Scenario Analysis of Demands in a Technology Market Using Leading Indicators

Mary J. Meixell
Bell Labs, Lucent Technologies

S. David Wu
Lehigh University

Report No. 98T-013

Scenario Analysis of Demands in a Technology Market Using Leading Indicators

Mary J. Meixell
Bell Labs, Lucent Technologies
Princeton, New Jersey

S. David Wu
Department of Industrial and Manufacturing Systems Engineering
Lehigh University
Bethlehem, Pennsylvania

Abstract

This paper proposes an approach to analyze demand scenarios in technology-driven markets where product demands are volatile, but follow *a few* identifiable life-cycle patterns. After analyzing a large amount of semiconductor data, we found that not only can products be clustered by life-cycle patterns, but in each cluster there exists leading indicator products that provide advanced indication of changes in demand trends. Motivated by this finding we propose a scenario analysis structure in the context of stochastic programming. Using the Bass growth model and a Baysian update structure, the proposed method streamlines scenario analysis by focusing on parametric changes of the demand growth model overtime. The Baysian structure allows expert judgement to be incorporated in scenario generation while the Bass growth model allows an efficient representation of time varying demands. Further, by adjusting a likelihood threshold, the method could generate scenario trees of different sizes and accuracy. The structure provides a practical scenario-analysis method for manufacturing demand in a technology market. We demonstrate the applicability of this method using real semiconductor data.

Keywords: Semiconductor Manufacturing; Scenario Analysis; Growth Models; Leading Indicators; Multistage Stochastic Programming; Bass Diffusion Model; Baysian Structure; Manufacturing Demand Analysis; Planning and Decision Making; Manufacturing Logistics

1. Introduction

In technology driven markets, demands are characterized by different drivers of technological innovation and a relatively short product life-cycle. Typical examples for such market environment are the high-tech industries such as semiconductor and electronics. The research is motivated by a yearlong study at Lucent Technologies. After analyzing their demand data for some 3,500 products, we found that these products follow a few (approximately six) life-cycle patterns and the products can be grouped according to these patterns using statistical cluster analysis. More importantly, after correlation analysis on historic shipment data we found that in each cluster there exists a subset of "leading indicator" products that give advanced indication of changes in demand trends. Based on these observations, we believe it is of practical importance to develop a demand model making use of the following information: the observed data points provided by the "leading indicator" product, and a pair of time-lagged growth models that relate the demand pattern of the leading indicator and the actual product demand. This demand model can be then embedded in a stochastic model for manufacturing decision-making. In this paper, we examine such demand model in the decision context of multistage stochastic programming.

Multistage stochastic programming models often suffer from an exponentially growing scenario tree, which causes major computational difficulties (Birge and Louveaux, 1997)(Kall and Wallace, 1994). Two basic approaches have been proposed for this problem: (1) *algorithmic approximation and sampling* methods that provide bounds or confidence intervals that direct the solution algorithm during optimization, and (2) *model simplification* schemes that control scenario generation based on problem-specific *a priori* analysis. Our proposed method belongs to the latter category. While the *algorithmic* approach involves structural and computational analysis *after* a model is formed, the *model simplification* approach seeks problem specific knowledge to reduce the model size *before* optimization is ever attempted. Clearly, the two approaches can be combined and each applied at a different stage when solving a real-world problem. We will concentrate our effort on developing a streamlined approach to model demand scenarios in a technological driven market.

A main thesis of this paper is that given the information provided by observing the leading indicator products and a growth model characterizing the technology life-cycle, we will

be able to analyze demand scenarios for a particular market segment. In essence, our approach uses the leading indicators as an instrument which estimates the parameters in the demand growth model. We assume the existence of a nominal demand growth curve, which represents the results of a demand forecast. We then define demand scenarios by various deviations from the nominal curve. This is done such that the dimensionality of the resulting scenario tree is significantly reduced when compared to a typical scenario tree, while most critical elements of demand variation is captured nonetheless.

2. Generating Demand Scenarios

The central concept of our proposed approach is the use of a leading indicator to model the process that drives demand over time. We assume the existence of a nominal demand curve for the leading indicator based on earlier forecasts; the curve is defined by specific parameter settings in a growth model. This nominal curve represents the analyst's best assessment of the product life cycle using historic information and other means of *a priori* analysis. We then devise a scenario tree structure based on possible parametric deviations from the nominal curve. We start our analysis on a two-stage scenario tree. In stage one, a set of possible deviations from the nominal curves is computed. In stage two, the Bayes rule is used to compute the posterior probability for each parametric change, condition on the stage-one demand realizations, and a likelihood function characterizing past behavior of parameter changes. Importantly, at any point in the defined scenario tree, the demand series associated with the cluster can be directly computed from the leading indicator demand model. The term "demand" is used throughout this paper to represent anticipated sales for manufactured products. We assume that influence of strictly binding capacity and competitive forces are outside of the scope of the demand model. For the cases where it does not, a growth model that includes the appropriate competitive and capacity factors will be needed and can be incorporated into a similar framework.

2.1 Use of a Growth Model

Growth models have been shown useful for describing demand for short-life-cycle, high-technology products (Kurawarwala and Matsuo, 1996). Traditional data-dependent time series models are not appropriate in this environment since the data available is often insufficient for

parameter estimations at an acceptable level of confidence. It has been shown that growth models can be used to describe the behavior of stochastic demand, without the sizeable data requirements typically associated with time-series models (Meade and Islam, 1998). Let $\tilde{R}_t$ be the demand for the t^{th} value in the series (t=1,2,...). (The item index is dropped for convenience.) At time t, R_1, R_2, ..., R_t have been observed and are known while R_{t+1}, R_{t+2}, ... are unknown. Let $\tilde{\Theta}$ be a set of parameters used to describe demand, $\tilde{R}_t$, whose values are also unknown. We use the Bass diffusion model (Bass, 1969) to define demand over a series of time periods, t, where $\tilde{R}_t$ is defined as follows:

$$\tilde{R}_t = m\left[\frac{1-e^{-(p+q)}}{1+\left(\frac{q}{p}\right)e^{-(p+q)}}\right] \qquad (1)$$

This is a deterministic statement of the diffusion model, appropriate for a specific realization of demand parameter values. The parameter m represents total life-cycle sales, and, the parameters p and q are shape parameters. m is simply a scale parameter. The p parameter is known as the *coefficient of innovation* and is intended to capture external growth effects – the growth in the number of innovators or adopters that buy the new item. The q parameter is the *coefficient of imitation* and is intended to capture internal growth effects – the growth in the number of imitators who are influenced by the innovators to buy the item. The Bass diffusion model is applicable in an environment such as the one we model here, where competitive forces and capacity restrictions are treated independently of the model.

To compute the demand over an interval of time (t-1,t), we use a discrete version of the Bass diffusion model as follows:

$$S(u) = m \cdot p + (q-p)\cdot S(t_{u-1}) - \left(\frac{q}{p}\right)S^2(t_{u-1}) \qquad for\ u = r...T \qquad (2)$$

where: $S(u)$ is the demand in the time interval (t_{u-1}, t_u)

r is the start time for the demand series

T is the final period of demand

A pair of models of this type is used in our framework to enhance the predictive capability of the scenario structure. A primary model describes demand for a cluster of items of interest, and has a set of parameters that we will designate as $(m,p,q)^{CL}$. The second model describes the demand for the cluster's leading indicator with parameters $(s,m,p,q)^{LI}$.

A location parameter, s, is added to the parameter set for the leading indicator to represent the lag of the leading indicator series relative to the cluster series. When the lag between the leading indicator and the cluster is greater than 0, information about the cluster demand can be derived from the observed behavior of the leading indicator demand. Let t_o^{LI} be the timing of the initial demand for the leading indicator, and t_o^{CL} be the timing of the initial demand for the cluster of interest. Also, let the time of the peak sales be t_m^{LI} for the leading indicator and t_m^{CL} for the cluster. The time periods of the final demands are t_f^{LI} and t_f^{CL} for the leading indicator and cluster, respectively. The lag value is the difference between the start times for the two series, $t_o^{CL} - t_o^{LI}$. We treat lag in this analysis as a stationary but unknown parameter. Figure 1 illustrates these relationships.

