

Christoph Kilger and Michael Wagner

The target of SCM is to fulfill the (ultimate) customer demand (Chap. 1). Customer demand does either explicitly exist as actual customer orders that have to be fulfilled by the supply chain, or it does exist only implicitly as anonymous buying desires (and decisions) of consumers. In the latter case, there is no informational object representing the demand.

Many decisions in a supply chain must be taken prior to the point in time when the customer demand becomes known. For example, replenishment decisions in a retail store are taken before a customer enters the store. Production quantities for make-to-stock products are determined prior to the point in time when the customer places orders. Decisions about procurement of raw materials and components with long lead times have to be taken before customer orders for finished goods using these raw materials or components become known. These examples describe decisions in a supply chain that have to be taken prior to the point in time when actual customer demand becomes known. Therefore, these decisions must be based on *forecasted customer demand*, also called *demand forecast*. The process of forecasting future customer demand is called *demand planning*.

The next section introduces a framework for demand planning processes, that helps to explain the structures and processes of demand planning.

C. Kilger (✉)

Ernst & Young GmbH, Wirtschaftsprüfungsgesellschaft, Heinrich-Böcking-Straße 6–8,
66121 Saarbrücken, Germany
e-mail: christoph.kilger@de.ey.com

M. Wagner

Paul Hartmann AG, Process Excellence Management, Information Technology,
Paul-Hartmann-Str. 12, 89522 Heidenheim, Germany
e-mail: michael.wagner@hartmann.info

7.1 A Demand Planning Framework

Forecasting future customer demand is quite easy, if there is just one product and one customer. However, in reality demand planning comprises often hundreds or even thousands of individual products and individual customers. In some cases, it is even impossible to list all products (e.g. in the case of configurable products) or to know all customers (e.g. in the consumer goods industry). Furthermore, demand planning usually covers many time periods, typically 12–24 months. Thus, an important aspect of demand planning is to define proper planning structures for products, customers and time. These structures are used to represent input to the forecasting process, historic transactional data and computed data like a statistical forecast or a forecast accuracy metric. Furthermore, aggregation and disaggregation of data takes place based on the pre-defined demand planning structures. Section 7.2 discusses demand planning structures.

In Sect. 7.3, we describe the demand planning process, which consists of the following steps:

1. Collection of input data like forecast data from former planning runs, historic customer orders, shipments, etc. and correction of historic data;
2. computation of further data, e.g. statistical forecast;
3. judgmental forecasting by the human planners, which review the planning situation and give their input (this might include planning of promotions);
4. consensus forecasting, consolidating the different views of the planners and dealing with exceptions;
5. planning of dependent demand, i.e. the demand for components of the finished goods (in case of product bundles, configurable products, etc.);
6. release of the forecast to further planning and execution processes, e.g. master planning, purchasing, allocation planning, collaborative planning.

In many situations a good forecast can be computed automatically from historic customer orders. This is called *statistical forecasting* and usually takes place in step 2 of the demand planning process (see above). Statistical forecasting uses sophisticated methods to create forecasts for a lot of items automatically. As there are many statistical forecasting techniques, each having multiple parameters influencing the results, it is hard to find the best statistical forecasting technique and to set the parameters properly. To support the selection of a statistical forecasting method and to estimate the parameters many APS offer so-called *pick-best functions*. Statistical forecasting is described in detail in Sect. 7.4.

As described at the beginning of this chapter, the task of demand planning is to support processes that need information on the customer demand, but have to be executed *prior* to the point in time when the customer demand becomes known. So far, this task seems to be quite easy. But, as Nahmias (2009) argues in his textbook, the main characteristic of forecasts is that they are usually wrong! Therefore, each planning step which is based on forecasted demand contains uncertainty to some extent. It is apparent that the accuracy of the forecast directly influences the quality of the processes using the forecast. In order to achieve a high forecast accuracy

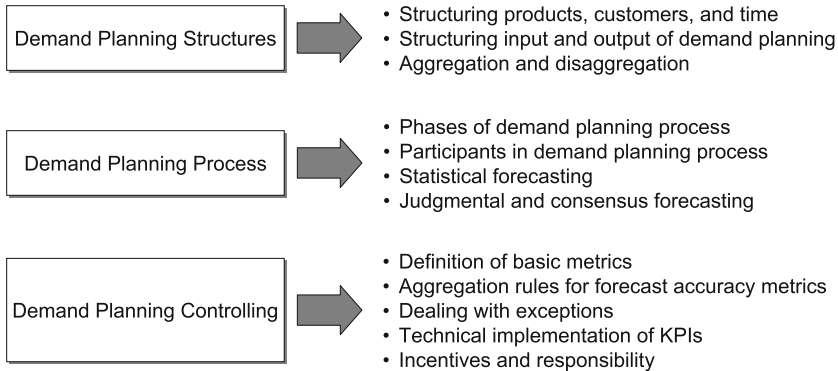


Fig. 7.1 Demand planning framework

it is necessary to implement appropriate controlling mechanisms, for example measuring the forecast accuracy (Eickmann 2004). Section 7.5 describes controlling mechanisms for demand planning. The overall demand planning framework is summarized in Fig. 7.1.

Forecasting, as described above, is not an actual planning or decision process as it “only” aims at predicting the future demand as accurately as possible. However, there are business decisions that influence the future demand, e.g. promotions and new product introductions. The impact on the demand plan might be assessed by the demand planning process as a what-if analysis or simulation run, and a subsequent master planning process is used to determine the corresponding supply plan. This feedback loop between demand planning and master planning is often called sales and operations planning S&OP (see Sect. 8.4). Based on the results of the master planning, adjustments to the business decisions and the demand plan can be decided.

The difference between planned and actual sales influences the service level of the whole supply chain. As this service level usually cannot reach 100 %, safety stocks are an adequate tool for improving customer service. The amount of safety stock required for reaching a desired service level is closely linked to the forecast accuracy. These and further additional features of demand planning (price-based planning, sporadic demand, lost sales vs. backorders, model selection and parameter estimation) are summarized in Sect. 7.6.

7.2 Demand Planning Structures

The task of demand planning is to predict the future customer demand for a set of items. The demand pattern for a particular item can be considered as a time series of separate values (Silver et al. 1998). For each item, there may be multiple time series, representing for example historic data, forecast data or computed data like the forecast accuracy. The selection of the right time series to be used in the demand

planning process depends on the answer to the question *What is being forecasted?* For example, a mid-term master planning process might require forecasted customer orders (customer requested date) for every product group, sales region and week. On the other hand, short-term replenishment decisions for finished products may be based on forecasted shipments (shipment date) for every product in daily time buckets, grouped by distribution center. The examples illustrate that it is necessary to clarify the requirements of all processes that will use the forecast before designing the demand planning structures.

In general each forecast consists of three components:

1. The *time period*, in which the forecasted demand is planned to substantiate as customer demand;
2. the *product*, that will be requested by the customer;
3. the *geographical region*, from where the customer demand will originate;

Thus, there are three dimensions along which forecast data can be structured: time, product and geography. In the following we discuss the structuring of forecast data along these dimensions, and conclude with considerations about the consistency of forecast data in complex demand planning structures.

7.2.1 Time Dimension

For demand planning time is structured in discrete *time buckets*, e.g. years, quarters, months, weeks, days. All demand planning data (actuals, forecast and computed measures) are represented as time series. Each time series consists of a sequence of time buckets. The period of time covered by the time buckets is called *demand planning horizon*.

The size of the time bucket depends on the requirements of the particular demand planning scenario considered. For example, a fast food chain that intends to forecast demand patterns within the next weeks will use daily time buckets. In consumer packaged goods industry and many other industries, the forecast is usually structured in months—as monthly buckets are well suited to capture seasonal demand patterns and drive buying, production and replenishment decisions. As the examples show, the selection of the size of the time buckets depends on the maximum resolution of the time dimension required by the processes that will use the forecast: Time buckets should be granular enough to prepare the supply chain for the fulfillment of the forecasted demand. On the other hand if time buckets are too granular one might easily run into performance problems.

In most APS time can be structured hierarchically. For example, forecast data that is entered in months can be aggregated to quarters and years and can be disaggregated to weeks and days. Aggregation and disaggregation rules are described in the next section.

