RC23407 (W0411-028) November 3, 2004
Mathematics

IBM Research Report

Forecasting Demand for IBM Semiconductor Products

Aliza R. Heching, J. R. M. Hosking, Ying Tat Leung
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Forecasting demand for IBM semiconductor products

Aliza R. Heching, J. R. M. Hosking and Ying Tat Leung

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

Abstract. We describe a procedure for forecasting demand for all of the semiconductor products of IBM Microelectronics Division (MD). We outline some of the particular problems in forecasting demand in the semiconductor industry, and we give considerable thought to how to define demand and how to measure forecast accuracy. Based on these considerations we develop an approach, called FIT, that uses simple time-series models to forecast demand for individual products; modifications permit intervention by users of the forecasts and ensure consistency of forecasts throughout a hierarchy of products. FIT forecasts have been extensively compared with those obtained from MD's existing forecasting process. We find that for lead time of three months or more, FIT forecasts are significantly more accurate than MD's forecasts, with forecast errors smaller by factors typically in the range 20% to 55%.

1 Introduction

In this paper we describe a forecasting method, the Forecast Improvement Tool ("FIT"), that we developed to help IBM's Microelectronics Division make forecasts of demand for its products. IBM Microelectronics Division manufactures thousands of products, and demand forecasts are needed for each of them. The products are generally arranged according to one or more product hierarchies. A marketing hierarchy represents an organization along the lines of business or financial responsibilities. A technology hierarchy represents an organization of the products along the lines of technology uniqueness and differentiation. At the highest level of the product hierarchy, products are grouped into product families; this grouping is generally determined by technology or by the intended uses for the products within each product family.

We focus on the implementation of FIT in three product families. The families were selected because they represent respectively custom products, commodity products, and products whose uniqueness lies between these two extremes. (For this latter case, most of the underlying intellectual property is owned by IBM, but the final product is owned by the customer.) The importance of selecting products that span the range between custom and commodity products is to ensure that the forecasting methods remain valid for all these types of products.

IBM Microelectronics Division's forecasting processes follow a monthly cycle and forecasts of monthly demand, updated monthly, are of prime interest. Forecasts are required of demand for individual products and for aggregations of products that correspond to nodes in the product hierarchy. FIT provides point forecasts and prediction intervals of monthly demand at each node of the product hierarchy. The basic approach uses independent time-series forecasts of demand at each node. The point forecasts are modified so that forecasts at different levels of the hierarchy are mutually consistent. The forecasts can be overridden by users of the forecasts, who are sales representatives and sales executives and may have detailed knowledge of customers' purchasing intentions that cannot be easily incorporated into an automatic forecasting procedure.

The structure of the paper is as follows. Section 2 discusses the nontrivial issue of how to define demand, and gives the definition that we use. Section 3 describes the nature of demand data in the semiconductor industry, and the challenges that it raises in obtaining reliable forecasts. Section 4 considers how best to measure forecast accuracy, another issue that needs careful attention in order to ensure that forecasts are useful for making business decisions. Section 5 describes how the overall forecasting procedure was developed, and lists the steps that comprise the final procedure. Section 6 shows how the forecasts have been used on IBM data, and compares the accuracy of the new forecasts with those of Microelectronics Division's existing processes. Section 7 summarizes our conclusions.

2 What is demand?

The difficulty in forecasting demand for all products, in particular those with constrained supply, begins with the question of how demand should be defined. A seemingly innocuous question, the problem of how demand is defined is actually quite complex. Should demand be defined as the actual final order placed by the customer (with its assigned ship date and order quantity)? Consider the following scenario: A customer calls an order contact center. The customer places a request for 500 units to be shipped on January 15, 2004. We refer to January 15, 2004 as the request date. The sales representative checks an automated inventory system and finds that there is not enough inventory (finished or scheduled for production) to meet the customer request. The sales representative responds that the customer can either order 500 units to be shipped on February 1, 2004 or can place an order for 300 units to be shipped on January 15, 2004. In response, the customer may (i) cancel his entire order and place an order with a competitor supplier, (ii) insist that the product be shipped on the original request date but place an order for 300 units instead of 500 units, or (iii) be willing to delay his order for two weeks and place an order for the originally requested volume of 500 units, however, these units will only be shipped on February 1, 2004. If demand is defined by the pair (assigned ship date, final order quantity) then in case (i) demand is zero, in case (ii) demand is (January 15, 2004, 300), and in case (iii) demand is (February 1, 2004, 500). One would be hard pressed to argue that any of these accurately reflects customer demand. In fact, if these values were used to predict future demand, it would result in underestimated future demand and the supplier will again be faced with inventory shortages for future customer demand. This example illustrates a phenomenon that the literature commonly refers to as censored data. By censored data we refer to the problem that when the customer service representative records zero demand (as in case (i)), demand for 300 units on January 15, 2004 (as in the case of case (ii)), or demand for 500 units of February 1, 2004 (as in the case of (iii)), true customer demand (for 500 units on January 15, 2004) is hidden. The data pertaining to the original customer request is not recorded anywhere.

The order process at IBM Microelectronics Division is similar to the one described above. The customer calls a customer order center with a demand described by a (request date, request quantity) pair. The sales representative responds with one or more possible (ship date, order quantity) pairs. The customer will either cancel his request or accept one of the (ship date, order quantity) pairs suggested by the sales representative. While the customer's original request quantity is not recorded in the sales system, the request date is recorded. Thus, demand (for the purposes of historical demand data) is defined by the pair (request date, order quantity), i.e., the timing of true customer demand is retained, but the magnitude of this demand may be distorted if IBM Microelectronics Division does not have enough capacity to meet all of the customer's original request for product. Given the data retained in the