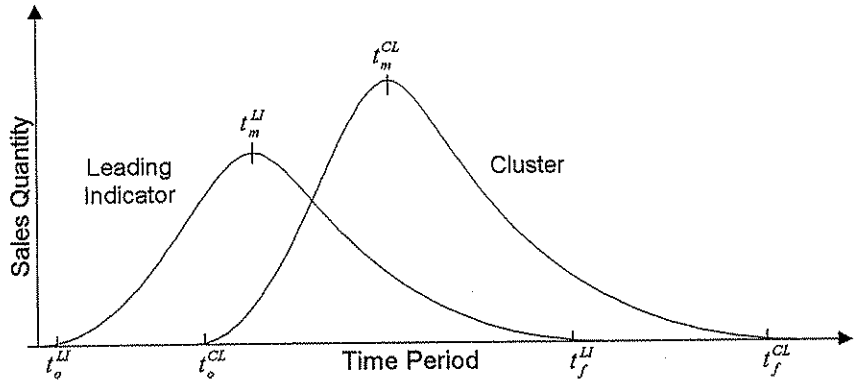


Figure 1. Time Based Relationship between the Demand Models

It is to be expected that actual demand will shift from the nominal curve, as illustrated in Figure 2 (a)-(c). For example, a demand series for a product may exhibit a "right-shift" in time when demand for a product is delayed. Demand may change in total life cycle demand, and exhibit an "upshift" or "downshift." Or, demand may shift in its growth patterns, and follow a "left-shift" or "right-shift" as illustrated. Furthermore, practitioners can often anticipate the form and assign probabilities to these shifts as alternative scenarios to demand.

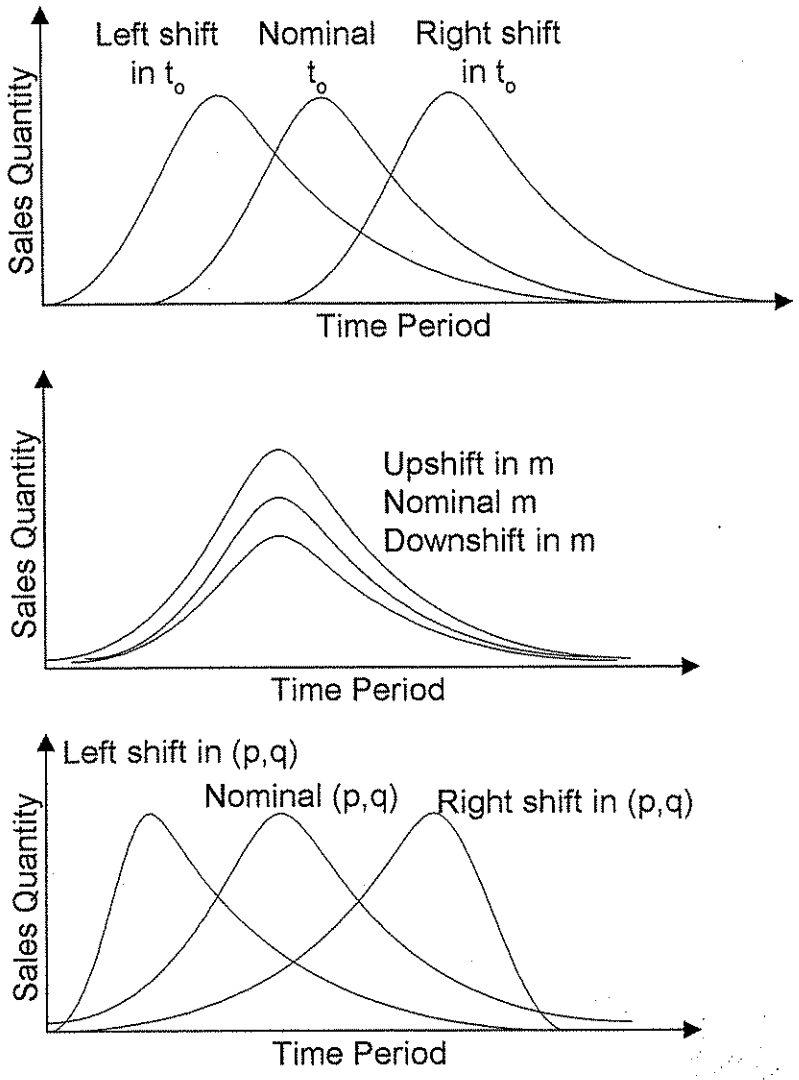


Figure 2. Time, volume, and shape shifts in demand

We incorporate the tendency for these shifts in demand to occur by treating the parameters of the demand model as random variables. The states of these random variables form the branch points on a scenario tree. We will denote the time shift in s as one of the possible states $\{s^-, s^0, s^+\}$. The volume shift, m, also has three possible states $\{m^-, m^0, m^+\}$. And the shape shift can be described as one of the possible states $\{pq^-, pq^0, pq^+\}$. The two shape parameters are grouped in this approach so that it is the pair, (pq) represented on the scenario tree. In working with this model, we found that appropriate skew changes can be modeled by allowing one of these parametrs, p, to vary. Consequently, both parameters are estimated in this approach, but the states in the scenario tree are specified based on the pair.

The parameter s is an integer-valued lag time between the leading indicator and cluster demands. The state s^0 is the nominal, expected lag time, s^- represents lag times shorter than the nominal, and state s^+ represents longer lag times. It is important that the states for each of the parameters are collectively exhaustive. So for the location parameter, s, we use the following state definitions:

s^-	*when* $s < s^0$
s^0	nominal, expected value of the lead time
s^+	*when* $s > s^0$

The states m^0, m^- and m^+, are the nominal, downshift and upshift in market size. We define these states by discretizing the continuous distribution for the m parameter, and then using a credibility interval to establish a range around the estimated nominal value, m^0. A credibility interval is a confidence set for a parameter and is analogous to the concept of confidence intervals in classical statistical inference. The interval is centered at m' and when m is normally distributed, 90% of the density is within $1.645 * \hat{\sigma}_m$ of the mean. The 90% credible set for m is $(m' - 1.645 \cdot \hat{\sigma}_m, m' + 1.645 \cdot \hat{\sigma}_m)$, where m' is an estimate of the mean. $\hat{\sigma}_m$ can be approximated using historic data on the distribution of demand for similar products. Figure 3 illustrates this concept.

Setting limits on the parameter m in this way is similar to the use of control limits in Statistical Process Control. Lin & Adams (1996), Roes, Does, & Schurink (1993), and Nelson

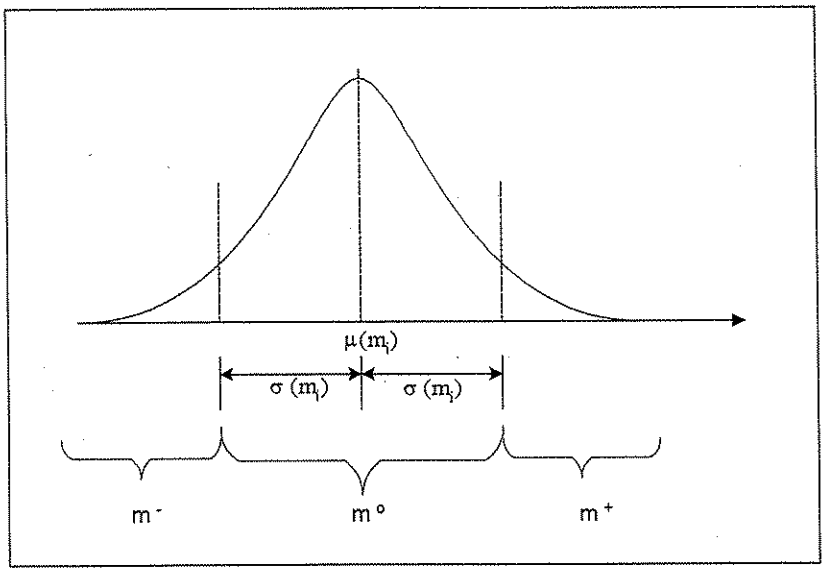


Figure 3. Credibility Limits around *m*

(1982) discuss the use of this idea in practice. Here, we apply the concept to determine when the observed value of m varies enough to claim that there has been a shift in the unknown m parameter.