Please note that the conversion of weekly into monthly forecast data and vice versa is not straight forward. Figure 7.2 illustrates the relationship between weeks and months. In order to convert forecast data between weeks and months, most APS disaggregate the forecast to the lowest level (days) and aggregate it from there to

Month 1				Month 2					
Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	

Fig. 7.2 Conversion of weeks and months

any time granularity. Another approach is to define so-called *split weeks* along the boundaries of months. In Fig. 7.2 weeks 5 and 9 would be split into two time buckets at the beginning and the end of month 2.

7.2.2 Product Dimension

Forecasting may take place on the level of SKUs (stock keeping units, e.g. final products) or on the level of product groups. Forecasting on SKU-level creates an individual forecast for each SKU, reflecting its individual demand pattern. Forecasting on product group level results in a more aggregated forecast. In most industries the number of SKUs is very large and prevents forecasting on SKU level. Please note that it is more difficult to create a highly accurate forecast on SKU level than on product group level—thus the forecast accuracy on group level is usually higher than on SKU level.

SKUs can be aggregated to product groups in multiple ways. Let us take the beverage industry as an example. Figure 7.3 shows multiple ways to form product groups from finished products. The left branch groups products by size and packaging. The middle branch shows the grouping by taste (Cola, Ginger Ale, Root Beer, etc.; Soft Drinks, Ice Teas, Juices, etc.). The right branch groups products by their style, i.e. whether they contain sugar (regular) or sweetener (diet). This grouping can be used to anticipate general trends of consumer demand.

The forecast can be entered on any of these aggregation branches and levels. For instance, the forecast planners from the sales organization would enter their forecast on the “subgroup” level, i.e. Cola, Ginger Ale, Root Beer, etc. Planners from the product management department would forecast the distribution of regular vs. diet beverages on the “style” level. On each level there may be one or multiple time series representing the forecast quantities.

Forecasts can easily be *aggregated* to higher levels. For example, the forecast of the sales planners can be aggregated to the product “group” level (Soft Drinks, Ice Teas, Juices, etc.) and to the top level (“Beverages”) by adding up the forecast quantities of the “subgroups” level. Forecast quantities can also be *disaggregated* to lower levels. Disaggregation of a forecast quantity to a lower level has to be defined by *disaggregation rules* (see e.g. Meyr 2012):

- *Even distribution:* The forecast quantity of the higher level is evenly distributed to the items on the lower level.

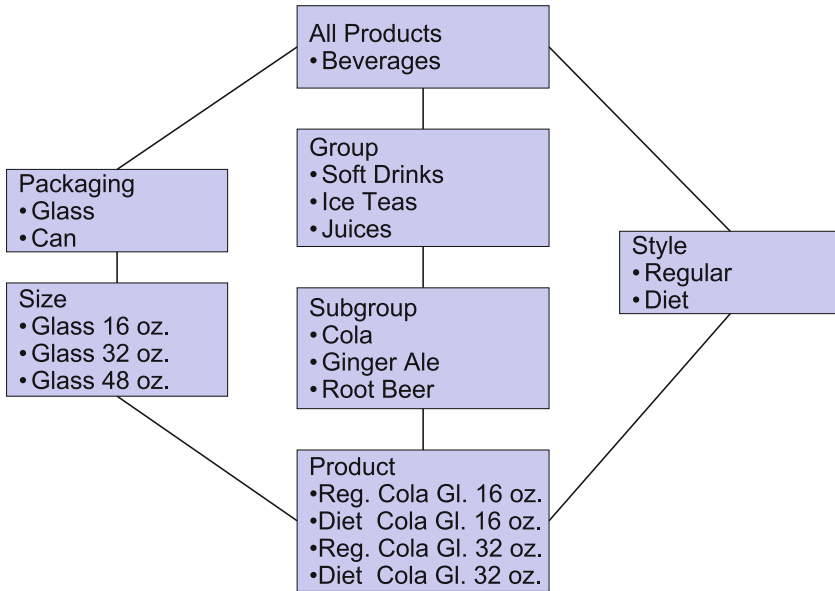


Fig. 7.3 Product dimension (example)

- *Existing quantities on lower level:* If there are already forecast quantities on the lower level, the percentage distribution of the instances is computed and applied to the forecast quantity on the higher level.
- *According to some other time series:* The distribution of values of some other time series is used to disaggregate the quantities from the higher to the lower level. For instance, the forecast on top level (“Beverages”) could be disaggregated to the “packaging” level by using the historic distribution of packaging styles from a time series representing historic sales. Figure 7.4 illustrates the disaggregation by some other time series.

In many cases the values used for disaggregation are taken from another time period, e.g. from the year before (as it is the case in the example shown in Fig. 7.4). In other cases data from the same time period is used. For example, the forecast on “subgroup” level could be disaggregated to “product” level using forecasted quantities for the packaging style and sizes and forecasted quantities for the consumer trends towards regular vs. diet beverages for the same time period. However, this will require more complex calculation schemes than the simple disaggregation rules described above. Many APS offer simple macro programming languages for this purpose.

Forecasting on the top product group and disaggregation of the forecast to lower levels is called *top-down forecasting*, forecasting on the lowest product level (SKU-level) and aggregation of the forecast to the higher product groups is called *bottom-up forecasting*. A combined approach, where first a planning level is selected

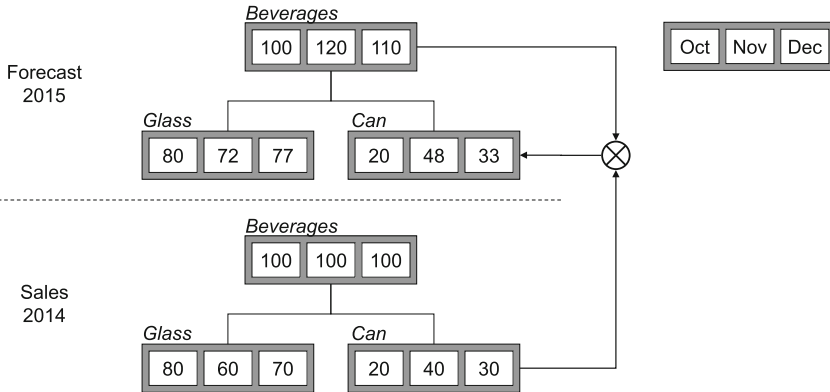


Fig. 7.4 Disaggregation by some other time series (example)

in the product hierarchy, a forecast is created on that level, which is then aggregated to higher product groups and disaggregated to SKU level, is called *middle-out forecasting*. The general term for demand planning on multiple product group levels is *hierarchical forecasting*.

7.2.3 Geography Dimension

The third dimension of forecasting is geography. As all demand originates from customers, customers form the lowest level of the geography dimension. Similar to products, customers may be grouped according to multiple aggregation schemes:

- Grouping by regions and areas supports the planning of regional demand
- Grouping by supply source (distribution centers, manufacturing plants, etc.) may be used to check the feasibility of the forecast against rough-cut capacity constraints
- Grouping by key account supports the consolidation of forecasts for international customer organizations, consisting of multiple national subsidiaries.

Figure 7.5 shows options to structure forecast by geography. Please note that similar aggregation and disaggregation rules can be applied to the geography dimension as described in the previous subsection for products.

In many modern APS there is no distinction between product dimension and geography dimension. Instead, planning structures are built up from *planning attributes* (sometimes also called *characteristics*). Planning attributes represent properties of the products used to structure the forecast and support the forecasting process by aggregation and disaggregation.