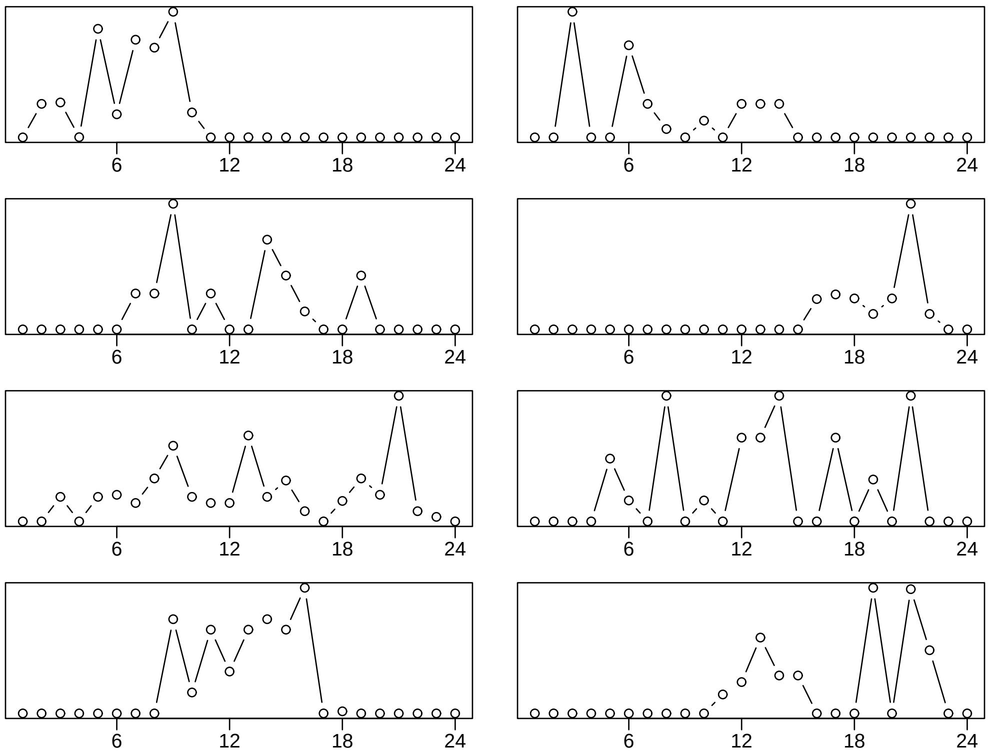

Figure 1: Monthly demand for some typical semiconductor part numbers.

sales systems, the pair (request date, order quantity) is the most accurate definition of demand possible.

3 Challenges in forecasting semiconductor demand

The semiconductor industry is a very challenging area in which to make statistical forecasts of demand for products, and IBM Microelectronics Division is no exception. Several aspects of the data make forecasting difficult. First, demand is highly variable and erratic. Some typical monthly demand series for individual part numbers are shown in Figure 1. Demand is very erratic, with large changes in volume from month to month. For demand at the part-number level, we found that the coefficient of variation of the monthly demand series was typically about 1.5 and the lag-1 autocorrelation was 0.15. (These values are averages over series that had at least a 10 month period between the first and last months with nonzero demand). Even at the product-family level, where the demand is aggregated over several hundred part numbers, the coefficient of variation is still typically about 0.5 and the lag-1 autocorrelation does not exceed 0.4.

The nature of the product life cycles also makes forecasting difficult. The lifespan of a product rarely exceeds 24 months and is much less in many cases. The life cycle rarely displays identifiable periods of growth, maturity and decline. The graphs in Figure 1 illustrate some typical patterns: demand may increase sharply when a product is introduced, it may decrease to zero within a month or two when a product reaches the end of its life, and there may be periods of low demand during a product's maturity, all of this superimposed on a high level of month-to-month variation.

At higher levels of aggregation, demand data series for product families have several years of available history. However, the semiconductor industry is evolving rapidly, and there must be doubt about whether relationships in the data that were valid three or four years ago are still useful for making forecasts today. There may have also been changes in the business process, e.g. the order-placement process and the organization of the sales force, that could affect the dependencies in the data that can be useful for forecasting. We therefore prefer to use a relatively short period of history as the basis for our forecasts. We judged that 24 months of history would be a suitable amount; this amount is subject to change, depending on business circumstances.

IBM Microelectronics Division manufactures thousands of products, and demand forecasts are needed for each of them. In the management of the business, the products are arranged in a hierarchy with each node in the hierarchy corresponding to the business responsibility of a manager or executive. Products are also grouped into families by technology or by the purpose for which the products are intended: there are several dozen of these product families. We use a set of hierarchies, one for each product family. In each hierarchy the top level is the product family, intermediate levels correspond to the business line managers' responsibilities, the penultimate level contains the individual products (part numbers) and the lowest level corresponds to combinations of part-number and customer. A typical hierarchy has eight levels and 500 nodes. Forecasts are required of demand at each node of each hierarchy, and each hierarchy of forecasts must be internally consistent, with the forecast at a parent node being equal to the sum of the forecasts at its child nodes.

Some part numbers have essentially a single customer, and even at the product-family level a few customers may account for the vast majority of sales. The demand patterns for these "dominant customers" can be quite different from those of the smaller customers: the dominant customers have their own buying patterns, e.g. a customer may tend to place large orders in particular months of the year. These customers' orders and potential requirements are closely monitored by the sales force, thus the sales representatives' judgemental forecasts often have greater accuracy than any statistical forecast could hope to achieve. Within each product family, we found it useful to construct separate hierarchies for each of the family's dominant customers and for all other customers combined.

Forecasts are needed for a range of lead times, and forecasts for different lead times have different uses. Forecasts at lead time 1 month influence short-term marketing decisions. Lead times of two or three months are the most important, and are used both in marketing and in production planning. Longer lead times (6–12 months) can be used for strategic marketing decisions production facilities' planning. For planning purposes, it is advantageous if forecasts of demand in a given time period are reasonably stable, e.g. this month's 3-month-ahead forecast should not be too different from last month's 4-month-ahead forecast. This constraint might be imposed on the forecasts when they are first computed, but we found it more convenient to apply the constraint afterwards: if a forecast of demand for a given time period is not markedly different from the previous month's forecast of demand in the same time period, it is reset to be equal to the previous forecast. The reasoning behind this is that forecasts should change only when there is a significant change in the business environment; fluctuations small enough to be regarded as noise should be ignored.

The nature of the data inevitably influences the choice of a forecasting procedure. Because there are many demand series to be forecast, there is insufficient time or resources to develop time-series models individually for each series. Because demand series at the part-number level are short, calibration forecasting models with many parameters is not feasible. It seems best therefore to concentrate on simple forecasting procedures, perhaps combining the forecasts given by a small number of such procedures. We adopted this approach.

4 Measuring forecast accuracy

4.1 What is forecast accuracy?

Irrespective of the forecasting procedure selected, the forecaster must determine a method for measuring the accuracy of the forecasts generated. Selection of an appropriate measure often depends upon the nature of the demand data (e.g., relative volumes of demand if forecasting demand for more than one product) or the objective of interest (e.g., minimizing variance in forecast revenue). Obtaining a single overall accuracy measure for a forecasting procedure is also a particular challenge with semiconductor demand data owing to the hierarchical structure of the data, the large number of products whose demand must be forecast, and the wide range of level and variability of demand between different products.

We begin this discussion by more explicitly defining what is meant by forecast accuracy. Let $F_{h,i,j,k}$ denote the forecast of demand in month k made at the end of month j for part number h in product family i. Often, companies will forecast demand for month k as early as twelve to twenty-four months prior to month k, with monthly updates of this demand forecast until month k is reached. For example, demand for the month of December 2004 will first be forecast in January 2004, and updated in each month February 2004,...,November 2004. We refer to the number of months between month j and month k as the forecast lead time. For example, the forecast made at the end of November 2004 for demand in December 2004 has a one month forecast lead time.

Let $A_{h,i,k}$ denote the actual order volume for part number h in product family i observed in month k. As discussed in Section 2, we use orders as a proxy for true demand, since true (uncensored) demand often cannot be measured for reasons such as back orders, lost orders due to stock-outs, etc.