Note that interpreting the continuous distribution for m in this way ensures that the scenarios are collectively exhaustive, as well as enabling a discrete representation of m on the scenario tree. So for the parameter m, we have the following state definitions:

m^-	$m < m^0 - 1.645\,\hat{\sigma}_m$
m^0	$m^0 - 1.645\,\hat{\sigma}_m \le m \le m^0 + 1.645\,\hat{\sigma}_m$
m^+	$m > m^0 + 1.645\,\hat{\sigma}_m$

Similarly, we define three possible states for the shape parameters – left-skewed, right-skewed, and central - and the nominal state can be any one of them. So, we define $(pq)^0$ as a curve with central tendency, $(pq)^-$ as a left-skew, and $(pq)^+$ as a right-skew.

pq^-	*left skew*
pq^0	central tendency
pq^+	*right skew*

The nominal parameter values for p and q can be estimated given a minimal amount of demand data at time t_0^{LI}. An efficient implementation of this step is to build over time a "catalog" of possible functional forms for demand curves as product growth curves are observed for individual products and their best-fit parameter values are estimated statistically. Product demand traits can be associated with the individual curves, and the curves can be classified using standard classification techniques. In this way, a set of relatively few and historically relevant growth curves can be collected for use in demand analysis for future products. Thus, identifying a specific state of the shape parameters p and q in the nominal curve is to simply select the parameter values of (pq) from a "catalog" of possible functional forms for demand curves.

Estimating the Parameters of the Growth Model

At various points of the scenario-tree generation procedure, the parameters of the growth model need to be estimated based on observed demand data. Specifically, the parameters must be estimated for the leading indicator model initially to provide the values for the nominal demand model. After demand through period t_m^{LI} is observed, the parameters must be estimated again to determine which state best describes the demand observed up-to-date. Moreover, the parameters of the cluster demand model are estimated to determine the parameter values for use in computing the demand series built into each scenario.

When enough data exists, the nominal lag time s^0 can be computed by analyzing the correlation between the leading indicator and the cluster. Oftentimes, a strong correlation exists at several different lag points for a specific leading indicator. In order to provide sufficient advanced warning via the leading indicator, it is best to choose from the correlation analysis the largest acceptable lag time. On the other hand, it is desirable that the chosen lag time provides a robust estimation of the demand curve over the entire life-cycle.

Throughout our analysis we assume that only limited demand data is available, (a minimum of 6 to 8 data points), and we use a standard estimation technique for the Bass diffusion model as the nominal prior state for the leading indicator. Standard approaches for parameter estimation include ordinary least squares estimation as in Bass (1969), maximum likelihood estimation as in Schmittlein and Mahajan (1982), and nonlinear least squares

estimation as in Srinivasan and Mason (1986). When only partial data is available, or when the estimation procedure yields unacceptable parameter values, an algebraic estimation technique as described in Mahajan & Sharma (1986) can be used.

When historical or forecast data does not exist, the nominal states for the leading indicator parameters may be based on independent market research and judgment alone. Methods for estimating m^0 are described in texts such as Kress and Snyder (1994) and Boyd, Westfall, and Stasch (1981). The parameters p^0 and q^0 can be selected based on an analogue concept that allows for choosing these parameters based on an qualitative evaluation of market and product features. Kalish and Lilien and (1986) present examples of the use of analogue products to select growth curve parameters in cases where no demand history has yet occurred.

2.2 The Scenario Tree and the Scenario Generation Procedure

The central concept in demand scenario generation is to use a leading indicator growth curve to model the demand in a broader market segment (cluster demand), and to analyze possible deviations of actual demand from a prespecified nominal model. We construct a two-stage scenario tree assuming that a nominal product life-cycle model is in place with state defined by the triplet $(s,m,pq)^{LI}$. Considering k possible deviations for each parameter, Stage 1 of the scenario tree represents 3^k possible deviations from the nominal state during period one. At the end of period one, an observation is made on the leading indicator demand up-to-date. Given this demand realization, Stage 2 of the tree considers 3^k possible parametric deviations from each Stage 1 state for the remainder of the product life cycle. While the Stage 1 probabilities are given by prior probability distributions associated with the parameter set, the Stage 2 probabilities are given by the posterior probabilities for each specified parameter state based on a Baysian update. Specifically, the Stage 2 probabilities are computed using the prior probabilities from Stage 1, an observation of leading indicator demand realization up to the end of period 1, and a likelihood function which describes the conditional dependence between the observation and the parametric state. Each leave node of the scenario tree represents a unique scenario, which characterizes two parametric changes of the growth model during the planning period. Assuming a stationary relationship between the leading indicator and the cluster demands, then at any point in the scenario tree the demand series associated with the entire cluster can be derived from the parameterized leading indicator demand model. Notice that the number of stages considered

depends on the number of demand observations. The scenario tree can be generalized to multiple stage by allowing multiple demand observations. The tree, however, grows exponentially as the number of stages increases.

We now summarize the main steps involved in constructing a two-stage scenario tree using the leading indicator growth model, where k deviations are considered for each parameter.

1. Based on a demand time-series associate with each product, group items within the problem scope into clusters such that the cluster members are more similar within the cluster than between the clusters.

2. Identify a leading indicator product for each cluster of interest.

3. Initialize the scenario tree structure. Compute the stage-one probabilities for each scenario as the prior probability associated with each the parameter set for the leading indicator demand.

4. Compute the stage-two probabilities for each scenario as the posterior probability result from a Bayesian update, using the prior probabilities and the likelihood function.

5. Compute the scenario probabilities.

6. Compute the cluster demand series given the 2-stage leading indicator demand model.

In the following, each steps of the above procedure are described in detail.

Clustering the Items

Clustering is the grouping of individual items into clusters such that the items within the clusters are more similar than across the clusters. In our application, the items within a cluster must be sufficiently similar such that their demands can be characterized by a certain leading indicator. There are three steps in establishing clusters: selecting a variable or set of variables that can be used to quantify similarity, selecting a method for clustering, and assessing the results. Gnanadesikan (1997) provides a discussion on clustering analysis. We assume that a partial demand time-series is available for each product and we characterize each time series by three types of similarity measures: First, consider a characterization of the demand time series. Specific measures in this category may include CV of sales, skewness, and the slope of the

change in the sales quantities over a certain period. Another type of similarity measure relates item demands to a specific leading indicator. Items would be grouped in this case according to their correspondence to a pre-selected demand model. For example, items that can be successfully predicted by the same growth pattern would be grouped together. The third type of similarity measure groups items according to their demands on resources. Items grouped by facility, equipment or by technology tend to create clusters that use the same resources during manufacturing. This is particularly relevant in modeling production planning problems, where there is typically a set of capacity related constraints that link usage of that facility (or equipment) to the available capacity. Modeling issues influence clustering because it is desirable to avoid cross-facility clusters.

Clustering represents an effort in the procedure to reduce the overall dimension of demand analysis. By clustering, we elevate the point of forecast to a higher level of aggregation, and so fewer individual models need to be calibrated. In some technology markets, the content of the clusters can be identified manually from insights of the technology dependency.

Identify leading indicators

Leading indicators are measures that are readily observable and that are related in some way to another variable of interest. In this application, the leading indicator is a product whose demand pattern predicts the overall cluster demand. Importantly, a leading indicator has a non-zero lag time relative to the cluster demand. We assume the lag time to be unknown but stationary over the product life cycle. Ideally, some historic demand data (e.g., shipment record) exists for the cluster and the leading indicator products, allowing a correlation analysis to identify leading indicators. Pairs of time series, for the leading indicator and the cluster demand, can be analyzed to measure how closely the series have tracked over history, and give a basis for selection of the leading indicator. In many industrial applications, the existence of leading indicator products is well known to the practitioners and they are often driven by technological dependency among products. For instance, in our analysis of Lucent's shipment data, it is clear that the sales of a certain modem chip provides excellent indication for an entire product family which make use of the .35 technology. The time series between the indicator and the whole cluster show a correlation of 0.98 at the lag of three to five weeks. Techniques that could be used to identify leading indicators vary depending on the amount of historic data available.

