RJ10379 (A0605-016) May 16, 2006
Other

IBM Research Report

Diagnosing and Tuning a Statistical Forecasting System: A Case Study

Ying Tat Leung
IBM Research Division
Almaden Research Center
650 Harry Road
San Jose, CA 95120-6099

Kumar Bhaskaran
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

Diagnosing and Tuning a Statistical Forecasting System: A Case Study

Ying Tat Leung
IBM Almaden Research Center
650 Harry Road
San Jose, CA 95120

Kumar Bhaskaran
IBM Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

April 2006

Abstract

Commercial forecasting systems are commonly used in manufacturing businesses to generate sales forecasts for thousands of products. These systems typically feature a number of built-in options to select and estimate the statistical forecasting model, from a pre-specified collection of models, and are targeted to provide a low cost and efficient method of forecasting. However, it is often observed that after prolonged use (e.g. two or more years), these systems suffer from serious performance degradation in terms of large forecast errors for a significant number of products. An immediate and tangible consequence of inaccurate forecasts is increase in inventory levels, rendering the entire supply chain less efficient. It is therefore important to diagnose and tune the statistical forecasting systems as part of their regular maintenance and operation.

This chapter describes a case study in which simple but useful tools were devised for the forecast practitioner to (i) diagnose a statistical forecasting system systematically and identify products that require forecast performance improvement; and (ii) tune the parameters of a statistical forecasting system to improve its overall forecast performance. The proposed tools were developed for a large manufacturer of consumer and industrial products in the USA.

1. Introduction and Background

Most, if not all, commercial enterprises require some form of demand forecasting for financial and operations planning. For financial planning, a high level forecast (e.g., in dollar value) of major product groups or geographies is sufficient. For operations planning, a more detailed forecast, such as forecast by product or even by product-location, is necessary. A manufacturing enterprise employing a make-to-stock strategy needs a demand forecast to plan what products and how much of each to build. A make-to-order manufacturer uses a demand forecast to plan the purchase of parts and materials and its production capacity. A retailer needs a demand forecast to determine how much of each product to stock at the different retail locations. Other service enterprises utilize a demand forecast to plan and locate their capacity (for both labor and equipment). We focus on the latter situation in this paper, namely detailed, product level forecasts that drive the planning of a supply chain.

Most often, a supply chain is engaged in the production of thousands or even tens of thousands of products, where each individual product may account for only a small portion of the total revenue. (See Fildes and Beard 1992.) Therefore it is neither practical nor economical to spend a lot of effort in forecasting a single product, except for the few top-selling products dominating a significant fraction of the total business. A relatively inexpensive and efficient way to forecast the sales of each of a large number of products repetitively is to use an automatic forecasting system. Such a forecasting system can be based on heuristics (e.g., as in "focus forecasting" (Smith 1991)), statistical methods (e.g., time series, regression; see Makridakis et al. 1997, Chapters 3-7), expert system like rules, or a combination. The most popular approach seems to be statistical methods, as evidenced by the large number of available software packages for statistical forecasting (Yurkiewicz 2004). Due to the large number of products involved, it is common that the system runs on a server computer and is part of a larger supply chain management system. In principle, these forecasting systems, once set-up, require minimum human intervention to operate and generate a forecast. This, however, only means that little human effort is required to produce some forecast; whether the forecasts are accurate or suitable for use is a different issue altogether. In addition, over time, the forecast performance of such a system tends to deteriorate if they are not diagnosed and tuned in a continuing basis. In this paper, we propose some practical ways to diagnose and tune the forecast performance of a typical statistical forecasting system under such an environment. Indeed, it has been shown, at least in one case of the Holt-Winters forecasting procedure, that an automatic version can be significantly improved by subjective modifications (Chatfield 1978).

We present a case study in which practical tools were developed to help an operations planner or a forecast analyst manage an automatic forecasting module of an integrated supply chain management system. Having gained wide acceptance in the last 15 years, these systems cover a wide range of supply chain management activities, including sales forecasting, inventory management, distribution requirements planning, master production planning, materials requirements planning, and even shop floor scheduling and control. Due to the enormous scope and the necessary complexity of such a system, the system designer has no choice but to limit individual modules to a relatively small set of key functionalities. In forecasting, for instance, there is usually very little provision for diagnosis and tuning such as forecast error analysis or parameter optimization. The simple but practical tools developed in this paper represent one way to fill this gap.

Specialized forecasting software provides more in-depth functionalities to forecast an individual product. Most notably, many of such specialized packages provide the ability to find optimal parameter settings of the chosen forecasting model (see Table of software survey in Yurkiewicz 2004 and other software surveys listed in Rycroft (1999)). But because of their standalone nature, significant integration effort has to be spent in utilizing these packages to develop an overall system suitable for managing thousands of products on a daily basis. For this reason they are not nearly as popular as forecasting modules of integrated supply chain management systems.

In this study, our manufacturing enterprise produces industrial and consumer items of a commodity and semi-commodity nature. It adopts a make-to-stock strategy and utilizes a demand forecasting module of a commercially available supply chain management system. At any one time there are more than 3000 active products for which demand forecasts are made regularly. We assume that this forecasting module will continue to be used in the future. The complex issues of whether there is much to be gained by switching to a specialized, standalone forecasting system, what kind of models should be considered, and how the system fits in the business process of forecasting are outside the scope of this work. (See, for example, Chambers et al. 1971, Jenkins 1982.) Our objective is to improve the forecast performance of the system in use. To this end, we assume that the principal quantitative measure of forecast performance is the mean squared error of the one-period-ahead forecast of the recent past 12 periods, i.e., we attempt to minimize this mean squared error. A key reason to use the mean squared error is that the direct cost of the forecast error is inventory in the supply chain, since the root mean squared error is typically used to calculate the necessary safety stock to serve a demand point.

Note that we use the terms "sales" and "demand" interchangeably. Strictly speaking, we should be performing demand forecasting as sales may be influenced by the actions or constraints of the business itself.

2. Role of Forecasting in Production Planning

The role of forecasting in production planning is characterized here in general in the context of supply chain planning and execution. Forecasting processes play a critical role in the demand planning function in a supply chain, which specifies how much demand we would like to satisfy, where, and at what selling price.