We define forecast accuracy as a measurement of the difference between $F_{h,i,j,k}$ and $A_{h,i,k}$. For any single product, forecast accuracy can be measured in different ways. For example, one can measure the accuracy of all forecasts that have a forecast lead time l? Or, one can measure the accuracy of all forecasts made for demand in month k (irrespective of each forecast's specific forecast lead time)? The choice typically depends upon the use of the forecast and how forecast inaccuracy will affect the business. In our case, we choose to select a fixed forecast lead time l and measure forecast accuracy for all forecasts made with lead time l. We compute accuracy measures for each part number-customer combination (the lowest level of the product hierarchy) and at each level of the product hierarchy up to the product family level. We assume that average accuracy of forecasts with forecast lead time l made for any single product is computed using a simple average across the time series. Let T denote the time horizon for each of these time series (i.e, T is the number of data points in these time series). The question of how to aggregate forecasts over multiple products, to produce a single measure of forecast accuracy for a set of products, is addressed below.

4.2 Forecast accuracy measures

We consider a number of different accuracy measurements and then determine those that are most appropriate given the nature of the forecast data and actual data and the intended uses for the data. Table 1 lists the different accuracy measurements considered. We define each of these measurements for forecast lead time l over a time horizon consisting of T periods.

We now provide explanations for each of the measurements described in Table 1 and point out potential pitfalls with using some of the measurements.

MAE is a measure of the absolute value of the differences between the forecast and actual demand values. It is difficult to compare MAE values across different levels of a product hierarchy, because the presence of bias in the forecast will greatly affect whether MAE increases or decreases with the level of aggregation. Further, the magnitude of MAE values for different products is significantly affected by the demand volume associated with each product. MARE represents an alternative measurement to MAE which avoids the impact of forecast bias. However, if the value of $A_{h,i,k}$ is zero and the value of $F_{h,i,k-l,k}$ is greater than zero, MARE will take on a value of infinity. Thus, MARE may not always be a useful measure of forecast accuracy, depending on the nature of the data. RMAE is a measure that avoids the possibility of obtaining a value of infinity for forecast accuracy unless, at the level of aggregation being considered, all demand observations are equal to zero. However, the usefulness

Table 1: Some forecast accuracy measurements.

	Technique	Accuracy Measure
MAE	Mean Absolute Error	$(1/T)\sum_{k=1}^{T} \lvert A_{h,i,k} - F_{h,i,k-l,k}\rvert$
MARE	Mean Absolute Relative Error	$(1/T)\sum_{k=1}^{T} \lvert A_{h,i,k} - F_{h,i,k-l,k}\rvert / A_{h,i,k}$
RMAE	Relative Mean Absolute Error	$(1/T)\sum_{k=1}^{T} \lvert A_{h,i,k} - F_{h,i,k-l,k}\rvert \Big/ (1/T)\sum_{k=1}^{T} A_{h,i,k}$
SMARE	Symmetric Mean Absolute Relative Error	$(1/T)\sum_{k=1}^{T} \frac{\lvert A_{h,i,k} - F_{h,i,k-l,k}\rvert}{\frac{1}{2}(A_{h,i,k} + F_{h,i,k-l,k})}$
SMARE′		$(1/T)\sum_{k=1}^{T} h(A_{h,i,k}, F_{h,i,k-l,k})$ where $h(0,0) = 0$ and $h(x,y) = \lvert x-y\rvert/\{\frac{1}{2}(x+y)\}$ if $x \neq 0$ or $y \neq 0$.
SMARE″		$(1/T_+)\sum_{k\in\Gamma} \frac{\lvert A_{h,i,k} - F_{h,i,k-l,k}\rvert}{\frac{1}{2}(A_{h,i,k} + F_{h,i,k-l,k})}$ where Γ is the set of time periods for which $A_{h,i,k} + F_{h,i,k-l,k} > 0$ and T_+ is the number of such time periods.
LMAE	Logarithmic Mean Absolute Error	$(1/T)\sum_{k=1}^{T} \lvert\log(A_{h,i,k}/F_{h,i,k-l,k})\rvert$
OLMAE	Offset Logarithmic Mean Absolute Error	$(1/T)\sum_{k=1}^{T} \left\lvert\log\left(\frac{c + A_{h,i,k}}{c + F_{h,i,k-l,k}}\right)\right\rvert$
TLMAE	Truncated Logarithmic Mean Absolute Error	$(1/T)\sum_{k=1}^{T} r(A_{h,i,k}, F_{h,i,k-l,k})$ where: $r(x,y) = \min(\lvert\log(\frac{A_{h,i,k}}{F_{h,i,j,k}})\rvert, \log C)$ if $x > 0$ and $y > 0$; $r(x,0) = \log C$ if $x > 0$; $r(0,y) = \log C$ if $y > 0$; $r(0,0) = 0$.
RMSE	Root Mean Square Error	$\sqrt{(1/T)\sum_{k=1}^{T}(A_{h,i,k} - F_{h,i,k-l,k})^2}$
WMAE	Weighted Mean Absolute Error	$\sum_{k=1}^{T} w_k \lvert A_{h,i,k} - F_{h,i,k-l,k}\rvert \Big/ \sum_{k=1}^{T} w_k$
WSL	Weighted Skew Loss	$\sum_{k=1}^{T} w_k L(A_{h,i,k} - F_{h,i,k-l,k}) \Big/ \sum_{k=1}^{T} w_k$ where $L(0) = 0$ but, in general, $L(x) \neq L(-x)$.

of RMAE as a measure of accuracy suffers from the same limitation as MAE in that its value is dominated by the contributions of the data series with the largest volumes. We now point out an additional potential limitation associated with using RMAE or MARE as measures of forecast accuracy. $A_{h,i,k}$ is present in the denominator of both of these measures, rendering the measures asymmetric with respect to the values of $F_{h,i,k-l,k}$ and $A_{h,i,k}$. More specifically, MARE and RMAE will take on large values in the case that $F_{h,i,k-l,k} > 0$ and $A_{h,i,k} = 0$, but not when $F_{h,i,k-l,k} = 0$ and $A_{h,i,k} > 0$. This asymmetry implies that a positive demand forecast when actual demand is zero is more harmful to the business than a forecast of zero when actual demand is positive.