However, once a credible leading indicator is identified, it is possible to reduce significantly the efforts in modeling demand scenarios. The rest of the procedure describes how this is accomplished.

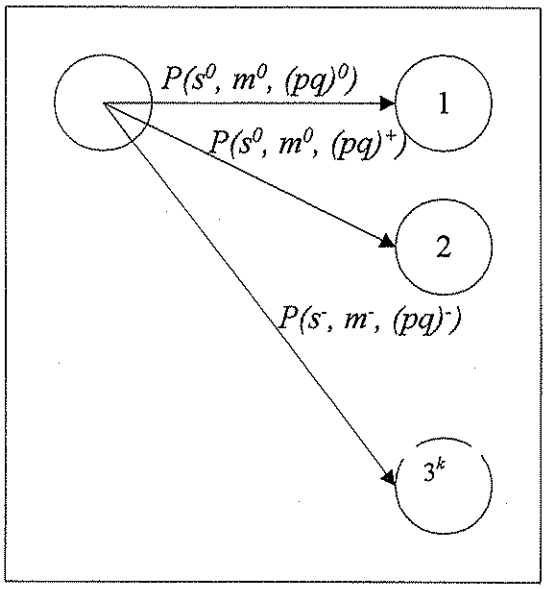


Figure 4. Stage 1 states and probabilities

Computing the Stage 1 Probabilities

The first stage probabilities associated with each scenario are the prior probability distributions for the parameter set of the leading indicator demand model. Since the scenarios are constructed on the leading indicator, we only need the state and the probability information for s, m, and (pq) for the leading indicator demand model. Figure 4 illustrates the states associated with the first stage scenarios. The states are defined as all possible combinations of deviations for each of the three parameters with prior probabilities, P(s, m, (pq)). In practice, the prior probabilities can be established based on statistical data analysis, market research and expert knowledge. The Bayesian framework allows both quantitative and qualitative information to be incorporated here. It is important that the prior distribution reflects all of the relevant information available to the decision-maker prior to observing a particular demand sample.

Historical data for previous product lines could be useful for assessing the prior probabilities for the *pq* parameters, for example. By referring to a possible set of demand curve

patterns, and noting the percentage of relevant products that tended to have central, left, or right skew, a simple estimate of the prior probabilities for each state is provided. In the absence of meaningful historical data, expert opinion on the probabilities associated with each deviation state can be used as the prior. The use of a credibility interval provides the needed prior probability for m. The area under the frequency distribution curve within the m^o range represents the prior probability for m^o, and since the normal distribution is symmetric, the probability for each of the m^- and m^+ states is half of the remaining area under the curve.

Since there are multiple uncertain parameters (s,m,pq) associated with this demand model, the prior probability associated with each branch of the scenario tree is a joint probability. To assess the joint priors, we assume that the distributions of the parameters are independent, i.e., $P(s,m,pq) = P(s) \cdot P(m) \cdot P(pq)$ where $P(s)$, $P(m)$, and $P(pq)$ are the marginal prior probabilities. The independence assumption is reasonable in cases where the total size of the market is not related to the growth pattern of the life cycle, or the lag time between the leading indicator and the cluster.

Computing the Stage 2 Probabilities

The probability associated with a Stage 2 node is the posterior probabilities for a specified parameter state based on its Stage 1 (prior) probability, an observation of demand realized up-to-date and a likelihood function. In other word, the Stage 2 probability is computed based on a standard Baysian update as illustrated in Figures 5 and 6. The likelihood function, $P(y/s_j,m_j,(pq)_j)$, describes the conditional dependence between the observed demand realization, y, and the parameter state $\Theta_j \equiv (s_j,m_j,(pq)_j)$. Formally stated, it is the likelihood that the parameter for the obseved sample is y if the parameter for the demand model is Θ_j. In the scenario tree, y is the up-to-date demand observed for the leading indicator at the end of period one. From this stream of observed leading indicator demands, a set of parameter values can be estimated, linking the observed demands to the states of the first stage on the scenario tree. Conceptually, this corresponds to a calibration of the entire nominal growth curve using the realized demand series. The result is a probabilistic assessment of the Stage 2 parametric state Θ_j. Conceivably, after observing a particular demand realization, a certain Stage 2 states become highly unlikely (with close to zero likelihood), thereby providing the scenario tree a basis to investigate more likely events. Specifically, the likelihood functions assigns a probability for each of the

