

Chapter 2

Forecasting

There are two main reasons why an inventory control system needs to order items some time before customers demand them. First, there is nearly always a *lead-time* between the ordering time and the delivery time. Second, due to certain ordering costs, it is often necessary to order in *batches* instead of unit for unit. These two reasons mean that we need to look ahead and forecast the future demand. A demand forecast is an estimated average of the demand size over some future period. But it is not enough to estimate the average demand. We also need to determine how uncertain the forecast is. If the forecast is more uncertain, a larger safety stock is required. Consequently, it is also necessary to estimate the forecast error, which may be represented by the standard deviation or the so-called Mean Absolute Deviation (*MAD*).

2.1 Objectives and Approaches

In this chapter we shall focus on forecasting methods that are suitable in connection with inventory control. Typical for such forecasts is that they concern a relatively short time horizon. Very seldom do we need to look ahead more than one year. Other types of forecasts may very well concern many years. In general, there are then two types of approaches that may be of interest:

- Extrapolation of historical data
When extrapolating historical data, the forecast is based on previous demand data. The available techniques are grounded in statistical methods for analysis of time series. Such techniques are easy to apply and use in computerized inventory control systems. It is no problem to regularly update forecasts for thousands of items, which is a common requirement in connection with practical inventory control. Extrapolation of historical data is the most common and important approach to obtain forecasts over a short horizon, and we shall devote the main part of this chapter to such techniques. The idea is to identify a simple demand pattern that can be assumed to describe the demand process. This demand model usually includes some stochastic element.

- Forecasts based on other factors

It is very common that the demand for an item depends on the demand for some other items. Consider, for example, an item that is used exclusively as a component when assembling some final products. It is then often natural to first forecast the demand for these final products, for example by extrapolation of historical data. Next we determine a production plan for the products. The demand for the considered component is then obtained directly from the production plan. This technique to “forecast” demand for dependent items is used in *Material Requirements Planning (MRP)* that is dealt with in Sect. 8.2.4.

But there are also other factors that might be reasonable to consider when forecasting demand. Assume, for example, that a sales campaign is just about to start or that a competing product is introduced on the market. Clearly this can mean that historical data is no longer representative when looking ahead. It is normally difficult to take such factors into account in the forecasting module of a computerized inventory control system. It is therefore usually more practical to adjust the forecast manually in case of such special events.

It is also possible, at least in principle, to use other types of dependencies. A forecast for the demand of ice cream can be based on the weather forecast. Consider, as another example, forecasting of the demand for a spare part that is used as a component in certain machines. The demand for the spare part can be expected to increase when the machines containing the part as a component are getting old. It is therefore reasonable to look for dependencies between the demand for the spare part and previous sales of the machines. As another example we can assume that the demand during a certain month will increase with the advertising expenditure the previous month. Such dependencies could be determined from historical data by regression analysis. (See Sect. 2.7.) Direct applications of such techniques are, however, very limited.

2.2 Demand Models

Extrapolation of historical data is, as mentioned, the most common approach when forecasting demand in connection with inventory control. To determine a suitable technique, we need to have some idea of how to model the stochastic demand. In principle, we could try to determine the model from analysis of historical data. In practice this is very seldom done. With many thousands of items, this initial work does not seem to be worth the effort in many situations. In other situations there are not enough historical data. A model for the demand structure is instead usually determined intuitively. In general, the assumptions are very simple.

2.2.1 Constant Model

The simplest possible model means that the demands in different periods are represented by independent random deviations from an average that is assumed to be relatively stable over time compared to the random deviations. Let us introduce the notation:

x_t = demand in period t

a = average demand per period (assumed to vary slowly)

ε_t = independent random deviation with mean zero

A constant model means that we assume that the demand in period t can be represented as

$$x_t = a + \varepsilon_t. \quad (2.1)$$

Many products can be represented well by a constant model, especially products that are in a mature stage of a product life cycle and are used regularly. Examples are consumer products like toothpaste, many standard tools, and various spare parts. In fact, if we do not expect a trend or a seasonal pattern, it is in most cases reasonable to assume a constant model.

Let us assume that the stochastic variables ε_t are independent and have the same distribution. Assume also that the parameter a is a constant that we do not know. The natural estimate of a is then the average of all previous demands x_t . However, if a is not constant and instead is varying slowly it should reasonably be better to use only a number of the latest demands. This is the basic idea both when using *Moving average* (Sect. 2.3) and *Exponential smoothing* (Sect. 2.4).