If such asymmetry is inappropriate, SMARE provides an alternative but similar measure of forecast accuracy which is symmetric in $F_{h,i,k-l,k}$ and $A_{h,i,k}$. Further, when relative forecast errors are small, SMARE takes on a value approximately equal to $|A_{h,i,k} - F_{h,i,k-l,k}|/A_{h,i,k}$. SMARE is bounded; its maximum value of 2 is attained when, for each time point k, either $F_{h,i,k-l,k}$ or $A_{h,i,k}$ (but not both) equals zero. When both $F_{h,i,k-l,k}$ and $A_{h,i,k}$ equal zero, SMARE can be adjusted in one of two ways: (i) SMARE$'$ takes on the same value as SMARE when $F_{h,i,k-l,k}$ or $A_{h,i,k}$ exceeds zero. In the case that both $F_{h,i,k-l,k}$ and $A_{h,i,k}$ equal zero, SMARE$'$ equals zero. (ii) SMARE$''$ represents a second modification of SMARE. SMARE$''$ ignores all periods when both $F_{h,i,k-l,k}$ and $A_{h,i,k}$ equal zero. The determination of whether to use SMARE$'$ or SMARE$''$ hinges upon the decision of whether a forecast should be considered "accurate" if both the forecast value and the actual demand value equal zero. If such a forecast should be considered accurate, then SMARE$'$ is the appropriate measure. On the other hand, if a forecaster should not be rewarded for such forecasts (i.e., such a forecast should not be considered a true accurate forecast) then SMARE$''$ should be used. LMAE is another accuracy measure that can be used in the case that relative error of forecasts is important. However, LMAE is inappropriate if $F_{h,i,k-l,k}$ or$A_{h,i,k}$ can take on a value of zero. Instead, OLMAE offsets the actual demand and forecast values to ensure that OLMAE can be used for any demand and forecast values. However, the use of this accuracy measurement requires a good determination of the offset value, c, which may be difficult to judge. Another modification that can be offered to increase usability of the LMAE measure is to truncate this measure any time $F_{h,i,k-l,k}$ or $A_{h,i,k}$ takes on a value of zero. TLMAE assumes the same value as LMAE in the case that both $F_{h,i,k-l,k}$ and $A_{h,i,k}$ exceed zero. Otherwise, the value of TLMAE is set to the log of some large value C.

RMSE is a commonly used measure of forecast accuracy. In special cases when the forecasts are unbiased and the demand is identically distributed, RMSE provides an estimate of the standard deviation of the forecasts. Sometimes, it may be preferred to assign greater weight to some forecast errors. For example, one may wish to assign greater weight (greater penalty or reward) to more recent data. WMAE is an accuracy measure that assigns a weight w_k to each time period k. This weight is assigned according to the relative importance of time period k.

The measures that have been listed thus far all assign the same penalty for overforecasting and underforecasting. Often, such an approach is invalid. If one considers manufacturing planning, overforecasting results in excess capacity reservation for a product and under forecasting means that there may not be sufficient product in stock to meet customer demand. If the cost of these two errors is not equal, it may not be appropriate to assign the same penalty to overforecasting and underforecasting. WSL assigns a different cost for overforecasting and underforecasting.

All of the measures described above were considered for assessing the accuracy of forecasts of IBM Microelectronics Division semiconductor demand. Both statistical desiderata such as robustness to outliers and the preferences of the business managers and forecasters were taken into account. It is difficult to find a single measure that is uniformly acceptable. Measures that involve absolute error have an economic interpretation in terms of dollars lost through forecast inaccuracy, but can be dominated by a few products or time periods when forecasts are very inaccurate; RMSE is even more dominated by a few extreme errors. Measures that cannot be computed when actual or forecast demands are zero are not suitable for the many products for which demand is often zero. Weighted, skew, and offset-logarithmic criteria have some appeal but involve parameters that cannot easily be chosen so that they are valid across the range of products. The overall most suitable measure was judged to be SMARE$''$, which has similarities with MARE when forecast errors are small, puts bounds on the influence of very large forecast errors, and contains adjustments that allow for zero values of forecast and actual demand.

4.3 Comparing performance of two different forecasting models

A standalone accuracy measure may not be of value if there is not some measure against which it can be compared. Theil's U coefficient (also referred to as Theil's U2) is a relative measure that is used to achieve this goal. The Theil U coefficient compares the accuracy of the forecast generated by the proposed forecasting scheme to that of a "naive" forecast. More specifically, Theil's U coefficient compares the RMSE of the forecasting model to the RMSE of the random walk model, and is defined as follows:

$$U = \frac{\sqrt{\sum_{k=1}^{T} (A_{h,i,k} - F_{h,i,k-l,k})^2}}{\sqrt{\sum_{k=1}^{T} (A_{h,i,k} - F^{RW}_{h,i,k-l,k})^2}} \tag{1}$$

where $F^{RW}_{h,i,k-l,k} = A_{h,i,k-l}$ is the forecast generated by the random walk model.

If Theil's U coefficient equals zero, the forecasts are perfectly accurate; the closer the coefficient is to zero, the better the model. In practice, values of 0.55 or less are considered quite good. If $U < 1$, the forecasts generated by the forecasting scheme have smaller error than those generated from the naive model. A value of one indicates that both models have the same measure of error. When $U > 1$, the forecasts generated by the forecasting scheme have greater error than those of the naive model.

We point out that the distribution of the U coefficient is not well understood. Thus, one cannot test the hypothesis that the forecasts generated by the forecasting scheme are significantly different from the forecasts obtained from a naive model.

Dependent upon the nature of the demand data, Theil's U coefficient may be useful. If the actual demand values are zero (at the level of aggregation being considered), the random walk will forecast zero demand and Theil's U coefficient will take on a value of infinity. We thus propose a truncated version of Theil's U coefficient, similar in spirit to TLMAE described above. Truncated Theil's U is defined according to the following formula:

$$\text{Trunc-U} = \frac{\sqrt{\sum_{k=1}^{T}(A_{h,i,k} - F_{h,i,k-l,k})^2}}{\min(\sqrt{\sum_{k=1}^{T}(A_{h,i,k} - F^{RW}_{h,i,k-l,k})^2}, C)}, \tag{2}$$

where C is appropriately defined by the user. We choose to set $C = 1$ for the following reason. If actual orders are equal in every period, then the denominator of (1) is zero. In this case, we wish to capture the exact inaccuracy of the proposed forecasting model. Setting the denominator in (2) equal to 1 (i.e., $C = 1$) achieves this objective.

Similar in spirit to Theil's U coefficient, one can define the relative absolute error ("RAE") of a forecasting model as follows:

$$\text{RAE} = \frac{\sum_{k=1}^{T}|A_{h,i,k} - F_{h,i,k-l,k}|}{\sum_{k=1}^{T}|A_{h,i,k} - F^{RW}_{h,i,k-l,k}|} \tag{3}$$

In cases that the denominator of RAE equals zero, we define a truncated RAE ("TRAE") as follows:

$$\text{TRAE} = \frac{\sum_{k=1}^{T}|A_{h,i,k} - F_{h,i,k-l,k}|}{\min(\sum_{k=1}^{T}|A_{h,i,k} - F^{RW}_{h,i,k-l,k}|, C)} \tag{4}$$

Again, we suggest setting $C = 1$.

5 Development of the forecasting procedure

5.1 General approach

The available data for forecasting are demand series for a range of products. Most of the series are short and highly variable, so demand forecasts at the level of individual products are not very reliable. However, the nodes of the product hierarchy consist of groups of products that might be expected to behave similarly, because the products share some common technology (in a technology hierarchy) or common customers or a common field of application (in a marketing hierarchy).