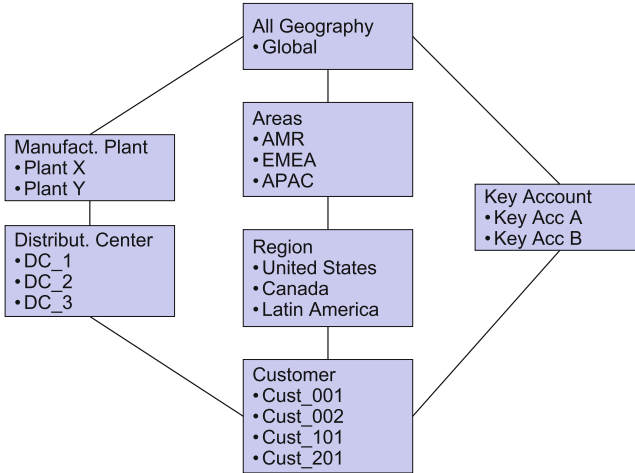


Fig. 7.5 Geography dimension (example)

7.2.4 Consistency of Forecast Data

As described in Sect. 7.3 demand planners usually select the best aggregation level to enter “their” forecast. Thus, forecast data may be entered on any level of the planning structures. As a consequence there may exist inconsistencies in the forecast data. As an example, consider the forecast quantities shown in Fig. 7.6. The forecast on “product” level is consistent with the aggregation on “subgroup” level, but there is an inconsistency between “product” level and the “packaging style and size” level. (The quantities in parentheses show the aggregated quantities from the “product” level.) A situation like this may occur if (1) forecast data is entered on “subgroup” level by some planner, (2) the forecast data is then automatically disaggregated to product level using the sales data from the year before, and (3) after that the forecast is changed on the level “packaging style and size”. There are two ways to keep forecast data over all levels of the planning structures consistent:

1. *Immediate propagation of changes:* All changes are aggregated to the higher levels and disaggregated to the lower levels applying pre-defined aggregation and disaggregation rules. Note that immediate propagation of changes might make changes to forecast data very slow as a lot of data has to be updated. Most APS enforce immediate propagation of changes as forecast data is stored only on the lowest level.
2. *Consistency checks:* Changes are entered into the APS without propagation to other levels. Aggregation and disaggregation rules are applied manually. The APS realigns the data on all levels and flags inconsistencies that cannot be resolved due to conflicting rules. These have than to be resolved manually.

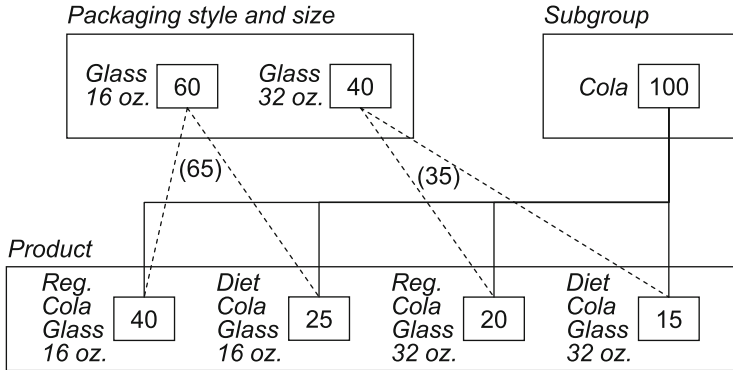


Fig. 7.6 Inconsistency of demand planning data

7.3 Demand Planning Process

The demand planning process consists of multiple phases (see e.g. Meyr 2012). Figure 7.7 shows a typical demand planning process that is used in many industries. The time scale shows the number of days needed to update the forecast in a monthly rolling forecasting process. The process starts in a central planning department with the *preparation phase*. In this phase the demand planning structures are updated by including new products, changing product groups, deactivating products that will no longer be sold (and therefore will not be forecasted anymore). The historic data is prepared and loaded into the demand planning module of the APS—e.g. shipments and customer orders. The accuracy for previous forecasts is computed (see Sect. 7.5 for details on the computation of forecast accuracy). In certain cases it is necessary to correct historic data before they may be used as input to demand planning. For example, shipment data must be corrected if stock-out situations occurred in the past—otherwise, these stock-out situations would potentially influence statistical forecasting methods using this time series as input.

In the second phase the *statistical forecast* is computed based on the updated historic data. Section 7.4 gives an introduction into statistical forecasting methods. When it comes to statistical methods and their application one typical question arises: How is the software able to make better forecasts than a human planner with years of experience in demand planning? The simple answer is that mathematical methods are unbiased. Empirical studies (see e.g. Makridakis et al. 1998) give evidence, that bias is the main reason why myopic statistical methods often produce better results. But that's only half of the truth, because information on specific events or changes (e.g. promotional activities, customer feedback on new products etc.) can lead to significant changes in demand patterns which might not be considered in standard time series analysis models. Therefore, it is necessary to combine the advantages of both worlds in an integrated demand planning process. For example, consider the demand planning process of a company selling beverages. In such

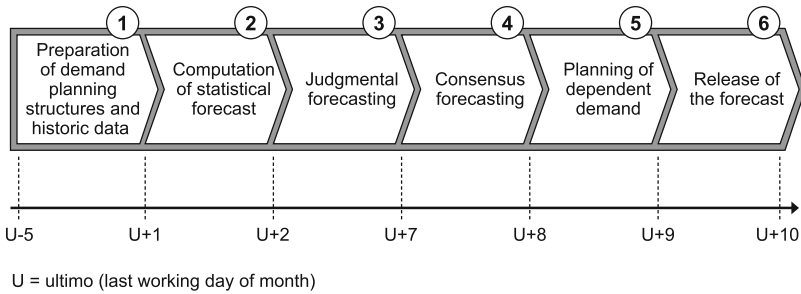


Fig. 7.7 Phases of a demand planning process

an environment the regular demand can be forecasted by a seasonal model quite accurately (refer to Sect. 7.4). But, the demand series are distorted by occasional additional demand due to promotional activities in some retail outlets. This effect can be estimated by the sales force responsible for the promotion, while the base line is forecasted by a seasonal model.

In the third phase of the demand planning process *judgmental forecasts* are created by multiple departments. Typical departments involved in judgmental forecasting are sales, product management, and marketing. Integration of statistical and judgmental forecasting is only reasonable, if information inherent in a statistical forecast is not considered in the judgmental process. In this case the information would be double counted and therefore the demand would be overestimated (or underestimated, if the judgment reduces the statistical forecast). In the following we describe some methods on how to integrate statistical forecasting and *structured judgment*. Non-structured judgment is often applied by demand planners, if they check the figures produced by a decision-support tool and “tune” the values using their sure instinct. But, for integration purposes it is necessary to structure judgment. Detailed process definitions and guidelines create a framework for such a structured judgment. Armstrong and Collopy in Wright and Goodwin (1998) describe the following five procedures for the integration:

- *Revised judgmental forecasts*: The first step in this procedure is made by demand planners, who create judgmental forecasts based on the knowledge of relevant data (e.g. historical data, causal factors etc.). Afterwards they are confronted with forecasts which are calculated using statistical methods. Then, the planners have the possibility to revise their initial estimate incorporating the new information. But, there is no predefined percentage to which extent each of the components has to be considered in the final forecast. This procedure often leads to more accurate forecasts than simple judgment not aided by statistical methods. Furthermore, it has the advantage that it leaves the control over the demand planning process to the human planner.
- *Combined forecasts*: As the above procedure assigns variable weights to the two forecasts, it is evident that these values are often biased or influenced by political means. A more formal procedure is assured by combining the two values

according to a predefined weighing scheme. Even if equal weights are assigned to judgmental forecasts and statistical forecasts, better results are possible.

- *Revised extrapolation forecasts:* Modifying statistical forecasts manually to take specific domain knowledge of the planner into account is common practice in a lot of companies. But, the revision process has to be structured accordingly. This means that the judgmental modification has to be based on predefined triggers (e.g. promotions, weather etc.).
- *Rule-based forecasts:* Rule-based forecasts are also based on statistical forecasts. But, the selection or combination of different forecasting methods is supported by structured judgments of experts. The rules used for the selection are derived from the specific knowledge of the experts or on past research. They are based on characteristics of time series or on causal factors. Rule-based forecasting improves simple extrapolation methods especially, if the series have low variability and low uncertainty.
- *Econometric forecasts:* Regression models are referred to as econometric forecasting methods, if the model selection process and the definition of causal variables is provided by structured judgment. Improvements are reported especially, if this procedure is applied to long-range forecasts. As bias could have much impact on the result of econometric forecasts, it is advisable to give the judgmental process a very rigid structure.

In practice the forecast resulting from the structured judgment processes is often discussed in a *consensus forecast meeting*. The goal is to reach a consensus about open issues like different opinions about the influence of a promotion to the sales quantities in a particular region. The degree to which the judgmental forecasts from the individual departments contribute to the consensus forecast may be derived from the average forecast accuracy the departments achieved in the past. Consensus forecasting and structured judgment needs to be supported by detailed feedback mechanisms which show the planners the quality of their inputs. Therefore, forecast accuracy reports have to differentiate between the quality of (automatic) statistical forecasting and judgmental forecasts.