2.2.2 Trend Model

If the demand can be assumed to increase or decrease systematically, it is possible to extend the model by also considering a linear trend. Let

a = average demand in period 0

b = trend, that is the systematic increase or decrease per period (assumed to vary slowly)

A trend model means that the demand is modeled as:

$$x_t = a + bt + \varepsilon_t. \quad (2.2)$$

During a product life cycle there is an initial growth stage and a phase-out stage at the end of the cycle. During these stages it is natural to assume that the demand follows a trend model with a positive trend in the growth stage and a negative trend in the phase-out stage.

2.2.3 *Trend-seasonal Model*

Let

F_t = seasonal index in period t (assumed to vary slowly)

If, for example, $F_t = 1.2$, this means that the demand in period t is expected to be 20 % higher due to seasonal variations. If there are T periods in 1 year, we must require that for any T consecutive periods $\sum_{k=1}^T F_{t+k} = T$. When using a multiplicative trend-seasonal demand model it is assumed that the demand can be expressed as

$$x_t = (a + bt)F_t + \varepsilon_t. \quad (2.3)$$

By setting $b = 0$ in (2.3) we obtain a constant-seasonal model.

In (2.3) it is assumed that the seasonal variations increase and decrease proportionally with increases and decreases in the level of the demand series. In most cases this is a reasonable assumption. An alternative assumption could be that the seasonal variations are additive.

Many products have seasonal demand variations. For example the demand for ice cream is much larger during the summer than in the winter. Some products, like various Christmas decorations, are only sold during a very short period of the year. Still, the number of items with seasonal demand variations is usually very small compared to the total number of items. A seasonal model is only meaningful if the demand follows essentially the same pattern year after year.

2.2.4 *Choosing Demand Model*

When looking at the three demand models considered, it is obvious that (2.2) is more general than (2.1), and that (2.3) is more general than (2.2). It may then appear that it should be most advantageous to use the most general model (2.3). This is, however, not true. A more general demand model covers a wider class of demands, but on the other hand, we need to estimate more parameters. Especially if the independent deviations are large, it may be very difficult to determine accurate estimates of the parameters, and it can therefore be much more efficient to use a simple demand model with few parameters. A more general model should be avoided unless there is some evidence that the generality will give certain advantages.

It is important to understand that the independent deviations ε_t cannot be forecasted, or in other words, the best forecast for ε_t is always zero. Consequently, if the independent deviations are large there is no possibility to avoid large forecast errors. Consider the constant model (2.1). It is obvious that the best forecast is simply our best estimate of a . In (2.2) the best forecast for the demand in period t is similarly our best estimate of $a + bt$, and in (2.3) our best forecast is the estimate of $(a + bt)F_t$.

In some situations it may be interesting to use more general demand models than (2.1)–(2.3). (See Sect. 2.9.) This would, however, require a detailed statistical

analysis of the demand structure. In practice this is rarely done in connection with inventory control.

One practical problem is that it is quite often difficult to measure demand, since only sales are recorded. If historical sales, instead of historical demands, are used for forecasting demand, considerable errors may occur in situations where a relatively large portion of the total demand is lost due to shortages. (See Sect. 2.10.5.)

2.3 Moving Average

Assume that the underlying demand structure is described by the constant model (2.1). Since the independent deviations ε_t cannot be predicted, we simply want to estimate the constant a . If a were completely constant the best estimate would be to take the average of all observations of x_t . But a can in practice be expected to vary slowly. This means that we need to focus on the most recent values of x_t . The idea of the moving average technique is to take the average over the N most recent values. Let

\hat{a}_t = estimate of a after observing the demand in period t

$\hat{x}_{t,\tau}$ = forecast for period $\tau > t$ after observing the demand in period t

We obtain:

$$\hat{x}_{t,\tau} = \hat{a}_t = (x_t + x_{t-1} + x_{t-2} + \cdots + x_{t-N+1})/N. \quad (2.4)$$

Note that the forecasted demand is the same for any value of $\tau > t$. This is, of course, because we are assuming a constant model.

The value of N should depend on how slowly we think that a is varying, and on the size of the stochastic deviations ε_t . If a is varying more slowly and the stochastic deviations are larger, we should use a larger value of N . This will limit the influence of the stochastic deviations. On the other hand, if a is varying more rapidly and the stochastic variations are small, we should prefer a small value of N , which will allow us to follow the variations in a in a better way.