We therefore decided on an overall approach that makes demand forecast separately at each node of the product hierarchy, and combines the forecasts at different nodes to ensure consistency of forecasts at different levels of the hierarchy (e.g. that the sum of the demand forecasts is equal at each level of the hierarchy).

5.2 Forecasting individual series

A number of forecasting techniques were tested on monthly demand data for the 26-month period from January 1999 to February 2001, for the three product families described in section 1. Because the demand series are erratic and, for many products, very short, it is difficult to fit complex time-series models, and still more difficult to test their forecasting performance on data not used in the fitting process. Ten simple forecasting techniques were therefore considered. They are listed in Table 2. They all need no calibration (i.e., they involve no parameters that need to be estimated from historical data), and require at most six historical data values in order to make forecasts. Thus from the test data we can compare forecast and actual values for 20 months at lead time 1 month, decreasing to 9 months at lead time 12 months.

We compared the performance of the Time Series models using forecasts at all levels of the product hierarchy. Typically, comparisons were made across all nodes at one level of the product hierarchy, for forecasts at a specific lead time. The greatest importance was attached to lead times of 3–5 months, since these lead times are most important for scheduling of manufacturing. The actual and forecast values were compared using the forecast accuracy measures described in Section 4.2, of which SMARE$''$ and TRAE were judged to be the most useful. By these criteria, the WMA and MQ were most often judged to be the best. MQ performed particularly well at the higher levels of one of the product hierarchies, in which demand often has a strong 3-month pattern with the third month in the quarter having unusually high demand. WMA was usually the best or close to the best in other cases. A3 and LM were typically a little worse than WMA, and the "Z" variants A3Z, LMZ, and WMZ, were typically a little worse still. The TX and TXR methods were not competitive: they often gave unreasonably high or low forecasts. The Z method, though never competitive according to the SMARE$''$ and TRAE criteria, sometimes gave the highest accuracy by the MAE criterion: this casts doubt on the value of MAE as a measure for choosing a forecasting technique.

Exponential smoothing was also considered as a forecasting technique. It models variations in trend, level, and seasonality, each of which involves calibration of one parameter, and requires an initialization period of at least 12 months. Thus few data points are left for evaluating the accuracy of out-of-sample forecasts. For many demand series, exponential smoothing appeared to give forecasts that were as accurate as the most accurate of the simple techniques. However, because these results are based on very limited data we feel that the use of exponential smoothing is not justified.

Table 2: Ten simple forecasting techniques.

	Technique	Description of forecasts
Z	Zero	All forecasts zero.
A3	3-month average	Average of last 3 months.
A3Z	3-month average with allowance for zeroes	Zero if at least 4 of last 6 months have zero demand; otherwise, average of last 3 nonzero demands.
LM	Local mean	Average of last 6 months, after discarding highest and lowest values.
LMZ	Local mean with allowance for zeroes	Zero if at least 3 of last 6 months have zero demand; otherwise, average of nonzero demands in last 6 months, after discarding highest and lowest values.
WMA	Weighted moving average	Moving average of last 6 demands, with linearly decreasing weights.
WMZ	Weighted moving average with allowance for zeroes	Zero if months in the last 6 with zero demands account for more than half of the weight in a sequence of linearly decreasing weights; otherwise, moving average of non-zero demands in last 6 months, with weights linearly decreasing over time.
TX	Trend extrapolation	Estimate trend by linear regression of last 6 demands, with time as explanatory variable.
TXR	Robust trend extrapolation	As TX but use robust regression (Matlab function `robustfit`).
MQ	Month-in-quarter	Extrapolate recent pattern of month-within-quarter variation. Details in text of Section 5.2.

The forecasts finally used were therefore WMA and MQ. We now give their detailed description. Let the available monthly demand data be $\{A_{t-h+1}, A_{t-h+2}, \ldots, A_t\}$, i.e. t denotes the current month and h the number of months of history.

Let $H_i = \min(h-1+i, 6)$, $i = 1, \ldots, 12$: this is the length of the moving average that we use for forecasting i months ahead. It is intended to be six but if there is insufficient history to permit this, we use as long a moving average as the available history permits. The forecast for month $t+i$, $i = 1, \ldots, 12$, is given by

$$F_{t+i} = \sum_{j=1}^{H_i} w_j^{(H_i)} \, A^*_{t+i-j}$$

where $w_j^{(H)} = 2(H+1-j)/\{H(H+1)\}$ and

$$A^*_k = \begin{cases} A_k & \text{if } \ k \le t, \\ F_k & \text{if } \ k > t. \end{cases}$$

The weights $w_j^{(H)}$ decrease to zero linearly in j and sum to 1; the forecast is this weighted average of the actual values, except that when actual values are required at future time points they are replaced by forecasts.

The MQ forecast aims to capture the most obvious seasonality in demand: a tendency for the three months of a quarter to have successively higher demand. The idea is to take the pattern of the two most recent quarters and propagate this into the future. This captures the effect, often present in semiconductor demand, of different quarters having different overall demand but similar patterns of relative demand in the months within the quarter. The forecast is computed by multiplying the most recent quarterly total of demand by a month-in-quarter coefficient based on each month's contribution to its quarter, for the two most recent quarters. The current month is regarded as the third month of a rolling quarter. The month-in-quarter coefficient for the jth month in the quarter is defined by

$$q_j = \frac{1}{2}\left(\frac{A_{t+j-6}}{A_{t-5} + A_{t-4} + A_{t-3}} + \frac{A_{t+j-3}}{A_{t-2} + A_{t-1} + A_t} \right), \qquad j = 1, 2, 3,$$

and the forecasts for lead times 1, 2, and 3 are given by

$$G_{t+j} = (A_{t-2} + A_{t-1} + A_t) q_j \,, \qquad j = 1, 2, 3.$$

A somewhat similar scheme, but based on the historical pattern of demand over months in a year rather than months in a quarter, was used in the "WineGlass" business planning tool of Wu et al. (1992). For the next three lead times, the month-in-quarter coefficients are updated to

$$q_j = \frac{1}{2}\left(\frac{A_{t+j-6}}{A_{t-2} + A_{t-1} + A_t} + \frac{G_{t+j-3}}{G_{t+3} + G_{t+2} + G_{t+1}} \right), \qquad j = 4, 5, 6,$$

and the forecasts for lead times 4, 5, and 6 are given by

$$G_{t+j} = (A_{t-2} + A_{t-1} + A_t)q_j\,, \qquad j = 4, 5, 6.$$

Similarly, for lead times 7–12 the month-in-quarter coefficients are given by

$$q_j = \frac{1}{2}\left(\frac{G_{t+j-6}}{G_{t+3} + G_{t+2} + G_{t+1}} + \frac{G_{t+j-3}}{G_{t+6} + G_{t+5} + G_{t+4}}\right), \qquad j = 7, 8, 9,$$