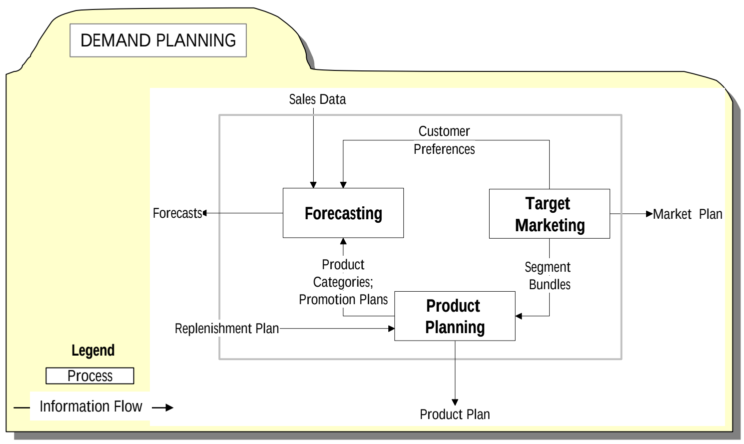


Figure 2.1: Business Processes in Demand Planning

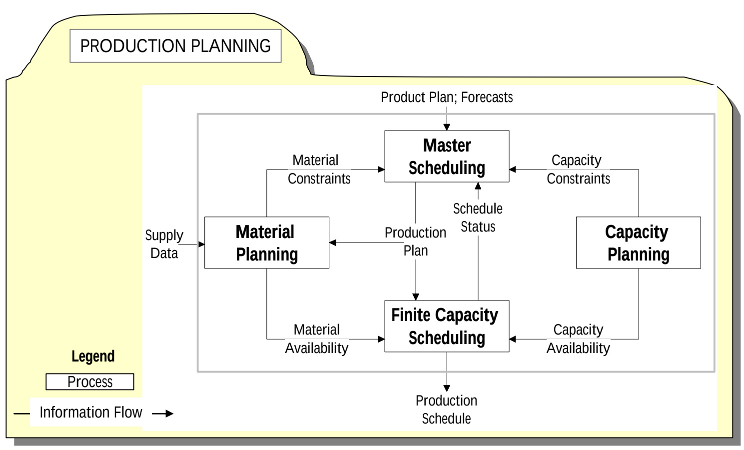


Figure 2.2: Business Processes in Production Planning

Demand planning is an aggregation of forecasting, target marketing, and product-planning processes (see Figure 2.1). The forecasting processes use sales data such as point-of-sales scanner data, customer and market segment information from target marketing, promotion plans and product categories from product planning to generate demand forecasts. These dependencies also highlight the collaborative nature of the forecasting activity in an enterprise and typically involve many role players in a supply chain such as the forecast analyst, the promotion planner, and the merchandizing manager. Forecasting processes are also supported by a number of demand-planning related analytics such as analysis of promotional elasticity and sensitivity of sales to promotional events, analysis of demand for products that are considered as slow movers, impact of clearance pricing, data mining of customer buying patterns etc.

The Master Scheduling process, in the production-planning function, uses the forecasts, the material constraints from material planning processes, the capacity constraints from the capacity planning processes, and the current schedule status from the finite capacity scheduling process to formulate the production plan (see Figure 2.2).

3. Practical Issues in Using a Statistical Forecasting System

A statistical forecasting system typically contains forecast procedures that determine how the user can interact with a forecasting model to generate a forecast. A "user" in this context refers to someone who is knowledgeable about the data and is using the mathematical methods contained in the statistical forecasting system to obtain forecasts. More importantly, the user is typically very knowledgeable about the manufacturing operation and the business, but is not someone who is an expert in statistics or mathematical modeling. More often than not, the users regard a forecasting model as a black box. How then do they use the system? In some cases the forecasting system vendors or system-integration consultants fine tune the software and set the parameters during installation, and provide users with canned recipes which they can then use to operate the system. In other cases the users install the system using default parameters programmed into the system by the vendor and let the system work. Ad hoc methods are then devised as needed to adjust the parameters, often with unpredictable but significant consequences on the supply chain performance. The forecasting system manuals accompanying the software generally offer very little help to the user. The software vendors encourage the formation of "user groups" which allow users to exchange information. Such a forum is useful to resolve procedural issues but do not provide technical information on what parameters or which forecast models are relevant for the data at hand. The diagnostic and tuning procedures reported here were developed to help the user maintain the forecasting system so that forecasts of acceptable quality can be obtained throughout the life of the system.

It is important to distinguish what is done here as part of the maintenance of a forecasting system and what is typically meant by "maintenance" from the system point of view. The latter includes maintaining the required data and the software itself. This is usually recognized as a significant effort and financial and human resources are allocated for it. Maintenance as referred to here is forecasting model maintenance which is different and complementary to the system maintenance effort. At present, model maintenance is unfortunately not routinely budgeted as regular activities. Consequently the forecast accuracy decreases over time. In some cases, this will eventually lead to the perhaps unjustified conclusion that the system is no longer adequate; a new forecasting system is purchased and the cycle starts all over again. Regular model maintenance will help avoid pre-mature abandoning of existing forecasting systems in which significant investments have already been made. It will also help direct efforts towards situations where more advanced models or methods are really needed while leveraging well known and well proven methods for most of the forecasts. Gung et al. (2002) discusses reasons and opportunities for new forecasting models.

Model maintenance is covered in the subsequent sections. Next we discuss some issues that typically occur in the ongoing application and use of forecasting systems. Addressing these issues involve subjective judgments and governance policies associated with forecasting in supply chain management.

Product lifecycles, especially short lifecycles, complicate the application of forecasting systems in at least two ways. First, the system has to support new product introductions. Complex, manual procedures to create a new product in the forecasting system can become too expensive very quickly. Second, in order to be able to obtain some reasonable forecast at the beginning of the product's life, it is common to borrow the sales history of another "similar" product or that of the predecessor product. If the new product has significant new features or if the market has changed (e.g., with the entry of a new competitor), using the history of the predecessor product may be questionable. Finding a similar product is not trivial. Subjective judgment often plays an important role. The forecasting system should at least provide reasonable reporting or data visualization to support the user in making the necessary judgments.

Even for mature products, market conditions can change over its lifecycle such that the historical sales data of the product need to be adjusted to reflect the changing conditions. For example, in a wholesale business, a significant customer is gained or lost; in a retail business, a new market segment may be gained through product bundling, or a new competitor has entered the market. Custom procedures to adjust the product history have to be devised based on the individual situation. Once again, subjective judgment plays an important role.

Supply chains feature multiple products and/or multiple geographic regions (or customer segments). This is set up as a product hierarchy in forecasting systems to support various business processes including marketing, financial planning, and production planning. Different levels of the hierarchy are used for different purposes. For example, for financial planning one uses the forecasts at the top one or two levels (most aggregated); for detailed production planning the forecasts at the bottom one or two levels (most detailed, say the bottom level being the SKU(Stock Keeping Unit)-location) are used. Clearly the forecasts within the entire hierarchy need to be consistent, in the sense that the sum of the lower level forecasts is equal to the higher level. In one common approach, the forecasting systems derive the forecast for each product or each node in the hierarchy using an independent model, then enforce consistency through taking the upper level as the reference forecast and splitting that forecast based on the lower level forecasts. Fliedner (2001) reviews different approaches in forecasting a hierarchy. Such a hierarchy makes the forecasts of the different products dependent on each other, even though the underlying forecasting models for each product may be independent. A consequence is that when some products are changed (e.g., new product introduction or product discontinuation), the entire hierarchy has to be re-run from the beginning of time in the forecasting system. Besides the additional effort required for re-running the forecasts, the new forecasts will not be identical to the historical ones, causing potential confusion. For example, historical forecasts on which business decisions had been made cannot be traced back and analyzed.

4. Diagnosis of the Statistical Forecasting System

4.1. Diagnosis of the overall system

The first question we ask of the performance of a statistical forecasting system is that how it compares to the case of no forecasting. Even though an automatic forecasting system is used, time and other resources are needed to manage the input data and the system itself. Therefore we expect some return on our investment. If we do not wish to perform any serious forecasting, one of the simplest and least expensive ways to obtain some input for planning is to use historical sales data of the recent past. For example, we can take the average and standard error of sales in the past 12 months as the forecast and the basis for safety stock calculation. Indeed, an informative way to evaluate the forecast performance is to perform such a comparison, using a histogram of the coefficient of variation (CV) of the historical sales and one-period-ahead historical forecasts[1], over the past 12 months, of all the products we are trying to plan. The cumulative plots of the same quantities indicate the percentiles. Figures 4.1 and 4.2 contain an example of these plots[2].