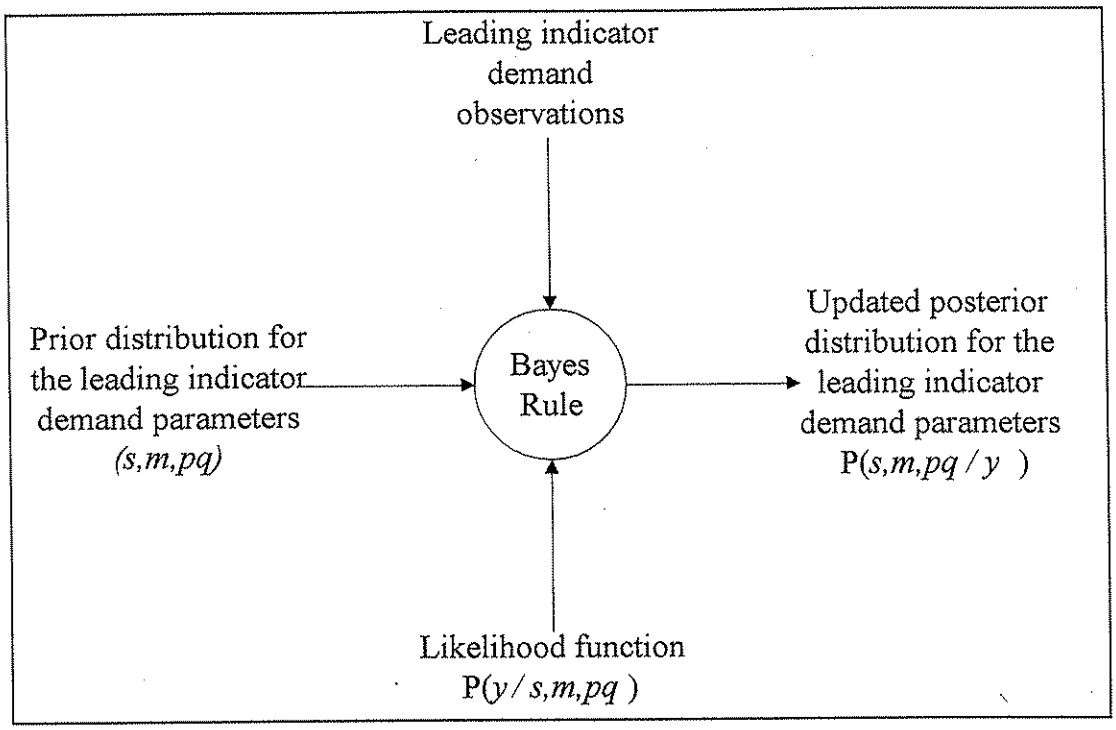


Figure 5. Baysian Update in the Demand Model

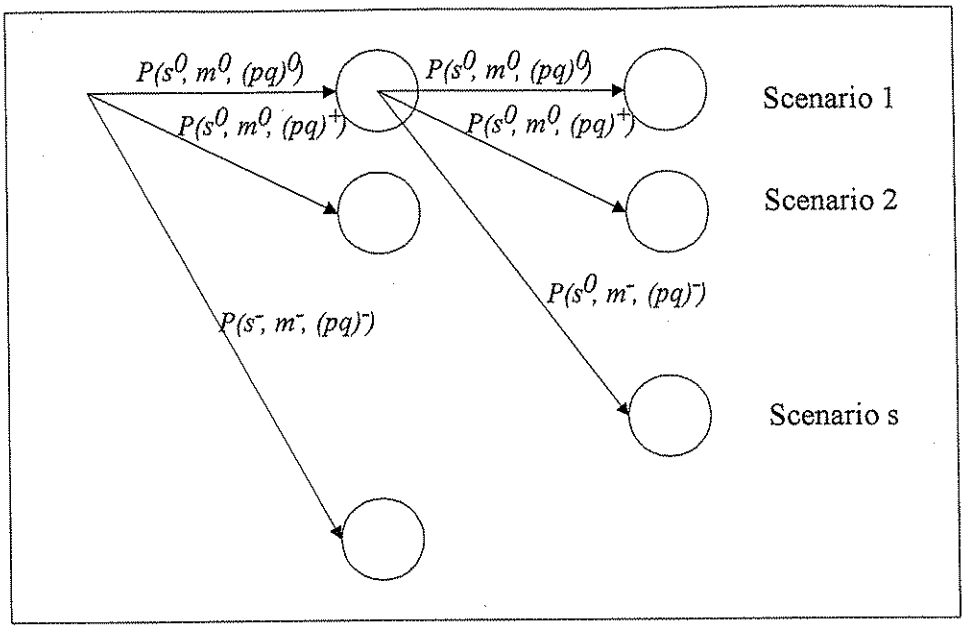


Figure 6. A two-stage scenario tree

parameter states associated with the actual demand observations, conditional on the true value of the parameter state being Θ_j. Because we use discrete probability in the tree, it is imperative that the prior states be relatively few in number, yet be collectively exhaustive in definition.

In actual implementation, the likelihoods can be assessed directly based on historical information and expert judgement. These functions would be revised over time as demand behavior is observed and rules can be derived that are in the form of likelihoods. These learned functions can be than expressed as decision rules with relative likelihoods. A rule of this type might specify that relative to other types of parameter value changes, the likelihood for a one-parameter state change from Stage 1 to Stage 2 is K times more likely than for two-parameter stage change.

With the prior probabilities and the likelihood functions for the leading indicator parameter states, the Stage 2 posterior probabilibilities are computed as follows:

$$\mathrm{P}\left(\mathrm{s}_j, m_j, (pq)_j / y\right) = \frac{P(y/\mathrm{s}_j, m_j, (pq)_j) \cdot P(\mathrm{s}_j, m_j, (pq)_j)}{\sum_{k=1}^{J} P(y/\mathrm{s}_k, m_k, (pq)_k) \cdot P(\mathrm{s}_k, m_k, (pq)_k)}$$

Computing the Scenario Probabilities

A scenario in the scenario tree is the sequence of two sets of parameter value states for the leading indicator growth model. The scenario probability is, thus, the joint probability of the events along the scenario path. The first set of parameters is the prior definition, followed by each possible second stage set of parameters. Each prior state has a prior probability associated with it, and the second stage probability is based on a Baysian update for that particular combination of first and second stage states and the demand realization. Thus, the scenario probabilities are computed as:

$$P(Sc(s)) = P(s,m,p) \cdot P[(s_2,m_2,p_2)/(s_1,m_1,p_1)]$$

The scenario structure described above can be easily implemented using a mathematical programming modeling language. For instance, Gassman & Ireland, (1995) describe an "arbitrary" scenario structure with fixed horizon using AMPL. This scenario structure carries random variable distributions, which may vary by both the time-period and prior history – a definition consistent with the demand scenario problem we consider here.

Computing the Cluster Demand Series

Since the ultimate goal of the scenario tree is to analyze demands of a broader market segment (i.e., the cluster demand) using leading indicator as a driver, we need to establish further the relationship between cluster demands and the states of the leading indicator model. As a premise of this study, we assume that a growth model exists for the demand cluster, and a stationary relationship is known between the leading indicator and the cluster demand growth curve (as depicted in Figure 1). Specifically, we need to determine the parametric state of the cluster demand growth curve for each given parameter state of the leading indicator model in the scenario tree. Given the derived parameters of cluster demand model, the demand series can be directly computed.

As a first step, we try to associate each scenario in the scenario tree with a parameter state for the cluster demand model. Specifically, the s parameter defines the time offset between the cluster and the leading indicator as follows: $t_0^{CL} = t_o^{LI} + s^{LI}$. The shape parameter *(pq)* of the cluster demand model is assumed to resemble that of the leading indicator demands. So, for example, a left-skewed demand curve for the leading indicator implies a left-skewed demand

curve for the cluster. The volume parameter m^{LI} for the leading indicator has a somewhat more complex relationship with the cluster volume parameter m^{CL}. This is the case because it is possible to have a strong but negative correlation between the cluster and its leading indicator. Using the coefficient of correlation ρ (between the leading indicator and the cluster demands), we can summarize their relationship as follows:

If $m_t^{LI} =$	**And** ρ	**Then** $\tilde{m}_{t+s}^{CL} =$
+	**> 0**	μ_{m+}^{CL}
0	**> 0 or < 0**	μ_{m}^{CL}
–	**> 0**	μ_{m-}^{CL}
+	**< 0**	μ_{m-}^{CL}
–	**< 0**	μ_{m+}^{CL}

Where μ^{CL} is the mean cluster demand estimated from the leading indicator model. The process for this statistical estimation is a separate research topic, which will not be discussed in detail here. Essentially, if enough data exists to estimate the parameters μ^{CL} using standard statistical estimation, generating the nominal values is not difficult. If not, other estimation approaches must be used. Examples of such approaches include selecting parameters based on analogous products, conducting market research to gain insight on the m parameter, e.g., Mahajan and Sharma propose an algebraic estimation when the timing of peak sales and the value of m can be estimated. Another practical alternative is to construct overtime a *catalog* of demand curve shape parameters.

Assuming a precise relationship can be correctly established between the leading indicator and the cluster demand parameters, the parametric state associated with the leading indicator model at any point in the scenario tree can be use to estimate the overall cluster demand. The strength of this approach obviously depends on the relationship one is able to establish between the leading indicator and the cluster demand. Based on our analysis on actual demand data at Lucent Technologies, it appears that a strong relationship can indeed be established in practice. In the following, we provide a numerical example of the scenario generation procedure using real data.

