasts are regularly replaced (or materialised) by firm orders as schematically illustrated in the next figure.



Combining forecasts and firm orders in the *Production Plan*

The question is then, which forecast to replace by a firm order? This is extremely difficult as forecasts are generally established on an aggregated level (for example product family) while firm orders are on the *finished product* level. Furthermore, a new firm order Y_t^k of quantity *Q,* of product *k,* in period *t,* may correspond to a forecast $F_{t'}^{k'}$ of quantity *Q',* of product *k'* in period *t.* Thus, the establishment of a *Production Plan* or a *Master Production Schedule* is both a very critical and challenging task. In practice, firms have developed their own tricks, based on experience that do a reasonable job, but these tricks do not really solve the issue. This topic will be discussed further in Chap. 4.

3.9 Table of contents