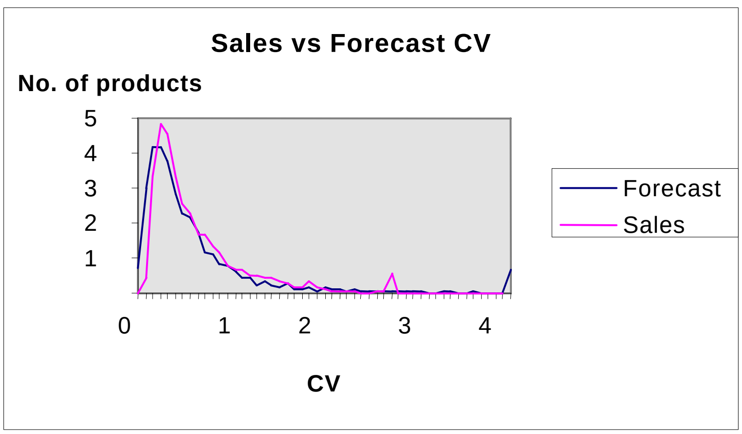


Figure 4.1: Sales and Forecasts CV Distribution Plot

[1] The coefficient of variation (CV) of sales is its standard deviation divided by its mean. The CV of the forecasts is defined to be the root mean squared error of the forecasts divided by the average forecast.
[2] The data used in all the examples of this paper were taken from a system in real life but were arbitrarily rescaled.

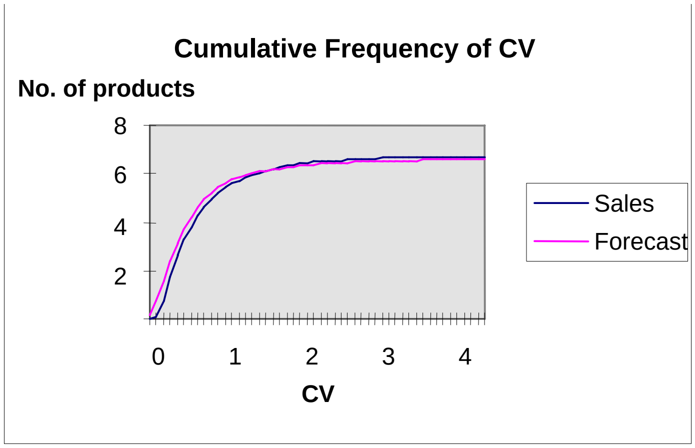


Figure 4.2: Cumulative Plot of Sales/Forecasts CV

From Figure 4.1, we can see that the forecasting system did shift the distribution of CV to the left, showing a general improvement (reduction) over the raw sales. However, the shift is relatively small, meaning that the improvement is not large. This is confirmed in Figure 4.2. Ideally, a good forecasting system should produce a steeper cumulative plot of forecast CV than that of sales.

The CV analysis can be repeated with deseasonalized sales data. An important usage of this CV analysis is that it shows a profile of the products in terms of their suitability or necessity of forecasting using a statistical model. The products with very high sales CV (outliers or the top few percentiles) are naturally difficult to forecast statistically with good accuracy. For these products, we may have to seek alternative methods of planning without relying heavily on sales forecasts. For example, if the sales volume of the product is low or the unit value of the product is very low, then one way to plan production is to follow a cyclic schedule with relatively large production runs (say, one production run per half year or a year). Then little to no safety stock will be necessary due to the large order size but the stocking cost is still low because of the low volume or low unit value.

For products with a small CV of sales (e.g. less than 0.2 for raw sales data and less than 0.1 for deseasonalized data), there is no need to use a sophisticated statistical model for forecasting. We can simply take the moving average of historical sales in the past several (up to 12) months and its standard deviation. Even if we had made good effort to forecast the sales statistically, the improvement in accuracy (in terms of reduction in CV) would not have been significant in practice.

The remaining products with sales CV's falling in the middle range are the ones for which we should try to tune the statistical forecasting models. They are most likely to result in significant forecast accuracy improvements with our effort; in other words, they give the best return of our investment in time and other resources.

The CV analysis also identifies which products need to be improved in terms of forecast accuracy. Excluding products with the largest or smallest sales CV as explained, we can start with products with the largest 20% of all forecast CV's and cross-check with our high volume or high dollar value products. (Similarly, we can cross-check with strategically important or other management selected products.) If there are products falling in both categories, these are the products with the most urgent need of forecast improvement. If there is no product in both categories, we can take the next lower 20% of all forecast CV's and so on. Once the products to be considered are selected, we next investigate their forecast performance.

4.2. Diagnosis of individual products

4.2.1. Sales Analysis

For a selected product, analysis of the historical sales data can give some insight into how a statistical forecasting model would behave. For example, if the historical sales is smooth, then we know a simple method like the moving average would be quite accurate and a very sophisticated or computationally intensive model is not necessary. Clearly, sales data analysis is useful in the selection of statistical models also; but we restrict our attention to whether our forecasting system in use (assuming its existence) is adequate. To start the sales data analysis, we consider the standard descriptive statistics of the sales data over the full historical horizon and the past 12 periods separately: sample mean, standard deviation, coefficient of variation, mean absolute deviation, mean absolute deviation as a fraction of the sample mean, the quartiles, the minimum and maximum. Besides the obvious patterns these quantities indicate, they are very useful in seeing quickly whether the past data are reasonable. For example, if the second or the third quartiles show zeroes, then we know that over half of our data are zero. This may mean that we do not have much past data, or our past sales were indeed very sporadic. In the former case, we can shorten the historical data horizon and in the latter case find a model which is particularly designed for intermittent demands (e.g. Johnston and Boylan 1996). In any case, the point of this exercise is to diagnose - to raise warning flags. The difference in the statistics between the full horizon and the recent past indicates the trend in the data pattern and whether our collection of data has improved so that a smaller horizon may lead to better forecasting results.

These simple methods are also useful during diagnosis of the overall system in the following way. Instead of presenting screens of historical sales data across products to the analyst, a table of selected summary statistics, say the second quartile and difference of it between the full horizon (e.g. 36 months) and most recent past (e.g. 12 months), across products is shown. In this way, it would be easier for the analyst to obtain general insight on the products and their relative behavior. Further, automatic warning procedures can be easily programmed to monitor a few summary statistic values.

For an overview of the data, it is useful to plot the historical sales data as a function of time for visual detection of data pattern and the data's general behavior. If the randomness of the data is not too high, then patterns like seasonality and increasing/decreasing trends can be detected by visual inspection. Otherwise, an autocorrelation plot of the full data set and data of the recent past can help in detecting seasonality in the presence of noise.

Figure 4.3 shows our implementation of the above sales data analysis in a spreadsheet. The tools provided are meant to help indicate the general pattern of the historical sales data, from which we may infer whether the class of statistical forecasting models we are using is appropriate (i.e., overkill or inadequate) or, moving back one step, whether we should use statistical forecasting for this product at all. Assuming that the answers to both questions are in the affirmative, we may move on to analyze the forecasts made by the automatic forecasting system for this product.