If we use one month as our period length and set $N = 12$, the forecast is the average over the preceding year. This may be an advantage if we want to prevent seasonal variations from affecting the forecast.

2.4 Exponential Smoothing

2.4.1 Updating Procedure

When using exponential smoothing instead of a moving average, the forecast is updated differently. The result is in many ways similar, though. We are again assuming a constant demand model, and we wish to estimate the parameter a . To update the

forecast in period t , we use a linear combination of the previous forecast and the most recent demand x_t ,

$$\hat{x}_{t,\tau} = \hat{a}_t = (1 - \alpha)\hat{a}_{t-1} + \alpha x_t, \quad (2.5)$$

where $\tau > t$ and

α = smoothing constant ($0 < \alpha < 1$)

Due to the constant demand model the forecast is again the same for any future period.

Note that the updating procedure can also be expressed as

$$\hat{x}_{t,\tau} = \hat{a}_t = \hat{a}_{t-1} + \alpha(x_t - \hat{a}_{t-1}). \quad (2.6)$$

We have assumed that $0 < \alpha < 1$ although it is also possible to use $\alpha = 0$ and $\alpha = 1$. The value $\alpha = 0$ means simply that we do not update the forecast, while $\alpha = 1$ means that we choose the most recent demand as our forecast.

2.4.2 Comparing Exponential Smoothing to a Moving Average

To be able to compare exponential smoothing to a moving average, we can express the forecast in the following way:

$$\begin{aligned} \hat{a}_t &= (1 - \alpha)\hat{a}_{t-1} + \alpha x_t = (1 - \alpha)((1 - \alpha)\hat{a}_{t-2} + \alpha x_{t-1}) + \alpha x_t \\ &= \alpha x_t + \alpha(1 - \alpha)x_{t-1} + (1 - \alpha)^2\hat{a}_{t-2} = \cdots = \alpha x_t + \alpha(1 - \alpha)x_{t-1} \\ &\quad + \alpha(1 - \alpha)^2x_{t-2} + \cdots + \alpha(1 - \alpha)^n x_{t-n} + (1 - \alpha)^{n+1}\hat{a}_{t-n-1} \end{aligned} \quad (2.7)$$

Let us now compare (2.7) to (2.4). In (2.4) the N last period demands all have the weight $1/N$. In (2.7) we have, in principle, positive weights for all previous demands, but the weights are decreasing exponentially as we go backwards in time. This is the reason for the name exponential smoothing. The sum of the weights is still unity.¹ When using a moving average, a larger value of N means that we put relatively more emphasis on old values of demand. When applying exponential smoothing, a small value of α will give essentially the same effect.

When using a moving average according to (2.4) the forecast is based on the demands in periods $t, t-1, \dots, t-N+1$. The ages of these data are respectively $0, 1, \dots$, and $N-1$ periods. The weights are all equal to $1/N$.

¹ Let $0 \leq x < 1$ and consider the infinite geometric sum $S(x) = 1 + x + x^2 + x^3 \cdots$. Note that $S(x) = 1 + x \cdot S(x)$. This implies that $S(x) = 1/(1 - x)$. The sum of the weights in (2.7) is $\alpha \cdot S(1 - \alpha) = 1$.

Table 2.1 Weights for demand data in exponential smoothing

Period	Weight	$\alpha = 0.1$	$\alpha = 0.3$
t	α	0.100	0.300
$t-1$	$\alpha(1-\alpha)$	0.090	0.210
$t-2$	$\alpha(1-\alpha)^2$	0.081	0.147
$t-3$	$\alpha(1-\alpha)^3$	0.073	0.103
$t-4$	$\alpha(1-\alpha)^4$	0.066	0.072

According to (2.5), or equivalently (2.7):²

$$\alpha 0 + \alpha(1-\alpha)1 + \alpha(1-\alpha)^2 2 + \cdots = \alpha(1-\alpha)S'(1-\alpha) = (1-\alpha)/\alpha \quad (2.8)$$

The forecasts are based on data of the “same average age” if

$$(1-\alpha)/\alpha = (N-1)/2, \quad (2.9)$$

or equivalently when

$$\alpha = 2/(N+1). \quad (2.10)$$

Consider, for example, a moving average that is updated monthly with $N = 12$. This means that each month in the preceding year has weight $1/12$. Consider then an exponential smoothing forecast that is also updated monthly. A value of α “corresponding” to $N = 12$ is according to (2.10) obtained as $\alpha = 2/(12+1) = 2/13 \approx 0.15$.