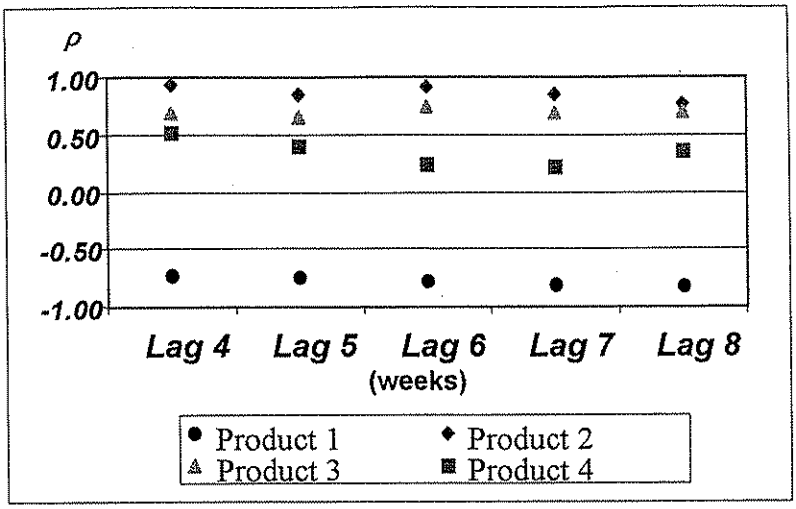
3. A Semiconductor Manufacturer Example

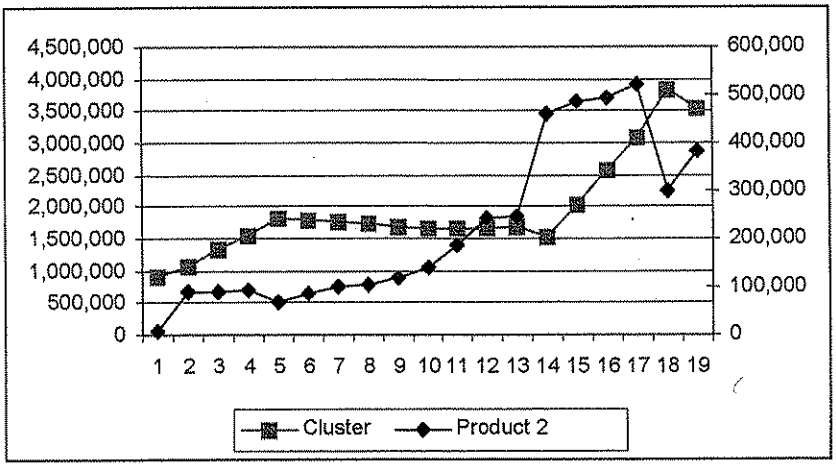
In this section we illustrate some of the finer points of demand scenario generation using a real-world data set. Our data consists of 29 weeks of sales data for 3,824 individual semiconductor items. Attributes identified for each item include the product family, business entity, and strategic business unit within which each item is managed. The associated technology and aggregate technology group is identified for each item.

Clustering Items and Identifying Leading Indicators

We first analyze the full set of 3,824 items to establish clusters that serve as the unit of aggregation in this scenario analysis. The particular similarity measure used in this example is resource-related based on technology category. The items are grouped according to their demands on resources. This similarity measure results in 6 clusters where the largest cluster consists of 1,543 items and the smallest one has 21 items. For the illustration we choose a medium size cluster with 915 items which are represented by relatively new technology products.

To identify the leading indicator for the technology cluster we perform a correlation analysis as follows: in each iteration we select a product in the cluster and its 29-week demand series. This time series is compared against the aggregate demand series for the entire cluster and a correlation is computed for each time-lag (s) from 0 to 8. This calculation is repeated for each product in the cluster. Figure 7 shows the correlation and the actual time series for two example products from the semiconductor test data. The data represents sales figures for the example cluster and several of its composing products. Since sub-groups are used as potential leading indicators, the total cluster sales quantities have been adjusted by removing from each period's sales figure of the sub-group's quantity. In this way, bias that might be introduced from a dominating sub-group is eliminated. We are in effect testing the products that comprise the technology group to see how well each predicts the rest of the technology group demand. Figure 7 shows the strength of the correlation between the cluster product and each of four different potential leading indicators. Note the varying strength of relationship among the different indicators at the different time lags. It appears that the Product 2 has a strong, positively correlated leading relationship with the technology cluster for all time lags from 4 through 8





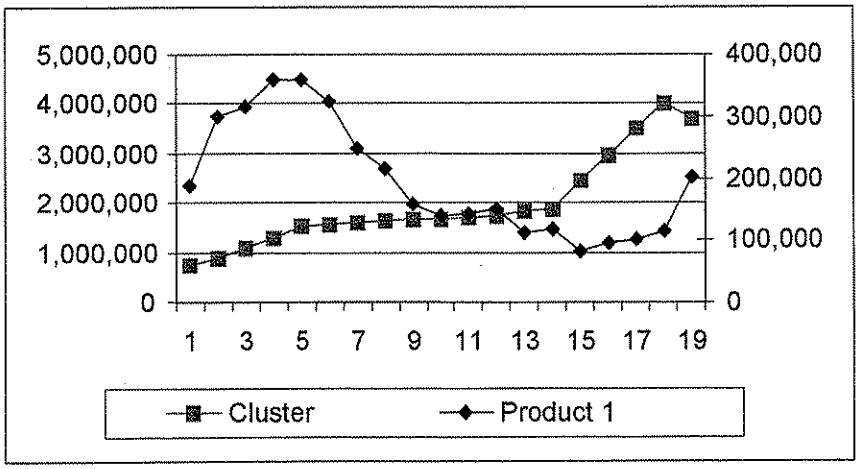


Figure 7. Correlation between product and the technology group

weeks. An 8-week lag is chosen ($s=8$) for the remainder of the example using Product 2 as the leading indicator.

Initialize Parameter Values and Stage 1 Probabilities

Given the leading indicator product and the historic data associated with it, we compute parameter values and the prior probabilities for the first stage states. To do this, a nominal demand model is specified for Product 2 by estimating the nominal values of p,q, and m using Mahajan and Sharma's algebraic estimation technique. Details of parameter initialization and some of the assumptions used are given in Meixell (1998). Table 1 summaries the parameter values for the leading indicator demand model. With the parameter settings, the priori probability for each state (s,m,pq) can be easily computed under the independence assumption.

Table 1. Parameter Values for the Leading Indicator Demand Model

	State (–)	Nominal State	State (+)
Parameter s	< 8 (0.2)	8 (0.4)	>8 (0.4)
Parameter m	≤ 7,015,000 (0.05)	7,335,000 (0.90)	≥ 7,655,000 (0.05)
Parameter p	.00578 (0.2)	.00231 (0.6)	.00058 (0.2)
Parameter q	.279 (0.2)	.279 (0.6)	.279 (0.2)

Computing the Stage-2 and the Scenario Probabilities

Using the prior probabilities, the demand realized up-to-date, and a likelihood function, we compute the conditional probabilities for each second stage state in the scenario tree. The likelihood function provides an entry for expert judgement in scenario tree construction. To illustrate this point, we use a set of simple rules for the likelihood calculation:

- The likelihood is set to 50 when the second stage state is the same as the first stage state, (i.e. all three parameters are the same in the two stages), set to 10 when two parameter states in the second stage equal the first stage, set to 1 when one parameter state in the second stage equals the first stage, and 0 when no parameter states are equal in the two stages
- The likelihood is set to 0 for impossible sequences of parameter states.

Using the above likelihood assumptions, when the observed data for the leading indicator suggests a demand state that is the same as the first stage state (s^0, m^0, pq^0), the posterior probability increases greatly (.8643) relative to the prior (.216). On the other hand, when the observed data suggest a completely different second stage state, the likelihood is zero, thus the posterior probability for that sequence is also zero. Some of the posterior probabilities are very small (0.0003) and can be eliminated from the tree. Carry through with this calculation results in only 29 scenarios with non-zero probabilities. The result calculated from the example is summarized in Table 2. The result is quite intuitive, given different possible demand observations y, and the likelihood function, the posterior probability provides an updated assessment of scenario probabilities. If the insights given in the likelihood rules are correct, the scenario tree has been successfully pared down to focus on a small fraction of more likely events given the observation.

Computing cluster demand given the leading indicator demand

As a last step, the demand series for the entire product cluster are computed for the leading indicator demand associated with each scenario. We first show how the parameter values may be established for each state of the cluster demand model. We illustrate the simple rules for determining the state of the cluster demand model in each stage for each scenario based on the state of the leading indicator demand model. With this information, the demand series can be directly computed.

We first establish a nominal demand model for the entire cluster with parameter values p, q, and m. This is typically a result of forecasting. To make sure the Bass diffusion model can indeed describe the demand growth pattern for the given set of data, we performed an ordinary least squares estimation. The P-value associated with the model is 0.0001, and so there appears to be a statistically significant relationship between the variables at the 99% confidence level. The R^2 statistic for the analysis indicates that 80.6% of the variability can be explained by the model. The standard errors are not reported here because this procedure underestimates the error. Mahajan, Mason, and Srinivasan (1986) report on the standard error problems with the estimation techniques for the Bass diffusion model.