Based on the consensus forecast *dependent demand* may be planned. The consensus forecast represents the demand for finished products (or product groups representing finished products). In many industries it is necessary to compute demand on component level from the consensus forecast. There are three applications for the computation of dependent demand:

- *Constrained availability of a key component:* If there is a key component that limits the supply of the products, it might be required to check the feasibility of the forecast based on the demand for that component resulting from the forecast (Dickersbach 2009). The pharmaceutical industry is a good example for this, as the supply of active ingredients is typically constrained and fixed for a long period of time.
- *Demand constraints that can be expressed by a key component:* In other industries like the computer industry or the automotive industry, overall market demand is constrained, and every finished product contains a specific key component: In computer industry this component is the processor, in automotive

industry every Diesel car contains a fuel injection pump. The conformance of the forecast with realistic market development can easily be checked using the overall demand for these key components.

- *Product bundling*: Especially in consumer goods industries, products are often bundled as part of a promotion. These bundles have an individual product number and are forecasted in the same way as “normal” products. However, it is important to understand that these products consist of other products and—therefore—influence the demand for the products of which they consist. These so-called cannibalization effects have to be analyzed and the forecast has to be adjusted accordingly (Dickersbach 2009).

Of course the dependent demand of components is also determined during master planning and materials requirements planning. However, it is much faster to compute and check dependent demand as part of the demand planning process and to update the forecast immediately.

The last step of the demand planning process is the formal approval and technical release of the forecast. This step makes the forecast available for other processes.

7.4 Statistical Forecasting Techniques

Forecasting methods were developed since the 1950s for business forecasting and at the same time for econometric purposes (e.g. unemployment rates etc.). The application in software modules makes it possible to create forecasts for a lot of items in a few seconds. Therefore, all leading APS vendors incorporate statistical forecasting procedures in their demand planning solution. These methods incorporate information on the history of a product/item in the forecasting process for future figures. There exist two different basic approaches—time series analysis and causal models. The so-called *time series analysis* assumes that the demand follows a specific pattern. Therefore, the task of a forecasting method is to estimate the pattern from the history of observations. Future forecasts can then be calculated from using this estimated pattern. The advantage of those methods is that they only require past observations of demand. The following demand patterns are most common in time series analysis (see Silver et al. 1998 and also Fig. 7.8):

1. Level model: The demand x_t in a specific period t consists of the level a and random noise u_t which cannot be estimated by a forecasting method.

$$x_t = a + u_t \quad (7.1)$$

2. Trend model: The linear trend b is added to the level model’s equation.

$$x_t = a + b \cdot t + u_t \quad (7.2)$$

3. Seasonal model: It is assumed that a fixed pattern repeats every T periods (cycle). Depending on the extent of cyclic oscillations a multiplicative or an additive relationship can be considered.

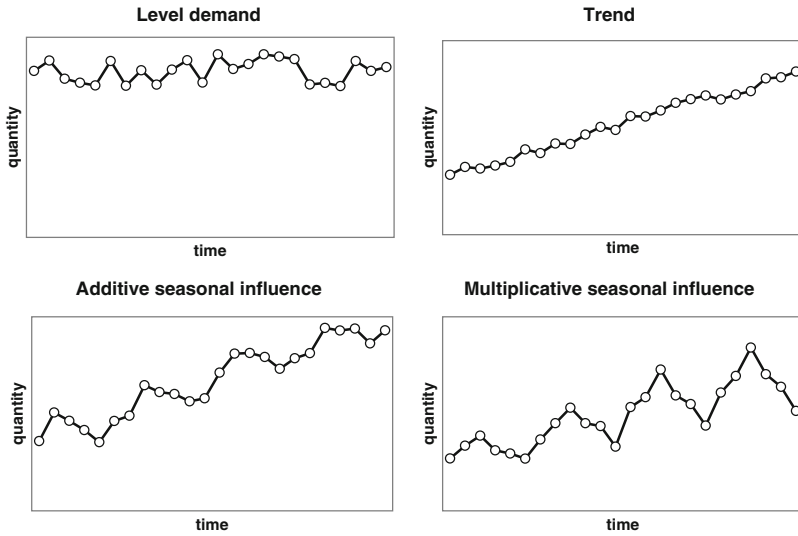


Fig. 7.8 Demand patterns

$$x_t = (a + b \cdot t) + c_t + u_t \quad \text{additive model,} \quad (7.3a)$$

$$x_t = (a + b \cdot t) \cdot c_t + u_t \quad \text{multiplicative model} \quad (7.3b)$$

where $c_t = c_{t-T} = c_{t-2T} = \dots$ are seasonal indices (coefficients).

The second approach to statistical forecasting are *causal models*. They assume that the demand process is determined by some known factors. For example, the sales of ice cream might depend on the weather or temperature on a specific day. Therefore, the temperature is the so-called independent variable for ice cream sales. If enough observations of sales and temperature are available for the item considered, then the underlying model can be estimated. For this example, the model might consist of some amount of independent demand z^0 and the temperature factor $z^1(t)$

$$x_t = z^0 + z^1(t) \cdot w_t + u_t \quad (7.4)$$

where w_t is the temperature on day t .

As for parameter estimation in causal models the demand history and one or more time series with indicators are needed, the data requirements are much higher than for time series analysis. Furthermore, practical experience shows that simple time series models often produce better forecasts than complex causal models (see e.g. Silver et al. 1998, p. 130). These tend to interpret stochastic fluctuations (noise) as “structure” and therefore, introduce a systematic error into the model. In the following two paragraphs the characteristics and the approach of the most frequently used forecasting methods are described.

Table 7.1 Weights of past observations in exponential smoothing for $\alpha = 0.2$

Period	0	-1	-2	-3	-4	...
Weight	0.2	0.16	0.13	0.10	0.08	...

7.4.1 Moving Average and Smoothing Methods

As each demand history is distorted by random noise u_t , the accurate estimation of parameters for the model is a crucial task. Also, the parameters are not fix and might change over time. Therefore, it is necessary to estimate under consideration of actual observations *and* to incorporate enough past values to eliminate random fluctuations (conflicting goals!).

Simple Moving Average. The simple moving average (MA) is used for forecasting items with level demand (7.1). The parameter estimate for the level \hat{a} is calculated by averaging the past n demand observations. This parameter serves as a forecast for all future periods, since the forecast \hat{x}_{t+1} is independent of time. According to simple statistics, the accuracy of the forecast will increase with the length n of the time series considered, because the random deviations get less weight. But this is no more applicable if the level changes with time. Therefore, values between three and ten often lead to reasonable results for practical demand series. But the information provided by all former demands is lost according to this procedure.

Exponential Smoothing. The need to cut the time series is avoided by the exponential smoothing method, because it assigns different weights to all (!) observed demand data and incorporates them into the forecast. The weight for the observations is exponentially decreasing with the latest demand getting the highest weight. Therefore, it is possible to stay abreast of changes in the demand pattern and to keep the information which was provided by older values. For the case of level demand the forecast for period $t + 1$ will be calculated according to the following equation:

$$\hat{x}_{t+1} = \hat{a}_t = \alpha \cdot x_t + \alpha(1 - \alpha) \cdot x_{t-1} + \alpha(1 - \alpha)^2 \cdot x_{t-2} + \dots \quad (7.5)$$

The parameter α is the smoothing constant, to which values between 0 and 1 can be assigned. For $\alpha = 0.2$ the weights in Table 7.4.1 are being used, if the forecast has to be made for period 1. Furthermore it is not necessary to store the whole history of an item as (7.5) can be simplified. The only data which needs to be kept in the database are the latest forecast and the latest demand value. Exponential smoothing for level demand patterns is easy to apply and requires little storage capacity. Therefore, it provides good forecasts for this kind of model and it also calculates reasonable forecasts for items which are influenced by high random fluctuations (Silver et al. 1998).

The exponential smoothing procedure for level demand can be extended to trend models and multiplicative seasonal models [see (7.2) and (7.3b)]. The method for the trend model is known as Holt's procedure (see e.g. Nahmias 2009). It smoothes both terms of the model, the level a and the trend component b with different smoothing constants α and β .