2.4.3 Practical Considerations and an Example

If the period length is one month, it is common in practice to use a smoothing constant α between 0.1 and 0.3.

We can see from Table 2.1 that the forecasting system will react much faster if we use $\alpha = 0.3$. On the other hand, the stochastic deviations will affect the forecast more compared to when $\alpha = 0.1$. When choosing α we always have to compromise.

If the forecast is updated more often, for example each week, a smaller α should be used. To see how much smaller, we can apply (2.10) as a reasonable approximation. Assume that we start with a monthly update and that we use the value of α “corresponding” to a moving average with $N = 12$, i.e., $\alpha \approx 0.15$. When changing to weekly forecasts it is natural to change N to 52. The “corresponding” value of α is obtained from (2.10) as $\alpha = 2/(52+1) \approx 0.04$.

When starting to forecast according to (2.5) in some period t , an initial forecast to be used as \hat{a}_{t-1} is needed. We can use some simple estimate of the average period

² Let $0 \leq x < 1$ and consider the infinite geometric sum $S'(x) = 1 + 2x + 3x^2 \dots = 1 + x + x^2 + \dots + x(1 + 2x + 3x^2 \dots) = S(x) + x \cdot S'(x)$. This implies that $S'(x) = S(x)/(1-x) = 1/(1-x)^2$.

Table 2.2 Forecasts obtained by exponential smoothing with $\alpha = 0.2$. Initial forecast $\hat{a}_2 = 100$

Period	Demand in period t , x_t	Forecast at the end of period t , \hat{a}_t
3	72	94
4	170	110
5	67	101
6	95	100
7	130	106

demand. If no such estimate is available, it is possible to start with $\hat{a}_{t-1} = 0$, since \hat{a}_{t-1} will not affect the forecast in the long run, see (2.7). However, especially for small values of α , it can take a long time until the forecasts are reliable. If it is necessary to start with a very uncertain initial forecast, it may be a good idea to use a rather large value of α to begin with, since this will reduce the influence of the initial forecast.

Example 2.1 The demand for an item usually fluctuates considerably. A moving average or a forecast obtained by exponential smoothing gives essentially an average of more recent demands. The forecast cannot, as we have emphasized before, predict the independent stochastic deviations. Table 2.2 shows some typical demand data and the corresponding exponential smoothing forecasts with $\alpha = 0.2$. It is assumed that the forecast after period 2 is $\hat{a}_2 = 100$.

In Table 2.2 the forecast immediately after period 3 is obtained by applying (2.5).

$$\hat{a}_3 = 0.8 \cdot 100 + 0.2 \cdot 72 = 94.4,$$

which is rounded to 94 in Table 2.2. (The forecasts should normally only be rounded in the final output.) Note that when determining \hat{a}_3 the demands in future periods are not known. Therefore at this stage, \hat{a}_3 serves as our forecast for any future period. After period 4 the forecast is again updated

$$\hat{a}_4 = 0.8 \cdot 94.4 + 0.2 \cdot 170 = 109.52.$$

If we compare exponential smoothing to a moving average there are some obvious but minor advantages with exponential smoothing. Because the average a , which we wish to estimate is assumed to vary slowly, it is reasonable to use larger weights for the most recent demands as is done in exponential smoothing. As we have discussed before, however, a moving average over a full year may be advantageous if we want to eliminate the influence of seasonal variations on the forecast. It is also interesting to note that with exponential smoothing we only need to keep track of the previous forecast and the most recent demand.

2.5 Exponential Smoothing with Trend

In practice, exponential smoothing (or possibly a moving average) is, in general, a suitable technique for most items. But there is also usually a need for other methods for relatively small groups of items for which it is feasible and interesting to follow up trends and/or seasonal variations.

2.5.1 Updating Procedure

Let us now instead assume that the demand follows a trend model according to (2.2). To forecast demand we need to estimate the two parameters a and b , compared to only a in case of a constant model. As before, we cannot predict the independent deviations ε_t . There are different techniques for estimating a and b . We shall here consider a method that was first suggested by Holt (1957). (Another technique based on linear regression is described in Sect. 2.7.) Estimates of a and b are successively updated according to (2.11) and (2.12).

$$\hat{a}_t = (1 - \alpha)(\hat{a}_{t-1} + \hat{b}_{t-1}) + \alpha x_t, \quad (2.11)$$

$$\hat{b}_t = (1 - \beta)\hat{b}_{t-1} + \beta(\hat{a}_t - \hat{a}_{t-1}), \quad (2.12)$$

where α and β are smoothing constants between 0 and 1.