4.2.2. Forecast Analysis

A natural first step is to consider a time plot of the historical sales and forecasts, together with the future forecasts, as shown in Figure 4.4. The general forecast-vs-sales pattern in the recent past (e.g. past 12 periods) provides a visual representation of major forecast inadequacies, if any, such as bias, time lag, over-reaction, over-damping effects, or missing seasonality. Standard descriptive statistics such as the average, standard deviation, mean absolute deviation, mean absolute deviation as a fraction of the average, quartiles, minimum and maximum of the recent past historical forecasts are useful for comparison with that of historical sales. For example, if sales are relatively stable (statistically stationary), there should be a reasonably close match of the averages and standard deviations. Matching the average is particularly important since a yearly or other aggregate forecast is often used for financial planning purposes.

One way to analyze the performance of the forecasting system is to see whether the forecasted change in sales from one period to the next matches well with the actual change. We use three tools for this purpose: the correlation coefficient, Theil's U-statistic, and Prediction-Realization (P-R) diagram for the forecasted and actual changes in the recent past. The correlation coefficient measures our forecasted changes against a perfect forecast. This measure gives some information on the absolute performance of our forecasts. However, a perfect forecast is unrealistic and we need an additional measure based on a more practical baseline. Theil's U-statistic is one such measure.

Let f_i and s_i be the forecast and actual sales for period i respectively. Theil's U-statistic (Theil 1966, pp26-32) is defined as

$$U = \sqrt{\frac{\sum_{i=1}^{n-1} (FPE_{i+1} - APE_{i+1})^2 \,/\, (n-1)}{\sum_{i=1}^{n-1} (APE_{i+1})^2 \,/\, (n-1)}},$$

where

$FPE_{i+1} = \dfrac{f_{i+1} - s_i}{s_i}$ is the forecasted relative change, and

$APE_{i+1} = \dfrac{s_{i+1} - s_i}{s_i}$ is the actual relative change.

This statistic essentially provides a comparison of the current forecasting method with the naive method of using the current period's sales as the forecast for the next period. When we use the naive method, $FPE = 0$ and $U = 1$. On the other hand, if the forecasted changes are perfect ($FPE = APE$), then $U = 0$. If we are using a computationally intensive method of forecasting but getting a value of U close to 1, we are not getting very good value of our forecasting investment.

The Prediction-Realization (P-R) diagram (see Figure 4.5) offers a pictorial view of the forecasted vs actual changes. Clearly, we would like all the points to lie in the first or third quadrant of the graph. A perfect forecast has all points lying on the $y = x$ diagonal line. Points appearing in the second or fourth quadrant indicate that even the direction of change was predicted incorrectly in those instances.

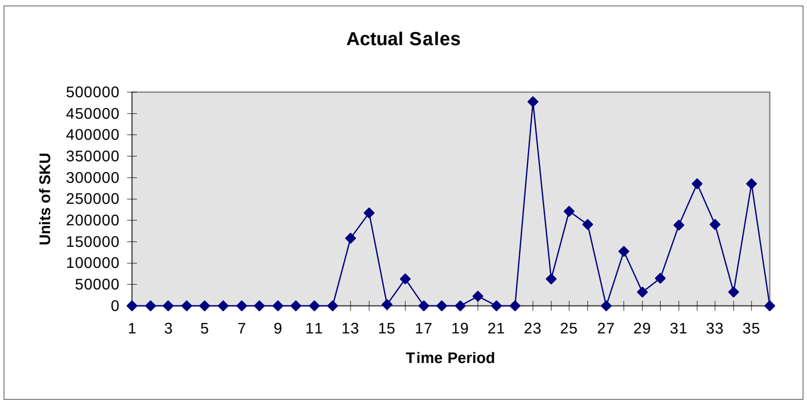
To determine whether the forecasting model is missing significant patterns in the historical sales, we use two simple tools: a time plot of the forecast errors and a plot of the autocorrelation coefficient of the forecast errors. If the forecasting model is adequate, then the forecast errors are random noise with no particular pattern. In addition, the autocorrelation coefficients should be small, showing a random pattern across lags. The bottom half of Figure 4.5 shows an example of these plots.

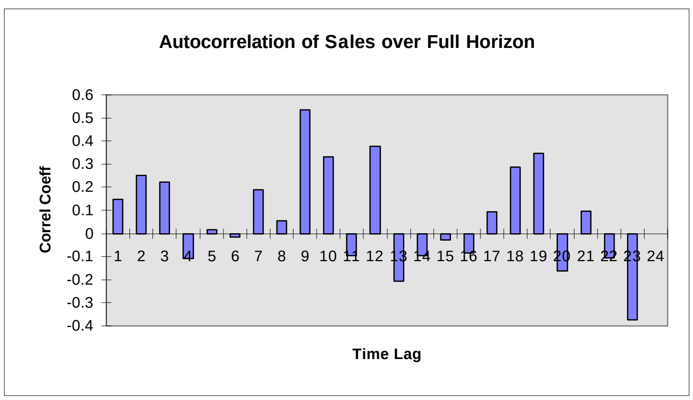
Descriptive statistics of sales per period over full horizon:

Average	72808.00	Minimum	0
Std deviation	114632.85	1st quartile	0
Coeff of variation	1.5745	2nd quartile	0
Mean absolute deviation	89607.11	3rd quartile	134640
MAD / Average	1.2307	Maximum	477360

Descriptive statistics of sales per period over last 12 periods:

Average	134640.00	Minimum	0
Std deviation	106318.25	1st quartile	31680
Coeff of variation	0.7896	2nd quartile	157680
Mean absolute deviation	92160.00	3rd quartile	198000
MAD / Average	0.6845	Maximum	285120