Winters introduced the seasonal model with exponential smoothing. A lot of lines of business are facing seasonal patterns, but don't incorporate it in forecasting procedures. For example, consider the manager of a shoe store, who wants to forecast sales for the next 2 weeks in daily buckets. As sales are usually higher on Saturdays than on Mondays, he has to take the weekly "season" into account. Winters' method is an efficient tool to forecast seasonal patterns, because it smoothes the estimates for the three parameters a , b and c . In contrast to the former two models the seasonal method needs far more data to initialize the parameters. For reliable estimates for the seasonal coefficients it is necessary to consider at least two cycles of demand history (e.g. 2 years). For more details on Winters' model see Chap. 29.

7.4.2 Regression Analysis

Where significant influence of some known factors is present, it seems to be straightforward to use causal models in the forecasting process. Regression analysis is the standard method for estimation of parameter values in causal models. Usually linear dependencies between the dependent variable x_t (e.g. the demand) and the leading factors (independent variables; e.g. temperature, expenditures for promotions etc.) are considered. Therefore, a multiple regression model can be formulated as follows (see e.g. Hanke and Wichern 2014):

$$x_t = z_0 + z_1 \cdot w_{1t} + z_2 \cdot w_{2t} + \dots \quad (7.6)$$

The ice cream model in (7.4) is called the simple regression model, as it only considers one leading indicator. Multiple linear regression uses the method of least squares to estimate model parameters (z_0, z_1, z_2, \dots). This procedure minimizes the sum of the squared difference between the actual demand and the forecast the model would produce. While exponential smoothing can consider all past observations, the regression method is applied to a predefined set of data. The drawbacks of such a procedure are the same as for the moving average model. Further, the weight of all considered values equals one and therefore the model cannot react flexibly to changes in the demand pattern.

As the data requirements of linear regression models are much higher than for simple time series models, it is obvious that this effort is only paid back, if the models are used for aggregate mid-term or long-term forecasts or for a few important end products.

The following example shows the application of linear regression for the ice cream model: Assuming that the ice cream retailer observed the following demands

Table 7.2 Demand and temperature data for the ice cream example

Period	1	2	3	4	5	6	7	8	9	10
Actual demand	43	45	54	52	54	55	43	33	52	51
Temperature (°C)	15	17	19	16	21	22	18	15	19	18

Table 7.3 Example forecasts using the linear regression model

Period	1	2	3	4	5	6	7	8	9	10
Model value	42	46	50	44	55	57	48	42	50	48

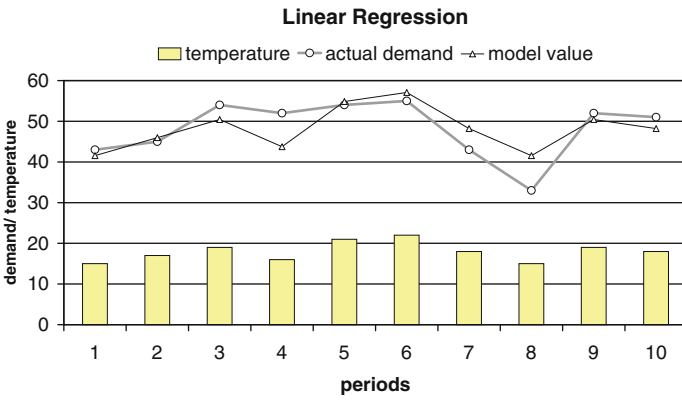


Fig. 7.9 Linear regression: results for the ice cream model

and temperatures (°C) over 10 days (Table 7.2) the linear regression will calculate the equation

$$\text{demand } x_t = 8.24 + 2.22 \cdot w_{1t} \tag{7.7}$$

with w_{1t} being the temperature on day t . Using (7.7) one can determine the forecasts (model value) which the model would have produced (see Table 7.3). But, for this it is necessary to be able to estimate the temperature reliably. Figure 7.9 shows the data and the resulting forecasts for the ice cream model.

7.5 Demand Planning Controlling

Demand planning controlling has the task to control the quality of the forecast and the quality of the demand planning process itself. The processes using the forecast as a foundation for their decisions (pre-production, purchasing, provision of additional capacity, etc.) need a quality measure to understand the accuracy of the forecast and the dimension of possible deviations of the forecast from the actual demand. No Master Planner would accept forecasts without being sure about the quality of the demand plan. Furthermore, the quality of the forecast is used as a feedback mechanism for the contributors to receive information about the quality of their contributions.

7.5.1 Basic Forecast Accuracy Metric

The first step in setting up a demand planning controlling is to define a basic metric for the accuracy of the forecast on some level of the demand planning structures. Based on this basic metric aggregated metrics can be computed. Note that aggregated measures cannot be computed directly on an aggregated level of the demand planning structures, as in this case shortage and excess planning would level out. A basic metric used to measure the forecast accuracy must have the following properties (Eickmann 2004):

- It must be summable. The domain of the metric must be positive. (Otherwise, positive and negative values would compensate when being aggregated.) The metric must be standardized (values between 0 and 100 %).
- All key figures (time series) required for the computation of the basic metric and for its aggregation must be available for all instances of the planning structures (all products, customers, time buckets, etc.). For example, a time series representing historic shipments might be not available for all instances of the planning structures (e.g. new products might not have historic shipments).
- It must be possible to get the buy-in of all involved departments in the organization regarding the definition of the basic metric. For example, if the “delivered quantity ex-works” is used as a reference to measure the forecast accuracy, sales might not commit to this metric—as sales cannot be made responsible for a low delivery service of production.

Furthermore, the level of the demand planning structures must be defined on which the basic forecast accuracy metric will be measured. Note that this level does not necessarily have to be the same as the lowest level of the demand planning structures (which is often only used to easily enter and structure the forecast data). For example assume that the total sales quantity per sales region is important to drive decisions in the supply chain, but the sales planners want to enter the forecast per customer. In this situation the basic forecast accuracy metric would be defined based on the sales quantity per sales region per article. Please note that shortage and excess planning per customer level out on the sales region level. It is important that the logistical

conditions are similar for all customers within the same sales region—otherwise the precondition mentioned at the beginning of this example would not hold, which states that supply chain decisions are driven by sales quantity per sales region, not per customer.

All accuracy measures are based on the forecast error $e_{t,r}$. It is defined as the difference between the forecasted quantity $\hat{x}_{t,r}$ and the actual quantity x_t : $e_{t,r} = \hat{x}_{t,r} - x_t$. The actual quantity x_t is the observed value of the time series (that is being forecasted) for time bucket t , e.g. shipments or customer orders based on customer requested date. The forecasted quantity $\hat{x}_{t,r}$ is the forecast for x_t that was created at time bucket r . Note, that in a rolling forecast scenario, there are multiple forecasts for the same actual quantity, each being created at a specific forecast run. The forecast error is influenced by the following parameters:

- *The time delta between forecast and actuals*: Forecasting is aiming at providing information about future shipments, sales etc. Normally, it is easier to tell the nearer future than the future that is far away. Thus, the forecast accuracy strongly depends on the time between the forecast creation and the time period that is being forecasted. For example, consider a forecast for the sales volume in June this year. The sales forecast for the month of June that has been created in March normally has a lower accuracy than the forecast created in May: $e_{\text{June, March}} > e_{\text{June, May}}$.
- *The forecast granularity*: The level of aggregation also has a strong impact on the forecast accuracy. Take sales forecast again as an example: It is easier to forecast the total sales volume for all products, for all geographic areas and for a complete fiscal year, than to forecast on a weekly basis low level product groups for all sales regions individually. Thus, the forecast accuracy normally decreases if the forecast granularity increases.

$e_{t,r}$ does not yet fulfill the rules for a basic forecast accuracy metric described above: it is not positive and not standardized. Thus we have to refine the definition of forecast error. Common refinements are the following:

$$\text{squared error } SE_{t,r} = e_{t,r}^2 \quad (7.8)$$

$$\text{absolute deviation } AD_{t,r} = |e_{t,r}| \quad (7.9)$$

$$\text{absolute percentage error } APE_{t,r} = \frac{|e_{t,r}|}{x_t} \cdot 100 \% \quad (7.10)$$

Note that $APE_{t,r}$ cannot be computed if the actual quantity x_t is zero (e.g. a product without any customer demand in time bucket t). In practice a forecast *accuracy* measure—e.g. according to the following equation—is often used rather than a forecast *error* measure:

$$\text{absolute percentage accuracy } APA_{t,r} = \max\{100 \% - APE_{t,r}; 0 \%\} \quad (7.11)$$

Implementations of this metric in APS may even consider the case that $x_t = 0$; in this case, $APA_{t,r}$ would be set to 0 %.