The deviational values for the p and q were assessed as with the leading indicator by testing various parameter settings until the desired shift was achieved. Note that, we vary the

Table 2. Computation of Scenario Probabilities

Observation y				Posterior Θ				Posterior	Stage1	Scenario
s	m	p,q		s^2	m^2	$(p,q)^2$		Probability	Probability	Probability
0	0	0		0	0	0		0.882	0.333	0.294
0	0	0		0	+	0		0.059	0.333	0.020
0	0	0		0	-	0		0.059	0.333	0.020
0	-	0		0	-	0		0.556	0.019	0.010
0	-	0		0	0	0		0.333	0.019	0.006
0	-	0		0	+	0		0.111	0.019	0.002
0	+	0		0	+	0		0.556	0.019	0.010
0	+	0		0	0	0		0.333	0.019	0.006
0	+	0		0	-	0		0.111	0.019	0.002
0	0	-		0	0	0		1.000	0.019	0.019
0	0	+		0	0	0		1.000	0.019	0.019
-	0	0		-	0	0		0.882	0.167	0.147
-	0	0		-	+	0		0.059	0.167	0.010
-	0	0		-	-	0		0.059	0.167	0.010
-	-	0		-	-	0		0.625	0.009	0.006
-	-	0		-	0	0		0.375	0.009	0.003
-	+	0		-	+	0		0.625	0.009	0.006
-	+	0		-	0	0		0.375	0.009	0.003
+	0	0		+	0	0		0.882	0.333	0.294
+	0	0		+	+	0		0.059	0.333	0.020
+	0	0		+	-	0		0.059	0.333	0.020
+	-	0		+	-	0		0.556	0.019	0.010
+	-	0		+	0	0		0.333	0.019	0.006
+	-	0		+	+	0		0.111	0.019	0.002
+	+	0		+	+	0		0.556	0.019	0.010
+	+	0		+	0	0		0.333	0.019	0.006
+	+	0		+	-	0		0.111	0.019	0.002
+	0	-		+	0	0		1.000	0.019	0.019
+	0	+		+	0	0		1.000	0.019	0.019
									Total=	1.000

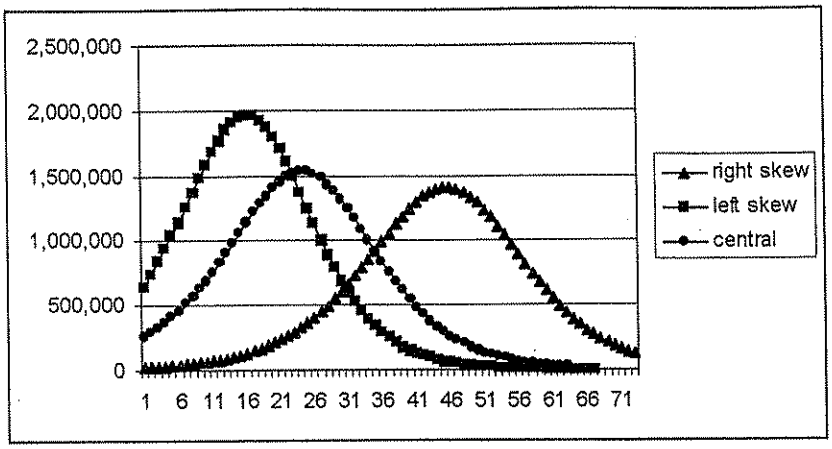


Figure 8. Estimated demand curves for the technology

parameter p, but leave the parameter q as a constant, yet still achieve the desired skew results. Figure 8 illustrates the three demand shape curves using the nominal value of m.

The deviational values for m were assessed using the 90% credibility limits as defined for the leading indicator demand model. The limits for the 90% range are (36,200,000 < m < 49,400,000). The down-shift and up-shift m values listed in the table are the approximate centroids of the range to the left and the right of the established credibility limits. The resulting parameter values for the cluster are presented in Table 3. These parameters can be used to generate the entire demand series for the cluster. For instance, when the leading indicator state is (s^+, m^-, pq^+) the corresponding cluster state would be (10, 34,300,000, .00607/.131).

Table 3. Parameter Values for the Cluster Demand Model

	State (-)	Central	State (+)
Parameter s	8	9	10
Parameter m	34,300,000	42,800,000	51,400,000
Parameter p	.0243	.0005	.00607
Parameter q	.131	.131	.131

4. Relating to Alternative Methods of Scenario Generation

The proposed procedure provides an approach for analyzing demand behavior via scenarios. The scenarios are defined based on parameter deviation from a nominal growth model across all stages. As with methods that simply enumerate all scenarios, this method too generates a tree that tends to grow exponentially with the number of stages. However, the procedure differs

from other approaches in how it streamlines the computational effort based on aggregate judgement on the parametric changes overtime. Like any other Baysian estimation methods, a different likelihood function will influence the relationship between prior probability, observed demand and posterior probability and therefore the particular subset of scenarios to be examined first. A key aspect of this approach is the use of observed demands to weight the relative criticality of remaining scenarios in the tree. This approach is especially beneficial for multistage production planning problems where the number of planning periods easily leads to intractable problem size, and a way of paring down the tree at every stage is essential. It is the likelihood function that serves as a learning mechanism for the knowledge accumulated overtime for the leading indicators.

Eppen, Martin, & Schrage (1989) present a straightforward approach for selecting scenarios for their capacity planning problem based on three views of demand (standard, optimistic, and pessimistic). They include these three states for each of five periods in the planning horizon. Although seemingly small, this problem results in a total of 243 scenarios.

By comparison, in production planning for example, practitioners typically plan for up to 16 weeks into the future. Using the simple enumeration approach just described, with three states for each of these 16 periods, a total of 3^{16} (over 43 million) scenarios would result. In our proposed approach to scenario selection, the initial tree contains $(3^k)^2$ or 729 scenarios for a 2-stage $k=3$ model that comprehends the entire 16 week planning horizon. In the general case of n stages, there are a maximum of $(3^k)^n$ scenarios. In our example, the use of simple likelihood rules trim the scenario tree by 96% from 729 to 29 scenarios. If a confidence interval is to be established for the likelihood function, then a similar trimming can be accomplished with a specified confidence. Obviously, the fewer scenarios used to model the stochastic demand environment, the larger the resulting standard error of the approximation. To control the desired level of model accuracy, the decision maker may simply introduce a threshold value for the likelihood associated with each scenario and evaluate only ones that are above the threshold. In reality, the threshold would be set at a level corresponding to a satisfactory number of scenarios for a specific problem. Table 4 illustrates the likelihood threshold for the semi-conductor example and the corresponding number of scenarios to be evaluated.

Table 4. Impact of the Threshold Value on the Number of Scenarios

Threshold Value	Number of Scenarios
0	729
1	243
10	136
50	29

Without actually solving the stochastic problem, the impact of the selected threshold on the error cannot be determined. So, in this strictly *a priori* approach to constructing a scenario tree, we evaluate the likelihood histogram, treat the threshold as adjustable, and move the threshold until an acceptable number of scenarios results.

5. Conclusions and future research directions

In this paper, we present a model of manufacturing demand for the purpose of generating scenarios to be used in a stochastic program. This demand model is built upon relationships between groups of time-lagged product demands. This is in contrast to earlier work (Wu and Meixell, 1998) that describe a demand process model that mimics the decisions in order-based supply chain systems. Each of these models is useful for analyzing demand behavior in the context of the specific problems for which they are developed.

We use a time-series demand model to describe demand because it enables a full series of demand quantities to be estimated based on a small set of parameter values. A Bass diffusion model is an example of such a demand model, and this research shows that numerous demand quantities can be estimated with three parameters. A full set of scenarios can be generated with just three possible states for each of these three parameters. We treat the parameters of the demand model in this paper as random variables. By doing so, the random event at each stage is the observed demand, which translates into an estimated set of model parameters. In this way, it is the demand model parameters that define the states and the structure of the scenario tree.

Leading indicators are used here to provide early-warning information about changes in trends in cluster demand in upcoming periods. Leading indicators contain timely, useful information that reduces the size of the scenario tree. The scenario tree is built on the parameters of the leading indicator, although we ultimately want a demand series for the related cluster. The

leading indicator information is timely when the lag time is sufficiently large between the leading indicator and the cluster of interest.