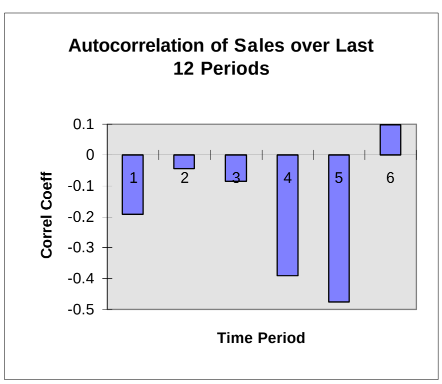


Figure 4.3: Sales Data Analysis

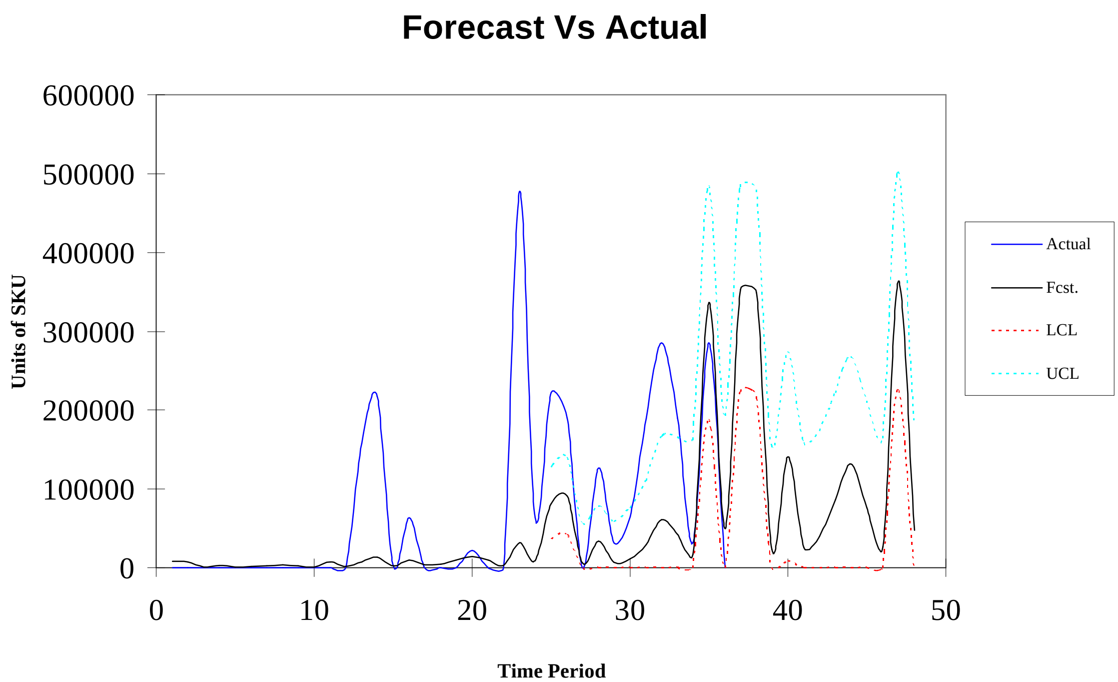
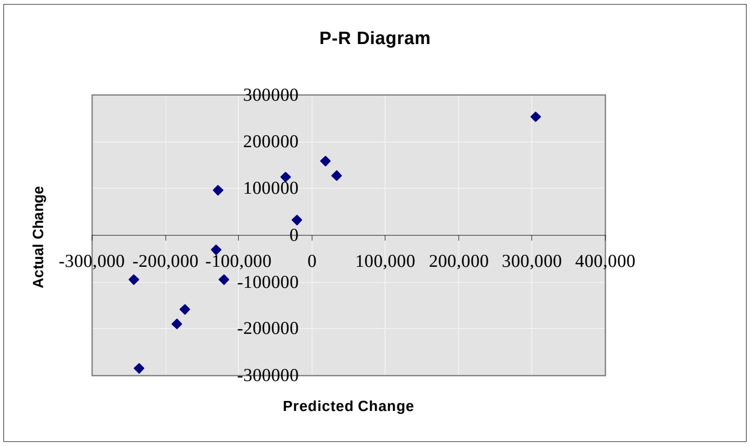
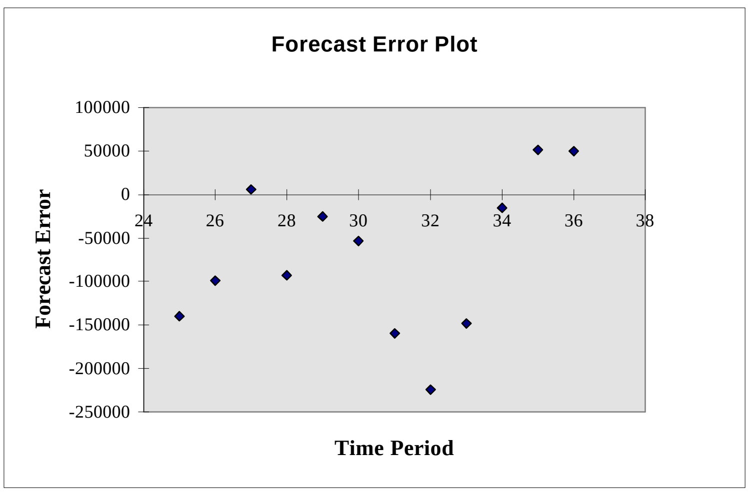


Figure 4.4: Sales/Forecasts Time Plot

Descriptive statistics of sales forecast per period over forecast horizon:			
Average	138605.29	Minimum	23638
Std deviation	136133.77	1st quartile	36294
Coeff of variation	0.9822	2nd quartile	81296
Mean abs deviation	108588.43	3rd quartile	192814
MAD / Average	0.7834	Maximum	363411

Correlation of forecast & actual change	**0.8473**
Theil's U-statistic	**0.4956**





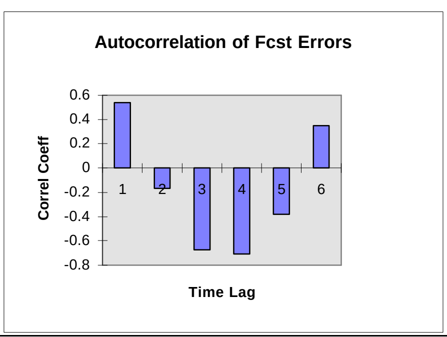


Figure 4.5: Forecast Analysis

5. Tuning of the Statistical Forecasting System

Based on an analysis of the forecasts as discussed in Section 4, one can determine whether to continue using the current forecasting model. Most often, the analysis results would be reasonably satisfactory for some products, but also identify other products as candidates for improving their forecast accuracy. The simple tool developed for tuning the parameters of the forecasting model was applied to a version of the well-known Winters' method (see, e.g., Makridakis et al. (1997), p.161), which is used in the supply chain planning system under study. We note however that the concepts and the tuning process are generally applicable for most statistical model based forecasting systems.

Winters' method represents the family of exponential smoothing models, a robust approach commonly used in practice. Exponential smoothing is featured in every one of the forecasting and supply chain planning systems available commercially (Yurkiewicz 2004). Sanders (1997) found from his survey of 350 U.S. manufacturers that 26% of low sales (<$100M) firms and 38% of high sales (>$500M) firms used exponential smoothing. In an older but more general survey of 500 U.S. corporations (not just manufacturers), Sanders and Manrodt (1994) reported that exponential smoothing was used by 13-20% of the respondents in forecasting up to 1 year ahead. Exponential smoothing has also performed favorably in past forecasting accuracy empirical studies (Makridakis and Hibon 1979, Makridakis et al. 1982). Because of its fast speed of computation, it is popular among retailers who have to generate a large number of forecasts (up to millions) on a regular basis. For a comprehensive review of exponential smoothing and related issues, see Gardner (1985).

In the remainder of this section, our discussion assumes the use of Winters' method. Our objective is to find the values of the model parameters such that the mean squared forecast error is minimized. It addresses the issue of choosing the smoothing parameters in a discussion of practical use of Winters' method in Chatfield and Yar (1988).

5.1. Tuning of individual products

Although dedicated forecasting software packages usually provide some parameter optimization capability (Yurkiewicz 2004), the forecasting modules of many integrated supply chain management systems do not provide such capability or such capability is very limited. For example, they may only have built-in criteria for parameter optimization such that the user cannot specify their own criterion. Choosing the appropriate criterion to measure forecast error is a non-trivial and important issue that impacts the performance of the business (see, e.g., Lee et al. 1993). A custom-built parameter optimization tool, such as the simple tool described here, can accommodate criteria that are chosen deliberately for the business.