$$q_j = \frac{1}{2}\left(\frac{G_{t+j-6}}{G_{t+6} + G_{t+5} + G_{t+4}} + \frac{G_{t+j-3}}{G_{t+9} + G_{t+8} + G_{t+7}}\right), \qquad j = 10, 11, 12,$$

and the forecasts by

$$G_{t+j} = (A_{t-2} + A_{t-1} + A_t)q_j\,, \qquad j = 7, \ldots, 12.$$

The WMA and MQ forecasts were combined to give an overall forecast of each demand series. We used a linear combination

$$C_{t+j} = \lambda F_{t+j} + (1 - \lambda)G_{t+j}$$

and estimated λ separately for each demand series, using the historical data for the series. The estimation used grid search over λ values between 0 and 1, with λ chosen so as to minimize the SMARE$''$ value of the combined forecast at lead time 3 months. This lead time was chosen because it is the most important for manufacturing decisions. A weighted combination of SMARE$''$ values at different lead times might also be a reasonable criterion. Because the MQ forecast needs 6 months of historical data and the estimation of λ uses a 3-month lead time, 9 months of historical data are required to compute the combined forecast. For demand series with less than 9 months of history, we use just the WMA forecast.

5.3 Forecasting the product hierarchy

It is important that forecasts at different nodes of the product hierarchy be mutually consistent, i.e. that the forecast at a parent node equal the sum of the forecasts at its child nodes. This could of course be achieved by making forecasts at the bottom level of the hierarchy and aggregating them to give forecasts at higher-level nodes. However, the overall accuracy of forecasts is greater at or near the top of the hierarchy, and we have more confidence in forecasts made by univariate analysis of the demand at a top-level node than in the aggregated forecasts of demand at the part-number level. We therefore used a top-down disaggregation procedure. Forecasts are made at each node of the hierarchy; then, working downwards from the top of the hierarchy, at each parent node the forecast demand quantity is distributed among the child nodes, proportionally to the magnitudes of the forecasts at the child nodes.

The product hierarchy is dynamic: old products are discontinued, new products are introduced, and products may be moved from one branch of the hierarchy to another during their lifespan. Even though a node (as a combination of values of classifying factors) is present in successive months, its typical level of demand may change if there is a change in the set of parts that are descendants of this node. This means that a forecast made with one month's hierarchy cannot in general be fairly compared with the actual demand at the "same" node in a later month. This in turn makes it difficult to provide meaningful forecasts, and to estimate the accuracy of the forecasts that are made.

To overcome these difficulties as much as possible, we use a new hierarchy each month. The historical data series at each node of the hierarchy is recomputed each month, and all historical forecasts are recomputed as though the current hierarchy had been in place permanently. The accuracy of these historical forecasts is estimated by comparing the forecasts with the recomputed historical data. We finally compute forecasts of future demand for the current hierarchy, using the calculated accuracy values to provide prediction intervals around the forecasts.

5.4 Manual intervention

The users of the forecasts are managers who have detailed knowledge of the sales or manufacturing process, and in many cases can judge that the time-series forecasts are likely to be inaccurate. The users can override the time-series forecasts and substitute their own judgmental forecasts. The users may also accept the technical forecast by indicating that it should be held fixed during the next reconciliation.

The changes made by the users may introduce inconsistency into the hierarchy of forecasts, so a further reconciliation step is made. It also proceeds from the top down, and allows for forecasts at some child nodes of a parent node to have been fixed by the users. From the forecast at the parent node, the forecasts at every "fixed" child node are subtracted; the remainder is distributed among the remaining child nodes, proportionally to the magnitudes of the forecasts at these nodes. If the remainder is negative, there is an inconsistency that must be addressed by the user, perhaps by increasing the forecast at the parent node. This reconciliation may go through several cycles of fixing of forecasts by the user and top-down distribution of forecast quantities until a satisfactory and consistent hierarchy of forecasts is achieved.

5.5 Overall forecasting procedure

The foregoing computations comprise the entire forecasting procedure, which we now summarize.

1. Build the hierarchy for the current month.

2. Generate forecasts (historical and current) for the current hierarchy. At each node of the hierarchy, and taking each month from the current month back through the previous 24 months as the forecast origin:

(a) Compute the WMA forecast for each lead time.

(b) Compute the MQ forecast for each lead time.

(c) Compute the optimal weight for the combined forecast, a linear combination of the WMA and MQ forecasts.

(d) Compute the combined forecast for each lead time.

3. Reconcile the forecasts.

4. Compute the accuracy of historical forecasts. The bias and MAE of the combined forecast are computed for each lead time and each node of the hierarchy. There may be insufficient data to estimate these quantities reliably: in such cases adjacent nodes at the same level of the hierarchy are merged until enough data points (at least 20) are available.

5. Compute prediction intervals for the current forecasts. The prediction intervals assume a normal distribution for the difference between forecast and actual values for the same lead time in different months, and use MAE to estimate the standard deviation of this normal distribution.[1]

6. Present the forecasts to the users; allow users to change the values of the forecasts.

7. Reconcile the forecasts a second time, to remove inconsistencies introduced by the users' changes. Steps 6 and 7 can be repeated until the users are satisfied with the entire hierarchy of forecasts.

6 Empirical testing

6.1 General approach

Once the overall forecasting models and method is determined, it is important to objectively measure the improvement offered by using the FIT forecasting methodology. We performed three phases of model testing and validation. The first phase ("Phase I testing") consisted of extensive testing of the accuracy of forecasts generated for all three product families at all levels of the product hierarchies. The overall sample size for model testing was large, due to the large number of products and customers considered. We computed forecast accuracy of forecasts for ten months of actual demand. Heavy emphasis was placed on the forecasts for lead times of one to six months, as these are the most important for the manufacturing planning and decision-making.

[1]The assumption of normality has not been justified by statistical analysis. It is made for convenience and for consistency with inventory management tools for which the demand forecasts may be used as input.

Forecasts with longer lead times are used for longer-term financial planning. The second phase of testing ("Phase II testing") consisted of additional validation of the results obtained during Phase I testing, with consideration of lead times of one to six months only. Phase II testing computed forecast accuracy of forecasts at all levels of the three product hierarchies for (a more recent set of) five months of actual demand. The third phase of testing ("Phase III testing") consisted of custom analysis of the Phase I and Phase II results, using yet another more recent set of data, to satisfy specific user requirements and specifications.

6.2 Phase I testing

For Phase I testing we collected historical sales data and IBM Microelectronics Division forecast data, with forecast lead time one to twelve months, for the months of July 2001 through April 2002. We used the FIT forecasting methodology to generate monthly forecasts (with one to twelve month forecast lead times) at all levels of the hierarchy for each of the three product families described in Section 1, as follows (L denotes the lead time, in months):

- Assume that we have data through June 30, 2001. On July 1, 2001 compute FIT forecasts for July 2001 ($L = 1$), August 2001 ($L = 2$), ..., May 2002 ($L = 12$).