The “average” \hat{a}_t corresponds to period t , i.e., the period for which we have just observed the demand. The forecast for a future period, $t + k$ is obtained as

$$\hat{x}_{t,t+k} = \hat{a}_t + k \cdot \hat{b}_t. \quad (2.13)$$

The most important difference compared to simple exponential smoothing according to (2.5) is that the forecasts for future periods are no longer the same. Note that the trend, i.e., the change per period, can just as well be negative.

The method means that \hat{a}_t is always adjusted to fit the present period. With this in mind, (2.11) is essentially equivalent to (2.5). As long as x_t is unknown, $\hat{a}_{t-1} + \hat{b}_{t-1}$ is our best estimate for the mean demand in period t . We determine \hat{a}_t as a linear combination of this estimate and the new demand x_t . The average difference between two consecutive values of \hat{a}_t should in the long run be equal to the trend. Therefore, we use these differences in (2.12) to update the trend by exponential smoothing.

It can be shown that if the demand is a linear function without stochastic variations, the forecast will, in the long run independent of the initial values, estimate the future demand exactly. When using simple exponential smoothing this is not the case. See Problem 2.4.

Table 2.3 Forecasts obtained by exponential smoothing with trend. The smoothing constants are $\alpha = 0.2$ and $\beta = 0.1$, and the initial forecast $\hat{a}_2 = 100$, $\hat{b}_2 = 0$

Period t	Demand in period t , x_t	Forecast for period $t + 1$ at the end of period t , $\hat{a}_t + \hat{b}_t$	Forecast for period $t + 5$ at the end of period t , $\hat{a}_t + 5\hat{b}_t$
3	72	94	92
4	170	110	114
5	67	102	102
6	95	100	100
7	130	107	109

2.5.2 Practical Considerations and an Example

The idea behind exponential smoothing with trend is to be able to better follow systematic linear changes in demand. As with exponential smoothing, larger values of the smoothing constants α and β will mean that the forecasting system reacts faster to changes but will also make the forecasts more sensitive to stochastic deviations. When choosing values in practice it can be recommended to have a relatively low value of β since errors in the trend can give serious forecast errors for relatively long forecast horizons. Note that the trend is multiplied by k in (2.13). It is therefore very unfortunate if pure stochastic variations are interpreted as a trend. When updating the forecast monthly, typical values of the smoothing constants may be $\alpha = 0.2$ and $\beta = 0.05$. When the forecasting system is initiated it is usually reasonable to set the trend to 0 and, as in exponential smoothing, let the initial \hat{a} be equal to some estimate of the average period demand. If the initial values are very uncertain it can be reasonable, also for exponential smoothing with trend, to use extra large smoothing constants in an initial phase.

Example 2.2 We consider the same demand data as in Example 2.1. Table 2.3 illustrates the forecasts when applying exponential smoothing with trend and looking one and five periods ahead, respectively. The smoothing constants are $\alpha = 0.2$ and $\beta = 0.1$. At the end of period 2, $\hat{a}_2 = 100$ and $\hat{b}_2 = 0$.

In period 3 we obtain from (2.11) and (2.12)

$$\hat{a}_3 = 0.8 \cdot (100 + 0) + 0.2 \cdot 72 = 94.4,$$

$$\hat{b}_3 = 0.9 \cdot 0 + 0.1 \cdot (94.4 - 100) = -0.56.$$

Our forecast for period 4 is then $94.4 - 0.56 = 93.84 \approx 94$. At the end of period 4 we obtain the real demand 170. Applying (2.11) and (2.12) again we get

$$\hat{a}_4 = 0.8 \cdot (94.4 - 0.56) + 0.2 \cdot 170 = 109.072,$$

$$\hat{b}_4 = 0.9 \cdot (-0.56) + 0.1 \cdot (109.072 - 94.4) = 0.9632,$$

The forecasts in Table 2.3 are rounded to integers.

Inventory Control

Axsäter, S.

2015, XIII, 268 p. 50 illus., Hardcover

ISBN: 978-3-319-15728-3