The basic forecast accuracy metrics must be aggregated in order to enable the controlling of the demand planning process. We distinguish between aggregation along the time dimension and aggregation along the product or geography dimension.

7.5.2 Aggregation of Forecast Accuracy by Time

There are many methods to aggregate the forecast accuracy or the forecast error by time. Each measure is calculated for a fixed horizon n (in the past) which has to be defined by the planner. If the horizon is short, then the value reacts fast to deviations from the average, but then it also might fluctuate heavily due to random demand variations. The following measures (for the first three measures see e.g. Silver et al. 1998) are common in practice:

$$\text{mean squared error } MSE_r = \frac{1}{n} \sum_{t=1}^n e_{t,r}^2 \quad (7.12)$$

$$\text{mean absolute deviation } MAD_r = \frac{1}{n} \sum_{t=1}^n |e_{t,r}| \quad (7.13)$$

$$\text{mean abs. perc. error } MAPE_r = \left[\frac{1}{n} \sum_{t=1}^n \frac{|e_{t,r}|}{x_t} \right] \cdot 100 \% \quad (7.14)$$

$$\text{mean abs. perc. accuracy } MAPA_r = \left[\frac{1}{n} \sum_{t=1}^n APA_{t,r} \right] \quad (7.15)$$

The MSE is the variance of the forecast error in the time horizon under consideration. In the Linear Regression forecasting procedure the MSE is used as the objective function which is minimized. As the error is squared in the formula, large deviations are weighted more heavily than small errors. Whereas the MAD uses linear weights for the calculation of the forecast accuracy. Further, the meaning of the MAD is easier to interpret, as it can be compared with the demand quantity observed. The main drawback of the two measures above is the lack of comparability. The values of MSE and MAD are absolute quantities and therefore, cannot be benchmarked against other products with higher or lower average demand. The measures MAPE and MAPA standardize the value based on the observed demand quantities x_t . The result is a percentage-value for the forecast error or accuracy, respectively, which is comparable to other products. The drawback of this calculation is the need for a positive actual. Therefore a rule for this case has to be defined.

The measures described above allow detailed analysis of the past, but they need to be discussed from the beginning each time they are calculated. In demand

planning tools for some 100 or 1,000 items one wants an automatic “interpretation” of the forecast error and therefore, might need an alert or triggering system. This system should raise an alert, if the statistical forecasting procedure no more fits to the time series or if the sales office did not provide the information on a sales promotion. Such an alert system can be triggered by thresholds which are based on one of the measures for the forecast accuracy. These thresholds are defined by the demand planner and updated under his responsibility. Besides the threshold technique some other triggering mechanisms have been developed which all are based on the forecast accuracy measured by MSE or MAD.

7.5.3 Aggregation of Forecast Accuracy by Product and Geography

In many industries, the units of measures, the sales quantities, the contribution margin, and the logistical conditions of all products considered in demand planning differ strongly. Thus, appropriate weighting schemes must be applied in order to aggregate the basic metric by product or by geography.

Simply computing the average of the basic forecast accuracy or forecast error for each instance of some aggregation level is not sufficient. For instance consider a product group with two articles A and B . Assume that in June A had sales of 1,000 and a forecast accuracy from May $APA_{A, \text{June}, \text{May}} = 100\%$ and B has sales of 10 pieces and a forecast accuracy of $APA_{B, \text{June}, \text{May}} = 0\%$. Using the average of both basic metrics results in a forecast accuracy for the product group of 50%—not reflecting the actual situation of the supply chain.

A common weighting factor for some product p is the sum of the forecasted quantity \hat{x}_p and the actual quantity x_p for that product related to the sum of forecasted quantity and actual quantity for all products¹:

$$\text{weight}_p = \frac{\hat{x}_p + x_p}{\sum_q (\hat{x}_q + x_q)} \quad (7.16)$$

The forecast accuracy of a product group G can then be defined based on the weight per product as

$$\text{forecast accuracy}_G = \sum_{p \in G} (APA_p \cdot \text{weight}_p) \quad (7.17)$$

This definition is robust against situations in which either the forecasted quantity or the actual quantity is zero. If for some product, there are no actuals ($x_t = 0$, e.g. consider a product with no customer orders in the respective time horizon), $\text{weight}_p > 0$ if $\hat{x}_t > 0$. The same proposition ($\text{weight}_p > 0$) holds if there is no

¹We omitted the indices t and r from x and \hat{x} in order to improve readability; precisely we should write $\hat{x}_{p,t,r}$ instead of \hat{x}_p and $x_{p,t}$ instead of x_p .

forecast or the forecast is zero ($\hat{x}_t = 0$) and $x_t > 0$. If both—actuals and forecast—are zero, the weight is zero and the product is not considered in the aggregated forecast accuracy.

Note that all aggregation types—by time, by product and by geography—can be combined; the respective formulas can easily be formulated by the reader. In many APS, macro languages are used to “customize” the aggregation schemes for the basic forecast accuracy metrics by all dimensions.

7.5.4 Forecast Value Added

If multiple departments are contributing to the forecast, the *Forecast Value Added* (FVA) can be measured (Gilliand 2002). This shows whether the effort of a specific step in the overall process pays off. Typically the first forecast (in most cases this is the automatic statistical forecast) is compared to a simple naive forecast. The naive forecast is simply created by using the most recent sales figure as forecast. Every successive step needs to improve (add value) to the forecast. As measure for the quality of a forecast one may use any suitable standardized forecast accuracy measure like MAPA. The FVA of the first forecast can then be calculated by subtracting the MAPA of the naive forecast from the MAPA of the first forecast. If this value is positive then one adds value by using the first forecast. This can be continued by measuring the MAPA of the revised forecast from marketing and comparing it to the MAPA of the statistical forecast and so on.

Based on the FVA the management can set targets and incentives for the participants in the forecasting process. This is a clear comprehensible system which contributes to the overall supply chain efficiency.

7.5.5 Biased Forecasts

In practice, sales (and other departments) tend towards overestimating future sales volume due to safety thinking. A larger forecast might lead to higher production and purchasing volume and—thus—to a better supply situation. This behavior results in a bias of the forecast which can be measured systematically, and based on that, controlled (and corrected). The bias can be measured by the mean deviation:

$$\text{mean deviation } MD_r = \frac{1}{n} \sum_{t=1}^n e_{t,r} \quad (7.18)$$

If $MD_r > 0$ the forecast is overestimated systematically. In this case, the forecast might be reduced by the bias MD_r , in order to “correct” the forecast and make it more realistic.

7.6 Additional Features

In this section additional features of demand planning based on APS are described, that have to be taken into account in order to address the specific demand planning needs of the supply chain.

7.6.1 Life-Cycle-Management and Phase-In/Phase-Out

In quite a lot of innovative businesses, like the computer industry, the life-cycles of certain components or products were reduced to less than a year. For example, high-tech firms offer up to three generations of a hard-disk every year. As common statistical forecasting procedures require significant demand history, it would take the whole life-cycle until useful results are gathered. But, since new products replace old products with almost the same functionality, it is plausible to reuse some information on the demand curve for the next generation.

Two main approaches are known in practice: The first one indexes the complete time series and determines the life-cycle-factor which has to be multiplied with the average demand to get the quantity for a specific period in the life-cycle (*life-cycle-management*). This method is able to stay abreast of arbitrary types of life-cycles. The only information needed for the application for new products is the length of the cycle and the estimated average demand. These two values are adapted continuously when observed demand data gets available during the “life” of the product.

The second approach (*phasing method*) divides the whole life-cycle in three phases. The “phase-in” describes the launch of a new product and is characterized by the increase of the demand according to a certain percentage (linear growth). Afterwards the series follows a constant demand pattern, as considered for the statistical forecasting procedures. During the “phase-out” the demand decreases along a specific percentage until the end of the life-cycle of the product. The only data necessary for the phasing model are the lengths of the phases and the in-/decrease-percentages.

For successful application of the above models it is necessary that the APS provides the functionality to build a “model library”. In this database life-cycles or phasing models are stored for each product group under consideration. Mostly only one life-cycle exists for the whole product group and this model is updated every time a life-cycle ends.