- Assume that we have data through July 31, 2001. On August 1, 2001 compute FIT forecasts for August 2001 ($L = 1$), September 2001 ($L = 2$), ..., June 2002 ($L = 12$).

- Continue.

Using the actual data collected through April 2002, we were able to test accuracy of FIT forecasts generated for the months through April 2002. Table 3 describes the FIT forecasts whose accuracy was measured during this validation stage. The leftmost column in Table 3 indicates the date at which the forecasts were generated. The subsequent columns indicate the months for which demand forecasts were generated.

Thus, Phase I testing consisted of ten sets of forecasts with lead time one, nine sets of forecasts with lead time two, eight sets of forecasts with lead time three, etc. Our next objective was to compute forecast accuracy for every node in the hierarchy, for each forecast lead time. We computed forecast accuracy as follows:

(i) Compute FIT absolute error (FIT forecast for month j minus realized demand in month j) for each month.

(ii) Compute IBM Microelectronics Division absolute error (IBM Microelectronics Division forecast for month j minus realized demand in month j) for each month.

Table 3: FIT forecasts testing during first validation stage.

	7/2001	8/2001	9/2001	10/2001	11/2001	12/2001	1/2002	2/2002	3/2002	4/2002
7/1/2001	$L = 1$	$L = 2$	$L = 3$	$L = 4$	$L = 5$	$L = 6$	$L = 7$	$L = 8$	$L = 9$	$L = 10$
8/1/2001		$L = 1$	$L = 2$	$L = 3$	$L = 4$	$L = 5$	$L = 6$	$L = 7$	$L = 8$	$L = 9$
9/1/2001			$L = 1$	$L = 2$	$L = 3$	$L = 4$	$L = 5$	$L = 6$	$L = 7$	$L = 8$
10/1/2001				$L = 1$	$L = 2$	$L = 3$	$L = 4$	$L = 5$	$L = 6$	$L = 7$
11/1/2001					$L = 1$	$L = 2$	$L = 3$	$L = 4$	$L = 5$	$L = 6$
12/1/2001						$L = 1$	$L = 2$	$L = 3$	$L = 4$	$L = 5$
1/1/2002							$L = 1$	$L = 2$	$L = 3$	$L = 4$
2/1/2002								$L = 1$	$L = 2$	$L = 3$
3/1/2002									$L = 1$	$L = 2$
4/1/2002										$L = 1$

(iii) Compute relative absolute error ("RAE") as (FIT absolute error)/(IBM Microelectronics Division absolute error). RAE measures the accuracy of FIT forecasts relative to IBM Microelectronics Division forecasts. A relative accuracy measure helps measure the benefit (in terms of improved accuracy) of using the FIT forecasting methodology instead of the forecasting process typically followed by IBM Microelectronics Division. The FIT forecasting methodology generates a forecast value for each period for each node in the product hierarchy. In the case that no IBM Microelectronics Division forecast was generated for a given month for a given node or that the IBM Microelectronics Division forecast equals zero, no RAE value is computed for that month-node combination.

(iv) For each forecast lead time, compute average RAE for each node by computing the simple average of the individual RAE values for the given node and forecast lead time.

(v) Winsorize (truncate) RAE to reduce the impact of outliers.

(vi) To compute average accuracy over a set of nodes, we compute the geometric mean of the relevant RAE values ("GMRAE"). RAE is a unitless measure, which is desirable in cases where demand volume can vary significantly from one part number to another. GMRAE is computed as the nth root of the product of n RAE values. Taking the geometric mean gives equal treatment to high and low RAE values: for example, RAE values of 2 and 0.5 make equal contributions to GMRAE. This ensures that inferences based on GMRAE do not depend on which forecast is in the numerator and which in the denominator in the definition of RAE in step (iii) above.

Thus, for each forecast lead time value we compute the (arithmetic) average RAE for each node at the lowest level of the product hierarchy (part number-customer) and then compute the GMRAE to determine average accuracy at higher levels of the product hierarchy.

Table 4: GMRAE values for Phase I testing.

		$L=1$	$L=2$	$L=3$	$L=4$	$L=5$	$L=6$
Product Family I							
	Dominant Customer	0.44	0.68	0.78	0.64	0.59	0.44
	All Other Customers	1.47	1.01	0.78	0.62	0.55	0.55
Product Family I	Total	1.41	0.99	0.78	0.62	0.56	0.55
Product Family II							
	Dominant Customer	1.11	0.97	1.37	0.42	0.21	0.43
	All Other Customers	1.37	0.74	0.63	0.55	0.50	0.55
Product Family II	Total	1.35	0.75	0.67	0.54	0.48	0.54
Product Family III							
	Dominant Customer	0.79	0.83	0.38	0.66	0.76	0.72
	All Other Customers	1.89	0.84	0.70	0.64	0.78	0.76
Product Family III	Total	1.69	0.84	0.64	0.65	0.78	0.76
Grand Total	Total	1.42	0.90	0.73	0.60	0.56	0.57

Since GMRAE is a relative measure, it measures the goodness of one forecasting method as compared with another. GMRAE values greater than 1 indicate that the IBM Microelectronics Division forecasts are more accurate than the FIT forecasts; values less than 1 indicate that FIT forecasts are more accurate than IBM Microelectronics Division forecasts.

Table 4 lists the GMRAE values calculated in Phase I testing. As discussed in Section 3, we created separate product hierarchies for the dominant customers and for all other customers within each product family. In Table 4 we report the GMRAE values for the highest level in each of these product hierarchies for each of the three product families. We also report the overall GMRAE values for each product family. Table 4 lists the GMRAE values for forecast lead times one through six months.

The results of Phase I testing vary with the forecast lead time. At lead time 1 month, IBM Microelectronics Division forecasts are, in general, more accurate. This is not surprising, because the IBM Microelectronics Division forecasts for lead time 1 month are typically set equal to the firm orders received from customers: they are a strong indicator of customers' purchasing intentions and one would expect them to have high accuracy. At lead time 2 months, the FIT and IBM Microelectronics Division forecasts are approximately equally accurate. At lead time 3 months, FIT forecasts tend to be more accurate than IBM Microelectronics Division forecasts, but not consistently so. At lead times greater than 3 months, FIT forecasts are, in general, more accurate than IBM Microelectronics Division forecasts; FIT forecast errors are smaller by factors typically in the range 20% to 55%.

6.3 Phase II testing

For Phase II testing we collected monthly actual demand data and IBM Microelectronics Division forecasts for May 2002 through September 2002. The purpose of Phase II testing was to discover whether the results obtained in Phase I continue to hold in this more recent period.