The main parameters in Winters' method are the three smoothing factors (one for each of the permanent, trend, and seasonality components; see the Appendix for details). We developed a simple three-option tuning process for the nonlinear optimization problem of minimizing the mean squared forecast error. (An implementation in Excel is shown in Figure 4.6.) Each option is more refined than the previous one, producing an improved parameter set, but requires a larger amount of computational time. The options are:

1. Grid search. It simply performs a search over a grid formed by the feasible ranges of a given set of parameters and chooses the point which yields the lowest mean squared error of the historical forecasts in the recent past (say 12 periods). We chose the feasible parameter set in the following way.

   The smoothing factors control the "responsiveness" of the forecasting model. When the factors are large (i.e., close to 1), the model reacts quickly to recent changes in sales. The forecasts may be more accurate, but are usually less robust in the long run. With large smoothing factors, the forecasts for a given future period generated in successive periods may be very different. This can create a certain degree of "nervousness" to the production plan driven by the forecasts. An organization needs to determine the extent of plan changes its production-distribution system can handle. For example, if the system is flexible because of investments in the production facilities, then we can use a larger smoothing factor to take advantage of the system flexibility. Otherwise, it may not be very useful to have highly adaptive forecasts while the production-distribution plans cannot be changed frequently. In the latter case the organization pays the price of having higher inventory costs. In addition, the nature of the business influences whether we need a responsive forecast. When sales are relatively stable, a large smoothing factor may actually decrease the forecast accuracy because the forecast will be too sensitive to occasional fluctuations. It is therefore prudent that the range of the smoothing factors be chosen based on the knowledge of the business and the capability of the organization.

The chosen range limits the grid search to a smaller set than the unit cube (which is the entire feasible set for the three smoothing factors). For our implementation we used a maximum of 0.45 for all three factors. The specific parameter sets for the grid search are then chosen by uniformly placing 125 points in the feasible cube. Obviously, the larger number of points will give better results but require more computational time. For our implementation, we find, after some experimentation, that 125 seems to be a reasonable number in terms of tradeoff between computational time and forecast error.

Grid search, although somewhat old fashioned and rather cumbersome, is a general approach for parameter optimization in Winters' forecasting method (Chatfield and Yar 1988).

2. Grid search with fine tuning. This option first runs a grid search as described, then takes the selected parameter set from the grid search as a starting solution for a nonlinear optimizer to find the final parameter values. In our implementation we used the Microsoft Excel Solver with a nonlinear optimizer based on the GRG2 algorithm (Lasdon et al. 1978).

3. Multiple-start optimizer for global optimization. Because the mean squared forecast error function of the three smoothing factors is likely to contain many local extrema, we couple the nonlinear optimizer with multiple starting points to attempt global optimization. We run the optimizer to find a local minimum using each point selected for the grid search described in option 1 as a starting solution. The final parameter set is chosen to yield the least of all the local minima. From our experience, when the sales data were fairly smooth (e.g., the sales data for a large class of products in the national market), the multiple-start optimizer would produce the same solution as the grid search with fine tuning. In other words, the GRG2 algorithm was able to find the global optimum with a single starting point.

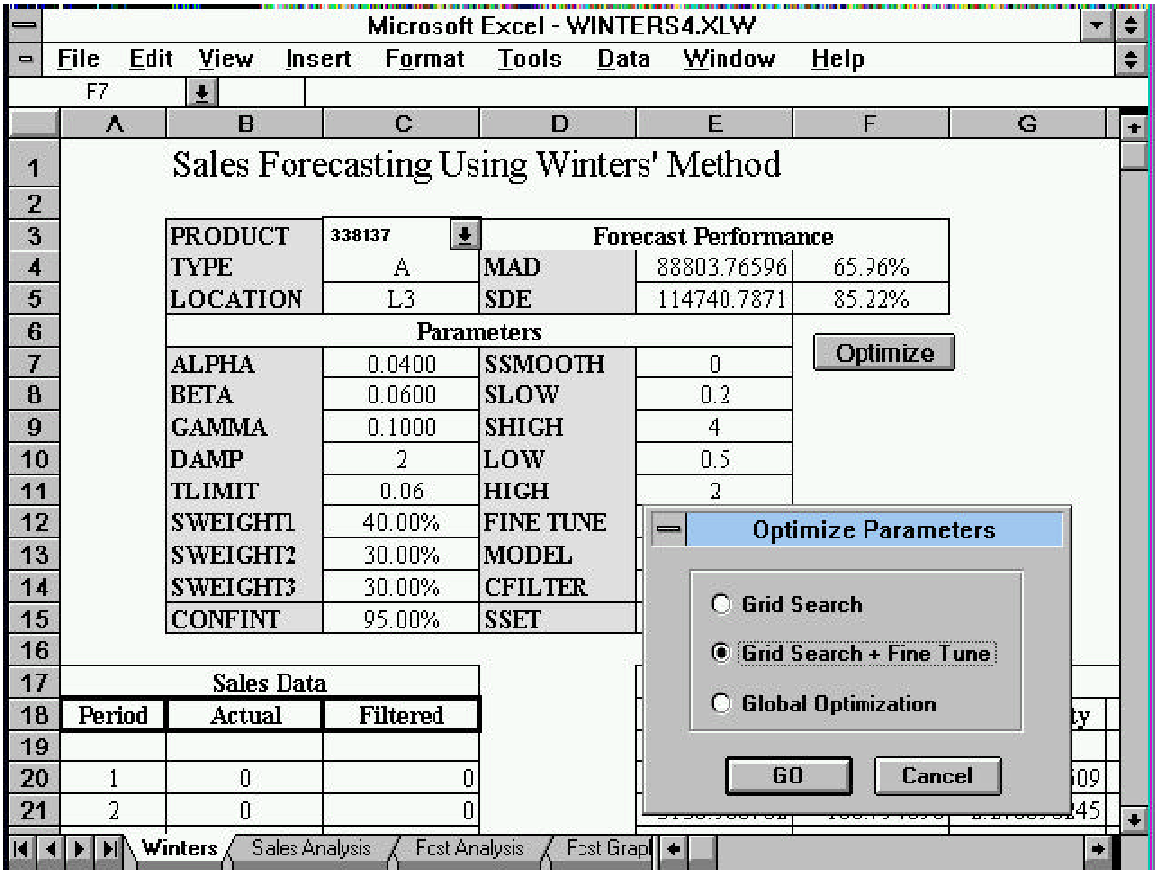


Figure 4.6: Implementation of tuning method in Excel spreadsheet

5.2. Tuning of the overall system

In the previous section, we discuss how we can gain some insight about the behavior of the forecasting model and tune it for a single product. Now we need to find a way to tune the system as a whole. The simplest but the most cumbersome way to optimize the system is to perform the above tuning process for each product and then upload the resulting parameters to the forecasting system. As we are dealing with many products, this may not be practical on a regular basis. We developed an alternative scheme as follows.

First, products that are important to the business need to be optimized individually. Often, these include the top sellers (or class A products in a Pareto classification of sales), products with consistently high inventory levels in the past (which can be identified using a Pareto analysis of the total inventory), products strategically important to the survival or growth of the business (e.g. products with advanced proprietary technology), and products which consume the largest amounts of the most critical resources of the business. In addition, forecast parameters for each of the large families of products (i.e. the top levels in the forecasting pyramid) should be tuned individually. In this way, we are assured that the overall volume of the business is forecasted reasonably well.