The proposed approach uses expert judgement about historical demand behavior to develop a scenario tree. The likelihood functions include information about observed sequences of parameter changes in the past, and the Bayesian update incorporates this information into the scenario tree probabilities. Because of these likelihood functions, a tree built with this approach will be at least the same and most likely considerably smaller than a tree built with a randomly sampled set of scenarios.

Future research in this area could include further work on the n-stage problem. Developing these concepts to allow for more than one demand update is likely to be both practical and challenging. In practice, information is typically acquired about actual demand every planning period. This demand information often contains considerable noise, however, so there may not be sufficiently more information to justify the computational implication of adding another stage to the scenario tree. Additional research is also warranted in the use of growth models to model uncertainty in demand in manufacturing planning problems. It is quite common to see time series models used for projecting demand in production planning problems (see for example, Silver & Peterson, (1985), Johnson & Montgomery (1974), Box, Jenkins & Reinsel (1994)) but there are few applications of growth models in this context. Investigation into modeling issues as well as application are both needed.

When the scenario tree is used in a stochastic production planning model, the issue of integrating the solution stage with the generation stage needs to be further explored. In the approach presented here, the scenarios are fully specified in advance of the solution stage for the stochastic programming problem. This is in contrast the scenario aggregation approach that iterates between the scenario generation and the solution stage (see Kall & Wallace (1994)). An interesting area of research would be to investigate a hybrid of the two approaches, and determine what could be gained by gathering information about the specific data by iterating through the solution stage while building the scenario tree as described here in this paper.

Acknowledgement

This research is supported, in part by National Science Foundation grants DMI-9634808, DMI-9732173 and a grant from Lucent Technologies.

References

Bass, Frank M (1969) " A New Product Growth Model for Consumer Durables" *Management Science*, volume 15, number 5, January, 1969, pages 215-227.

Berger, J. (1980) *Statistical Decision Theory: Foundations, Concepts and Methods*. Springer-Verlag, New York.

Birge, John R. and Frangçois Louveaux, (1997) *Introduction to Stochastic Programming*, Springer-Verlag New York Inc., New York.

Box, George E.P., Gwilym M. Jenkins, Gregory C. Reinsel, (1994) *Time Series Analysis: Forecasting and Control*, 3rd edition, Prentice Hall Inc., Englewood Cliffs, New Jersey.

Boyd, Harper W., Ralph Westfall, and Stanley Stasch (1981) *Marketing Research: Text and Cases*, 5th edition, Richard P. Irwin, Inc., Homewood, Illinois.

Dupacova, Jitka, (1994) "Applications of stochastic programming under incomplete information." *Journal of Computational and Applied Mathematics*, volume 56, pages 113-125.

Eliashberg, Jehoshua, and Rabikar Chatterjee, (1986) "Stochastic Issues in Innovation Diffusion Models," in *Innovation Diffusion Models of New Product Acceptance*, Vijay Mahajan and Yoram Wind, editors, Ballinger Publishing Company, Cambridge, Massachusetts.

Eppen Gary D., R. Kipp Martin, and Linus Schrage. (1989) "A Scenario Approach to Capacity Planning." *Operations Research*, volume 37, number 4, pages 517-527, July-August.

Escudero L.F., P.S. Kamesam, (1993A) "MRP Modeling via Scenarios," *in Optimization in Industry*, T.A. Ciriani and R.C. Leachman, editors, John Wiley & Sons, Inc., New York.

Escudero, L.F., P.S. Kamesam, A.J. King, R.J.B. Wets, (1993B) "Production Planning via Scenario Modeling," *Annals of Operations Research*, Vol. 43, pages 311-335.

Gassmann H.I. and A.M. Ireland, (1995) "Scenario formulation in an algebraic modeling language," *Annals of Operations Research*, 59, 45-75.

Gnanadesikan, R. (1997) *Methods for Statistical Data Analysis of Multivariate Observations*, 2nd edition, John Wiley & Sons, Inc., New York.

Johnson, Lynwood A., Douglas C. Montgomery. (1974) *Operations Research in Production Planning, Scheduling, and Inventory Control.* John Wiley & Sons, Inc. New York.

Kalish, Shlomo, and Gary L. Lilien, (1986) "Applications of Innovation Diffusion Models in Marketing," in *Innovation Diffusion Models of New Product Acceptance*, Vijay Mahajan and Yoram Wind, editors, Ballinger Publishing Company, Cambridge, Massachusetts.

Kall, Peter, and Stein W. Wallace (1994) *Stochastic Programming*, John Wiley & Sons, New York.

Kress, George, and John Snyder (1994) *Forecasting and Market Analysis Techniques: A Practical Approach,* Quorum Books, Westport, Connecticut.

Kurawarwala, Abbas, and Hirofumi Matsuo, (1996) "Forecasting and Inventory Management of Short Life-Cycle Products," *Operations Research,* volume 44, number 1, January-February, 1996.

Mahajan, Vijay, Charlotte H. Mason, & V. Srinivasan (1986) "An Evaluation of Estimation Procedures for New Product Diffusion Models," in *Innovation Diffusion Models of New Product Acceptance*, Vijay Mahajan and Yoram Wind, editors, Ballinger Publishing Company, Cambridge, Massachusetts.

Mahajan, Vijay, and Subhash Sharma (1986) "A Simple Algebraic Estimation Procedure for Innovation Diffusion Models of New Product Acceptance" *Technological Forecasting and Social Change*, volume 30, pages 331-345.

Meade, Nigel and Towhidul Islam (1998), "Technological Forecasting- Model Selection, Model Stability, and Combining Models," *Management Science,* Vol. 44, No. 8, pp.1115-1130.

Meixell, Mary J. (1998), *Integrating Manufacturing Planning Decisions with Supply Chain Management: Models and Analysis*, Unpublished Ph.D. dissertation, Department of IMSE, Lehigh University, Bethlehem, Pennsylvania.

Lilien, Gary L., Ambar G. Rao, Shlomo Kalish (1981) "Bayesian Estimation and Control of Detailing Effort in a Repeat Purchase Diffusion Environment", *Management Science*, 27, pages 493-506.

Lin, Winnie S.W. and Benjamin M. Adams (1996) "Combined Control Charts for Forecast-Based Monitoring Schemes", *Journal of Quality Technology*, volume 28, number 3, pages 289-301, July.

Nelson, Lloyd S., (1982) "Control Charts for Individual Measurements", *Journal of Quality Technology*, volume 14, number 3, pages 172-173, July.

Press, S. James, (1989), *Bayesian Statistics: Principles, Models, and Applications*, John Wiley & Sons, Inc., New York.

Roes, Kit C.B., Ronald J.M.M. Does, and Yvonne Schurink, (1993) "Shewhart-Type Control Charts for Individual Observations", *Journal of Quality Technology*, volume 25, number 3, pages 188-198, July.

Schmittlein, D., and V. Mahajan (1982). "Maximum Likelihood Estimation for an Innovation Diffusion Model of New Product Acceptance." *Marketing Science*, volume 1, pages 57-78.

Silver, E.A. and R. Peterson. (1985) *Decision Systems for Inventory Management and Production Planning*. 2nd ed., John Wiley & Sons, New York.

Srinivason, V. and C.H. Mason, (1986) "Nonlinear Least Squares Estimation of New Product Diffusion Models." *Marketing Science*, volume 5.

Thomas, R.J. (1985) "Estimating Market Growth for New Products: An Analogical Diffusion Model Approach" *Journal of Product Innovation Management*, volume 2, pages 45-55.

Wagner, Harvey M. (1975) *Principles of Operations Research with Applications to Managerial Decisions*, 2nd edition, Prentice-Hall, Inc. Englewood Cliffs, New Jersey.

West, Mike and Jeff Harrison (1997) *Bayesian Forecasting and Dynamic Models*, 2nd edition, Springer-Verlag, New York.

Winkler, Robert (1972) *An Introduction to Bayesian Inference and Decision*, Hold, Rinehard, & Winston, Inc., New York.

Wu, S. David and Mary J. Meixell (1998) "Relating Demand Behavior and Production Policies in the Manufacturing Supply Chain," *IMSE Technical Report 98T-007*, Lehigh University.