We collected historical sales data and IBM Microelectronics Division forecasts, with lead time 1 month through 12 months, for the months of May 2002 through September 2002. We used the FIT forecasting methodology to generate monthly forecasts, also with lead time 1 month through 12 months, at all levels of the hierarchy for each of the three product families, as follows:

- Assume that we have data through April 30, 2002. On May 1, 2002 compute FIT forecasts for May 2002 ($L = 1$), June 2002 ($L = 2$), ..., April 2003 ($L = 12$).
- Assume that we have data through May 31, 2002. On June 1, 2002 compute FIT forecasts for June 2002 ($L = 1$), July 2002 ($L = 2$), ..., May 2003 ($L = 12$).
- Continue.

We measured the accuracy of FIT forecasts generated for the months through September 2002, in the same manner as in Phase I. The general character of the accuracy results was the same as in Phase I.

6.4 Phase III testing

Phase III testing consisted of customized testing according to requests of individual users. We met with the head of IBM Microelectronics Division's Worldwide Tactical Marketing Team, responsible for all final IBM Microelectronics Division forecasts, as well as with the lead forecasters for each of the three product families included in our study.

For Phase III testing we used all of the historical data from July 2001 through September 2002 (i.e., the data used for Phase I and Phase II testing). We performed custom analysis of the pilot results, according to each individual user's requirements and preferences. We analyzed all of the data or focused on specific part numbers or customers or on specific time periods between July 2001 and September 2002, according to the specific requests of each of the individual users. The principal analyses were as follows:

(i) We analyzed accuracy results at different levels of aggregation such as at the part number-customer level, the customer level, and the technology level. The goal of the analysis was to identify customers for which FIT forecasts performed well relative to IBM Microelectronics Division forecasts. If FIT forecasts are highly accurate (relative to IBM Microelectronics Division forecasts) for smaller customers, for example, the forecasters can consider accepting the recommended FIT forecast for smaller customers thereby freeing time to generate highly accurate Microelectronics Division forecasts for larger customers.

We found FIT forecasts were, in general, more accurate than IBM Microelectronics Division forecasts for the class of "small" customers. (These are customers with small order sizes.) This can allow for a precise forecast-adoption rule: for all customers defined in the class of small customers, automatically adopt the FIT forecast.

(ii) We studied the impact of reservations on forecast accuracy. Often, IBM Microelectronics Division specifies an amount of product that would be produced and held in inventory as a buffer (typically, in case customer demand exceeded the forecast demand). The quantity produced in excess of forecast demand is referred to as "reservation." Thus, it could be the case that while forecast demand is inaccurate, the inclusion of these reservations will improve the accuracy of IBM Microelectronics Division forecasts. We performed this analysis by increasing the IBM Microelectronics Division forecast value by any reservation amount associated with that forecast. We then computed the forecast accuracy of this "adjusted" IBM Microelectronics Division forecast value. We found that including reservation quantities did not, in general, improve IBM Microelectronics Division forecast accuracy; forecast accuracy improved for a single forecast lead time for two customers only.

(iii) We tested whether the forecasts tended to be too low or too high. GMRAE, like most forecast accuracy measures, applies equal penalties to overforecasting and underforecasting. However, one can argue that these should be considered separately, as they have different impacts on business processes and on a business's success. We computed separate accuracy measurements for periods with overforecasting and periods with underforecasting. We counted the number of times that FIT forecasts were greater or less than actual demand. We also counted the number of times that IBM Microelectronics Division forecasts were greater or less than actual demand. Both IBM Microelectronics Division and FIT forecasts were too high more often than they were too low. However, the number of times that FIT overforecasts was less than the number of times that IBM Microelectronics Division overforecast. Also, FIT underforecasts more often that IBM Microelectronics Division underforecasts. We then studied the magnitude of overforecasting and underforecasting error. When we compared the relative percentage error in forecasts volumes (percentage error between actual and forecasted values), we found that FIT percentage error was on average smaller than the IBM Microelectronics Division percentage error, both in cases where forecasts were too high and in cases that forecasts were too low. The implications of these two types of forecast errors for inventory management and customer satisfaction are likely to affect the decision of which forecast to adopt.

Overforecasting can lead to higher inventory costs but underforecasting can lead to customer dissatisfaction. For each product, IBM Microelectronics Division must decide which error is more costly and accordingly decide which forecast to adopt or use as a basis for manufacturing and planning decisions.

Note that there is nothing inherent in the definition of FIT that would lead to consistent overforecasting or underforecasting. The observed tendency of FIT to overforecast may reflect business conditions during the period for which forecasts were computed; there is no reason to expect it to continue to hold when business conditions change. The tendency of IBM Microelectronics Division to overforecast may, however, be systemic, and reflect the optimism of the marketing personnel who make the forecasts.

7 Conclusions

The forecasting process at IBM Microelectronics Division is complicated and time consuming. At the beginning of each month, monthly forecasts for each part number-customer combination are generated for the upcoming 12–18 month horizon. Forecasts are reviewed and updated weekly within each month. Consequently, thousands of forecast figures must be determined each month. We have developed FIT, an automatic procedure that computes these forecasts, and we have made extensive comparisons of the FIT and IBM Microelectronics Division forecasts.

The forecast comparisons yield two main results. First, IBM Microelectronics Division forecasters can accurately predict short-term demand but at longer lead times FIT provides more accurate demand forecasts. This phenomenon can be explained in part by the fact that IBM Microelectronics Division forecasts for demand one or two months ahead are often set to equal the actual orders placed by the customers. Second, IBM Microelectronics Division forecasters can predict demand by large customers more accurately than demand by small customers. This greater accuracy can be attributed to the fact that forecasters will often invest significant time working with large customers to determine what their future product needs will be.

Our results indicate that IBM Microelectronics Division would gain significant benefits from including FIT in its forecasting procedure. FIT should be used where it is more accurate than IBM Microelectronics Division's existing forecasts: specifically, to forecast demand with longer lead times and to forecast demand for smaller customers. The FIT forecasts could either be adopted without revision or could be used by IBM Microelectronics Division's forecasters as a point of reference for generating the final subjective forecast of demand. By using FIT to generate forecasts for customers or lead times that FIT has shown to forecast more accurately, forecasters can allocate time to other forecasting responsibilities, such as improving their demand forecasts for larger customers. The result will be improved forecast accuracy and time savings.

Acknowledgements

Many people helped us with data preparation, programming assistance, and discussion of IBM Microelectronics Division's forecasting requirements and statistical modeling. In particular we are grateful to Jonathan Wong, Sandy Anderson, and Sarah Hood (IBM Microelectronics Division), and Charlie Hu, Samir Sharma, and Mike Levanoni (IBM Research Division).

References

Wu, L. S.-Y., Hosking, J. R. M., and Doll, J. M. (1992). Business planning under uncertainty: will we attain our goal? *International Journal of Forecasting*, **8**, 545–557.

Armstrong, J.S. and F. Collopy (1992). Error measures for generalizing about forecasting methods: empirical comparisons. *International Journal of Forecasting*, **8**, 69–80.