For the rest of the products, we need to classify them into a manageable number of categories according to the sales pattern of the products. (The actual number of categories will depend on the resources available for forecasting and the relative importance of the sales forecast to the business.) For example, we can group all products whose sales are known to be highly seasonal or, in the other extreme, very “flat”, products in the same stage in their life cycles (new products, mature products, etc.), products which are by their nature often sold together, products sold according to some common industry practice. For each of the category, we choose a representative product and optimize its forecasting parameters as above. These parameter values are then used for all products in the same category.

In general, parameter tuning need not be performed very frequently. For the Winters' method, Chatfield and Yar (1988) suggested updating the parameters every year or two. For the types of products in hand, we also feel that once a year is a reasonable interval. Many forecasting systems have built-in forecast error tracking, so that the system will issue a warning when its tracking criterion is violated. Even with such features, we still suggest that a yearly parameter tuning activity be used. This will ensure that we are getting the most of the forecasting system and will also serve as a forecast monitor as well.

6. Conclusions

The use of automatic demand forecasting modules in the context of an integrated supply chain management system provides many advantages and is now widely popular. Maintaining the forecasting system, however, takes more effort than routine data and software maintenance. It requires regular effort in maintaining the forecasting models throughout the life of the system and such activities need to be planned and budgeted for.

We illustrate the type of model maintenance that is needed through a case study. Simple tools were developed to diagnose where forecast accuracy improvements may lie and to subsequently tune the forecasting models in these cases. The tools developed complement the statistical forecasting capabilities provided by a commercial supply chain management system. While the latest forecasting software packages may provide similar parameter optimization capabilities, we have not seen any that provides similar forecast diagnostic functions. Our work also accompanies the list of features described by Fildes et al. (2003) as desirable design points of a forecasting system.

References

Chambers, J.C., S.K. Mullick, and D.D. Smith (1971). “How to choose the right forecasting technique,” *Harvard Business Review*, July-August 1971, 45-74.

Chatfield, C. (1978) “The Holt-Winters forecasting procedure,” *Applied Statistics* 27, 3, 264-279.

Chatfield, C. and M. Yar (1988). “Holt-Winters forecasting: Some practical issues,” *The Statistician* 37, 129-140.

Fildes, R. and Beard, C. (1992). “Forecasting systems for production and inventory control,” *International Journal of Operations and Production Management* 12, 4-27.

Fildes, R., Goodwin, P. and Lawrence, M. (2003). “Design features of Forecasting Support Systems and their effectiveness,” Lancaster University Management School Working Paper 2003/066, *Decision Support Systems*, forthcoming.

Fliedner, G. (2001). "Hierarchical forecasting: Issues and use guidelines," *Industrial Management & Data Systems* 101, 1, 5-12.

Gardner, E.S., Jr (1985). "Exponential smoothing: The state of the art," *Journal of Forecasting* 4, 1, 1-28.

Gung, R.R., Y.T. Leung, G.Y. Lin, and R.Y. Tsai (2002). "Demand forecasting TODAY," *OR/MS Today*, December 2002.

Jenkins, G.M. (1982) "Some practical aspects of forecasting in organizations," *Journal of Forecasting* 1, 31-49.

Johnston, F.R. and J.E. Boylan (1996). "Forecasting for items with intermittent demand," Journal of the Operational Research Society 47, 113-121.

Lasdon, L.S., A.D. Waren, A. Jain, and M. Ratner (1978). "Design and testing of a generalized reduced gradient code for nonlinear programming," *ACM Transactions on Mathematical Software* 4, 34-50.

Lee, T.S., F.W. Cooper, and E.E. Adam, Jr. (1993). "The effects of forecasting errors on the total cost of operations," *Omega International Journal of Management Science* 21, 5, 541-550.

Makridakis, S. and M. Hibon (1979). "Accuracy of forecasting: An empirical investigation (with discussion)," *Journal of the Royal Statistical Society* (A) 142, 97-145.

Makridakis, S. et al. (1982). "The accuracy of extrapolation (time series) methods," *Journal of Forecasting* 1, 111-153.

Makridakis, S., S.C. Wheelwright, R.J. Hyndman (1997). *Forecasting: Methods and Applications*, 3rd Edition, John Wiley.

Rycroft, R.S. (1999). "Microcomputer software of interest to forecasters in comparative review: Updated again," *International Journal of Forecasting* 15, 93-120.

Sanders, N.R. (1997). "The status of forecasting in manufacturing firms," *Production and Inventory Management Journal*, Second Quarter, 1997, 32-36.

Sanders, N.R. and K.B. Manrodt (1994). "Forecasting practices in US corporations: Survey results," *Interfaces* 24, 2, 92-100.

Smith, B.T. (1991) *Focus Forecasting and DRP*, Vantage Press, New York.

Theil, H. (1966) *Applied Economic Forecasting*, North Holland Publishing Co., Amsterdam.

Yurkiewicz, J. (2004). "Forecasting: Predicting your needs," *OR/MS Today*, December 2004.

Appendix: Forecasting Processes & Winters' Forecasting Method

The Forecasting Process

The forecasting method is best understood in the context of the overall forecasting process. Figure A.1 provides a simple functional overview of the overall forecasting process. It has essentially two main components: a hierarchical statistical forecasting process and a management interaction to adjust the statistical forecasts based on business considerations (also referred to as management overrides). The resultant forecast is then what primes the production and distribution planning processes. The ensuing description is focused on the hierarchical statistical forecasting process.

The hierarchical statistical forecasting is based on independently forecasting each level of the forecast pyramid, i.e., using independent forecasting models, and then reconciling the forecasts across the levels using a mechanism referred to as "forcing". First, the basic statistical forecasting model used for any single entity at any level (for example, an item at level L3 or an item at a particular stock location at level L2) is described. This is followed by a description of the forcing mechanism.