7.6.2 Price-Based Planning

In some industries—for example in the mineral oil industry—demand quantities strongly correlate with market prices. The quality of the products (different fuel grades, diesel, etc.) is fully specified and products from different suppliers can easily be interchanged. Second, there are spot market structures that make demand and

supply transparent, leading to a “free” formation of prices. Third, products can be stored, such that the demand for supply is—within certain bounds—dependent from the consumption of the products.

As a consequence, suppliers may sell nearly *any* quantity of their products—as long as they meet (or undercut) the current market price level. In this environment demand planning must include the planning of price levels, as prices are a major factor influencing demand. To include price planning into a demand planning environment, additional time series are needed:

- *Price levels*: There are multiple price time series that might be of interest for the demand planning process, e.g. market price, sales price, differential price (spread between market and sales price, might be negative!), average price level of competition, contracted prices.
- *Revenues*: The revenue can either be entered manually or computed by the product of price and quantity.
- *Exchange rates*: In international markets multiple currencies are involved (US Dollar, Euro, etc.). Usually, one currency is used as standard and all other currencies are transformed into the standard currency. For this purpose the exchange rates have to be known over time.

Incorporating these time series into a demand planning framework requires some further considerations. Only revenue and quantities can be aggregated—prices cannot be aggregated: What is the price of a product group G consisting of two products A and B , A having a price of 100 and B of 10? Clearly, the average price of the product group can only be computed from the aggregated revenue and the aggregated quantity:

$$\text{Price}_G = \left(\sum_{p \in G} \text{Revenue}_p \right) / \left(\sum_{p \in G} \text{Quantity}_p \right) \quad (7.19)$$

After prices have been aggregated to the higher levels of the demand planning structures, one might want to manually adjust prices on aggregated levels. As described in Sect. 7.2 time series may be disaggregated to lower levels of the demand planning structures using disaggregation rules. In order to bring the demand planning structures into a consistent state after the change of a price information, the following procedure can be applied:

1. Disaggregate the price time series to the lowest level using some disaggregation rule (e.g. based on existing price information on lower levels).
2. Adjust the revenue data on all levels by computing the product of quantity and updated price and store this value in the revenue time series.

Please note that price-based planning includes scenarios where the price may be influenced (e.g. by setting the market price for finished products) and the market demand results from that, scenarios where the price of raw material (e.g. crude oil) is set by suppliers, impacting the own product pricing and—by that—the resulting demand, and combinations of both scenarios.

7.6.3 Sporadic Demand

We call a time series sporadic (intermittent), if no demand is observed in quite a lot of periods. Those demand patterns especially occur for spare parts or if only a small part of the demand quantity is forecasted; for example the demand for jeans in a specific size on 1 day in a specific store might be sporadic. The usage of common statistical forecasting methods would produce large errors for those items. Additional judgmental forecasting would not increase the quality, because the occurrence of periods with no demand is usually pure random and therefore not predictable. Furthermore, sporadic demand often occurs for a large amount of C-class items, for which it would be appreciable to get forecasts with low time effort for human planners.

Efficient procedures for automatic calculation of forecasts for sporadic demand items were developed. These methods try to forecast the two components “occurrence of period with positive demand” and “quantity of demand” separately. For example, Croston’s method (see Silver et al. 1998 or Tempelmeier 2012) determines the time between two transactions (demand periods) and the amount of the transaction. The update of the components can then be done by exponential smoothing methods. Significant reduction of the observed error is possible, if the sporadic demand process has no specific influence which causes the intermittent demand pattern. For example, the frequent occurrence of stockouts in a retail outlet could produce a time series that implies sporadic demand.

7.6.4 Lost Sales vs. Backorders

Forecasts are usually based on the demand history of an item. But, while industrial customers (B2B) often accept backorders, if the product is not available, the consumer (B2C) won’t. Therefore, the amount of observed sales equals the amount of demand in the backorder case, but in the lost-sales case the sales figures might underestimate the real demand. For forecasting purposes the demand time series is needed and therefore, must be calculated from the observed sales figures. This problem frequently occurs, if forecasts for the point-of-sales (retailers, outlets) should be calculated.

There are two generally different solution approaches for the problem of forecasting in presence of lost sales: The first one tries to calculate a virtual demand history which is based on the sales history and the information on stock-outs. The forecasts can then be computed on the basis of the virtual demand history. This approach delivers good results, if the number of stock-outs is quite low. An alternative solution to the lost-sales problem is the usage of sophisticated statistical methods which consider the observed sales as a censored sample of the demand sample (see e.g. Nahmias 2009). For these methods it is necessary to know the inventory management processes which were/are applied for the products under consideration.

7.6.5 Model Selection and Parameter Estimation

The selection of the forecasting model and the estimation of the necessary parameters should be updated more or less regularly (e.g. every year) but not too often, as this would result in too much nervousness. APS often provide some kind of automatic model selection and parameter estimation. This is called *pick-the-best option*. The user only has to define the time-horizon on which the calculation should be based. The system then searches all available statistical forecasting procedures and parameter combinations and selects the one which produces the best forecast accuracy in the specified time-segment. As a result the user gets a list with the forecasting method and the corresponding parameters for each product/item he should implement. Therefore, the demand planner doesn't have to check if a model fits the time series under consideration (e.g. "Are the sales figures really seasonal or does the system only interpret random fluctuations?") and can use the toolset of statistical methods like a black box.

But, practical experience shows that the long-term performance is better and more robust, if only 1–3 forecasting methods with equal parameter settings for a group of products are applied. This follows from the following drawbacks of the described automatic selection:

- The time-horizon should cover enough periods to get statistically significant results. But often the history of time series is relatively short when demand planning is introduced first.
- The criterion for the evaluation is mostly one of the forecast accuracy measures described above. However, those values don't tell you anything about the robustness of the models' results.
- For the selection procedure three distinct time-segments are necessary: In the first segment the models components are initialized. For example for Winters' seasonal model 2–3 full seasonal cycles (e.g. years) are necessary to calculate initial values for the seasonal coefficients. The second segment of the time series history is used to optimize the parameter values. Therefore, the parameters are changed stepwise in the corresponding range (grid-search) and the forecast accuracy is measured. The optimized parameters are used to get forecasts for the third time-segment which is also evaluated using the forecast accuracy. This accuracy value is then the criterion for the selection of the best forecasting model.

The setting of the length of each of the time-segments has significant influence on the result of the model selection. Mostly the user has no possibility to change those settings or even can not view the settings in the software.

Therefore, the automatic model selection can guide an experienced demand planner while searching for the appropriate proceeding. But, it is not suggestive to use it as a black box.

7.6.6 Safety Stocks

Most APS-providers complement their demand planning module with the functionality for safety stock calculation. This is intuitive since the forecast error is one of the major factors influencing the amount of stock which is necessary to reach a specific service level. The calculation of safety stocks is quite complex, as there exist many different formulas each for a specific problem setting. The demand planner's task hereby is to check whether the prerequisites are met in his application. While this chapter cannot provide a fully detailed overview on safety stocks and inventory management, we want to focus on the functionality which can be found in most APS. For further information the reader is referred to one of the inventory management books by Silver et al. (1998) or Nahmias (2009).

Most software tools offer safety stock calculations for "single-stage inventory systems". This means that it is assumed that there exists only one single stocking point from which the demand is served. Multi-stage or multi-echelon systems (e.g. distribution chains with DC- and retailer-inventories) on the other hand have the possibility to store safety stocks on more than one stage.

For single-stage systems the amount of necessary safety stock ss is generally determined by the product of the standard deviation of the forecast error during the risk time σ_R and the safety factor k :

$$\text{safety stock } ss = k \cdot \sigma_R \quad (7.20)$$

Assuming that the variance of the forecast error in the future is the same as in the past, σ_R can be calculated by multiplying the standard deviation of the forecast error (calculated from past time series) σ_e with the square root of the risk time \sqrt{R} . The length of the risk time depends on the inventory management system. The following two systems have to be distinguished:

- *Periodic review system*: In such an environment the inventory position is reviewed only every t time periods (review interval). Each time the inventory is reviewed, an order is triggered and sent to the supplying entity (e.g. the production department, the supplier). The delivery is assumed to be available after the replenishment lead-time L . Therefore, the risk time equals the sum of the review interval and the replenishment lead-time: $R = L + t$.
- *Continuous review system*: In continuous review systems the point in time at which an order is released is triggered by a predefined reorder point. If the inventory position falls below the reorder point, an order of a specific quantity q is released. The risk time in a continuous review system equals only the replenishment lead-time L : $R = L$.

But that is only half of the safety stock formula. The safety factor k represents all other determinants of the safety stock. In the following the determinants and some of their values are explained:

- *Service level*: For the service level quite a lot of definitions exist. The most common ones are the following:

- Cycle- or α -service level: α is defined as the fraction of periods in which no stock-out occurs. Therefore, the safety stock has to ensure the probability (which fits the companies business objectives) of no stock-out during the replenishment cycle.
- Fill rate (β -service level): The fill rate is the order quantity of a product which can be met directly from stock.
- Order fill rate: While the fill rate considers the smallest unit of measurement of a product, the order fill rate counts *complete* customer orders served from stock.
- *Review interval or order quantity*: In periodic review systems the review interval is fixed and the order quantities depend on the estimated demand in an order cycle. For continuous review systems the opposite applies, as the order quantity is fixed and the length of the order cycle depends on the demand. But, if the demand is approximately level, both parameters can be converted in each other by the simple relation:
order quantity $q = \text{demand } d \cdot \text{cycle length } t$.

The required parameter can be calculated by minimizing the ordering costs and the holding costs for the lot-sizing stock. This computation can be made by applying the well-known economic order quantity (EOQ)-formula (e.g. Silver et al. 1998).

- *Demand distribution function*: The distribution function of the observed demand is usually approximated by a standard distribution known from statistics. One of the most common distribution functions is the normal distribution. The distribution parameters (mean and variance) can easily be calculated from a sample of demands from the historic time series.

All these parameters have to be combined in a formula which stays abreast the requirements of the business under consideration. Now, it should be clear that an APS-tool can only provide safety stock calculations if specific assumptions are met. But, if all parameters are user-definable the software can cover a wide range of different settings. Therefore, it is necessary to transfer the inventory management rules which are applied in the company to the standard parameters which are needed in the software. And that is the challenge of the demand planner.

This section on safety stock calculation gives only a short impression on the complexity of inventory management. The inspired reader can gather more information in one of the inventory management books listed below.

7.6.7 What-If Analysis and Simulation

In Sect. 7.1 the typical steps of a demand planning process are listed (see also Fig. 7.7 on p. 134): data collection, computation of further data, judgemental forecasting, creating a consensus forecast, planning of dependent demand, and release of the forecast to subsequent planning processes. In this workflow the final forecast is a result of the preceding steps—each step adding more information until the consensus forecast reflecting all information on the anticipated customer

demand is reached. In some cases the forecasting process does not only reflect the anticipation of customer demand but captures decisions to shape the demand. Examples of demand shaping decisions are

- Price promotions
- Bundling of products
- Regional promotions
- New product introduction.

In these cases it may be necessary to assess the impact of the demand shaping decision on the supply situation by means of master planning (Chap. 8), e.g. in a sales & operations planning scenario (Sect. 8.4). If, for example, a price promotion is planned for a specific product group, a higher demand on that product group might be forecasted (otherwise, reducing price as a promotional activity would make no sense). The forecast including the increased demand for that product group is passed on to master planning. If, for sake of the example, master planning detects a constraint—insufficient capacity or material availability to fulfill the increased constraint—, the supply plan will reflect a lower volume than the demand plan. In this case it would be reasonable *not to reduce price as a promotion* as there will be no increase of the sales volume due to supply restrictions. Thus, the workflow will go back to demand planning with the information that the increased forecast should be reduced and the price promotion should be canceled. This iteration loop is an example of a what-if analysis that helps to assess the impact of a demand shaping forecast decision.

The other examples in the list above also require what-if analysis and simulation of a shaped forecast by master planning including a feedback to the demand planning process:

- Consider a scenario with two products being bundled as a promotion to generate additional demand. If one of the products is not available in sufficient quantities, the effect on sales will be lower than planned for. The what-if scenario will reveal that additional demand forecasted due to the bundling-promotion should be reduced.
- A similar situation might occur with regional promotions of specific products. For example, in the car tires industry, a company wanted to strengthen sales in Spain by a marketing campaign. The additional demand of summer tires for Spain was planned in the forecasting process. Unfortunately, the campaign took place in September and October, when demand for winter tires in northern and central European countries was at a peak due to start of the winter season. Both product groups—summer tires and winter tires—share the same resources and an overload situation on these resources occurred. As winter tires have a higher margin than summer tires they are prioritized by master planning and all required resources are allocated to winter tires, reducing the supply of summer tires for the Spanish market. This information will be fed back to demand planning in order to adjust the timing or concept of the marketing campaign and the corresponding demand plan.
- The third example concerns a situation where a new product generation is being introduced to the market. This happens regularly in the electronics industry,

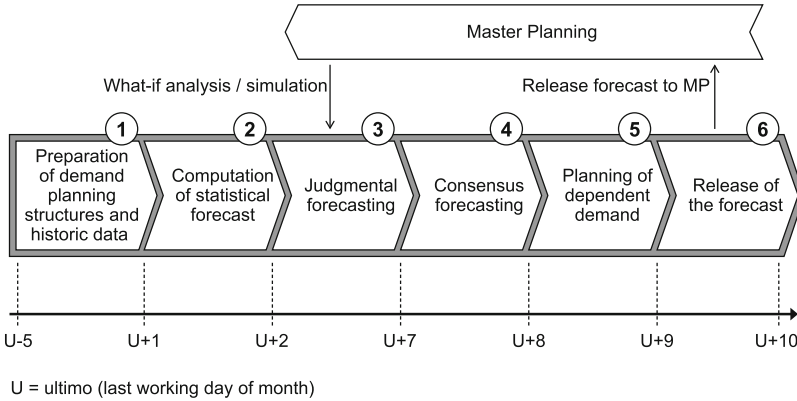


Fig. 7.10 Demand planning process with what-if analysis feedback loop

where products of a new product generation usually are cheaper and have better features than those of the preceding generation. In the demand planning process the demand of the products of the former generation will be reduced according to the phase out plan of those products, and the demand for the new product generation will increase over time. Assume that there is still a large inventory for some products of the old generation. In master planning we will detect the large period of cover for the old products. This might lead to a revised decision regarding the timing of the phase in of the new product generation and the phase out of the old one. This information will then be fed back to demand planning and the forecast will be adjusted according to the new timing of the introduction of the new product generation.

The examples illustrate that the feedback loop in a what-if analysis is required if the demand is actively influenced by business decisions, e.g. promotions and new product introductions. In this case the master planning process simulates the impact of the demand plan on the supply chain, indicating constraints, excess stock situations, lower supply than demand, etc. This information will be fed back to the demand planning process to be judged by the planners. As a result business decisions might be revised. After the step of judgemental forecasting, a new consensus forecast will be created, dependent demand will be planned, and the new forecast will be released to subsequent processes. Figure 7.10 summarizes the demand planning process with a feedback loop due to a what-if analysis.

References

Dickersbach, J. (2009). *Supply chain management with SAP APO* (3rd ed.). Berlin/New York: Springer.

Eickmann, L. (2004). Bewertung und Steuerung von Prozessleistungen des Demand Plannings. *Supply Chain Management*, 4(3), 37–43.

- Gilliand, M. (2002). Is forecasting a waste of time? *Supply Chain Management Review*, 6, 16–24.
- Hanke, J., & Wichern, D. (2014). *Business forecasting* (9th ed.). Harlow: Pearson Education Limited.
- Makridakis, S. G., Wheelwright, S., & Hyndman, R. (1998). *Forecasting: Methods and applications* (3rd ed.). New York: Wiley.
- Meyr, H. (2012). Demand planning. In H. Stadtler, B. Fleischmann, M. Grunow, H. Meyr, & C. Sürie (Eds.), *Advanced planning in supply chains — Illustrating the concepts using an SAP APO case study* (Chapter 4, pp. 67–108). New York: Springer.
- Nahmias, S. (2009). *Production and operations analysis* (6th ed.). New York: McGraw-Hill.
- Silver, E., Pyke, D., & Peterson, R. (1998). *Inventory management and production planning and scheduling* (3rd ed.). New York: Wiley.
- Tempelmeier, H. (2012). *Bestandsmanagement in Supply Chains* (4th ed.). Norderstedt: Books on Demand.
- Wright, G., & Goodwin, P. (1998). *Forecasting with judgment* (1st ed.). New York: Wiley.