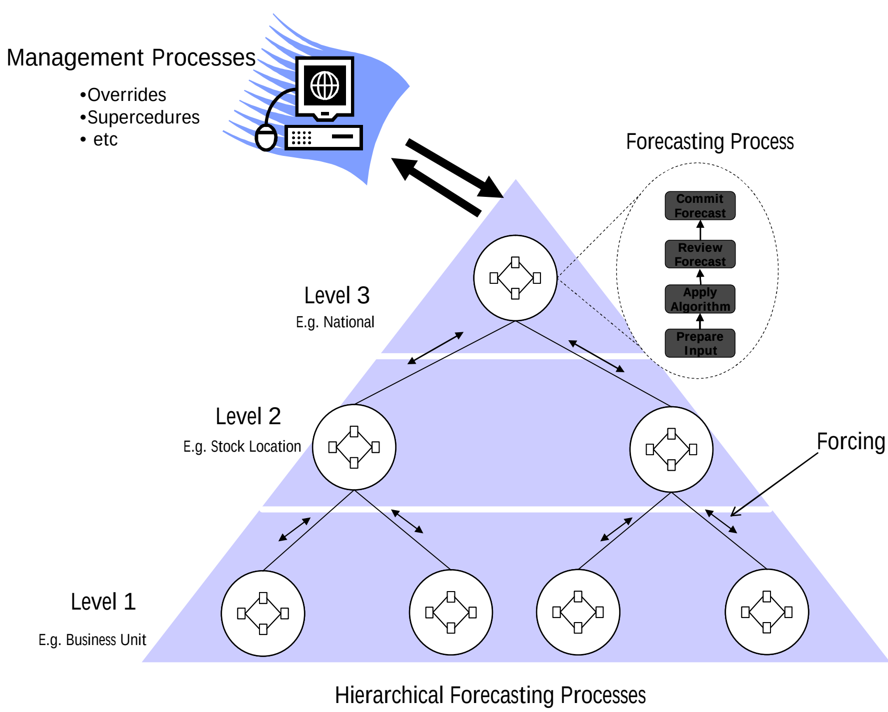


Figure A.1: Functional Overview of Overall Forecasting Processes

The Statistical Forecasting Model

The statistical forecasting model considered here is an embellishment of the basic Winters' exponential smoothing forecasting technique. Exponential smoothing is commonly used in forecasting product demand that is then used as input to production and distribution planning in manufacturing and supply chain management systems. Among exponential smoothing techniques the Winters' method (with extensions) is typically used to account for trend and seasonal variations in demand.

Assumptions

The description of the forecasting model is predicated on the following assumptions.

1. The forecasting system filters the actual demand to remove outliers based on a parameter referred to as the "demand filter factor" and the mean absolute deviation of the demand. It is assumed for describing the forecasting model that the actual demand has been filtered and is available as data. The filtering procedure is therefore outside the scope of the forecasting model described here.
2. The forecasting model is described assuming that the model is RESTARTED (i.e., all historical and future forecasts are computed from scratch using the entire array of historical data available). When the model is REINITIALIZED (i.e., historical forecasts computed previously are retained; future forecasts are computed from scratch using the entire array of historical data available), the model is applicable except that part of the forecast array still contain previously computed values. It may be noted that in this case the standard deviation of the error is based on the previously computed forecast values (F_t). When the model is operating under NORMAL condition (i.e., neither restarted nor reinitialized) the algorithmic step of generating future forecasts using the Winters equations is applicable. However, all other steps of the algorithm are in such a case irrelevant.
3. The forecast periods are assumed to be uniform, i.e., either they are of same time duration or can be considered as such. This assumption must be relaxed to accommodate the calendar effect such as (4-4-5 week pattern) calendar.

The Forecasting Algorithm

A1 **Regression**

A two-phase regression approach is performed on the actual demand to calculate *the initial permanent and trend components*. In the first phase, the actual raw demand is deseasonalized. A simple regression is then performed on the deseasonalized raw demand to fit a trend line. This trend line represents the fitted deseasonalized demand and yields the initial permanent and trend components. A double moving average method (also referred to as the centered moving average) is used in deseasonalizing the raw demand. This procedure is similar to that contained in the Census II procedure employed by the U.S. Department of Commerce.

A2 **Computing Seasonality**

If the model is seasonal, then the deseasonalized demand from step A1 is used to determine *seasonality ratios* for periods {t=1, 2... N}; these ratios are adjusted if warranted to comply with the lower and upper limits on seasonality (SL and SU respectively). They are then weighted using seasonality weights (W_y) and normalized to obtain *normalized seasonality factors* for periods {t=1, 2…12}. If seasonality smoothing indicator $\gamma > 0$ then an additional step is carried out to smooth these seasonality factors based on either a three or a five month moving average. If the model is non-seasonal, then all the seasonality factors are directly set equal to 1 bypassing the procedure described above.

A3 **Generating Historical Forecast**

Using the initial permanent and trend components from A1 and the seasonality factors from A2, Winters' basic equations are applied to generate the historical system forecast $\{f_t, t=1, 2\ldots N\}$. Since the model is assumed to be restarted (by assumption 2), $[F_t = f_t, \forall t]$. The following steps are carried out as the Winters' equations are used to roll the initial permanent and trend components from period to period. The permanent and trend components in period t=N are referred to as the *current permanent and trend components* respectively.

- Enforcing Non-negativity
  If the permanent component in any period is less than zero then it is set equal to zero.
- Checking Trend Limit
  The trend limiting factor χ is used to check if the trend component is within the pre-specified percentage of the permanent component. If the limit is violated then the permanent and the trend components are adjusted.
- Computing Seasonality Factors
  In the case of seasonality model, basic Winters' smoothing equation is applied on seasonality factors computed in A2 to obtain seasonality factors for the following 12 periods. These are then used to compute the seasonality factors for the next 12 periods and so on. Upon computation of each set of 12 seasonality factors the following step is executed.
- Adjusting Computed Seasonality Factors
  The seasonality factors are adjusted if they violate the lower (**SL**) and the upper (**SU**) limits. They are then normalized and additionally smoothed if $\gamma > 0$.

A4 Generating Future Forecast

- Using the current permanent and trend components as well as the seasonality factors (in the case of seasonal model) from A3, modified Winters' overall forecast equation is applied to determine the future forecast. The modification to the Winters' equation includes the *trend dampening factor* τ. The following additional step is carried out to scale the forecasts if necessary.
- Checking Forecast Reasonableness
  If the total forecast demand for the future twelve periods exceed (fall short of) the product of the forecast reasonable factor HI (LO) and the total forecast demand for the recent past twelve periods, then the forecast demand for the future twelve periods are appropriately scaled.

The Forcing Process

The forcing process is an important element of the hierarchical forecasting process and applies to both statistical forecasts as well as to any management overrides of forecasts in slightly different ways. Recall that for any item, the forecast pyramid consists of three levels: level L3 is the apex of the pyramid and represents for example the national forecasts, level L2 is the next lower level and represents the forecasts for example at all stocking locations for the item, and level L1 is the lowest level and represents the forecasts at all stocking locations for each business unit for the item. Forcing affects the forecasts at all three levels and is meant to produce a consistent forecast for items at all three levels. The forcing process is described for the case when there are no management overrides. The presence of the management overrides involves more complex calculations.

When there are no management overrides the forcing is applied to the system forecasts (generated by Winters' method described above). This typically happens at the end of the month forecast calculations. Let f_t, $f_{t(j)}$, and $f_{t(j,k)}$ be the unforced system forecast in period t (independently generated) for any item at levels L3, L2, and L1 respectively, where index j stands for the stocking locations and k stands for the business units. Let ST be the set of stocking locations for the item (i.e. $j \in ST$) and BU be the set of business units (i.e. $k \in BU$). Since the statistical forecasts are usually more accurate at the upper levels of the pyramid than the lower levels, forcing proceeds downwards, i.e., forecasts are "forced down." The forced forecasts (ff) at each level are then given as follows:

$$ff_t = f_t$$

$$ff_{t(j)} = f_{t(j)} * [ff_t / \Sigma_{(j \in ST)} f_{t(j)}]$$

$$ff_{t(j,k)} = f_{t(j,k)} * [ff_{t(j)} / \Sigma_{(k \in BU)} f_{t(j,k)}]$$

Following forcing the total forecast for the item at all three levels will be consistent:

$$ff_t = \Sigma_{(j \in ST)} ff_{t(j)} = \Sigma_{(j \in ST)} \Sigma_{(k \in BU)} ff_{t(j,k)}$$