MODELING THE EVOLUTION OF DEMAND FORECASTS WITH APPLICATION TO SAFETY STOCK ANALYSIS IN PRODUCTION/DISTRIBUTION SYSTEMS

DAVID C. HEATH AND PETER L. JACKSON

School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York 14853

In this paper, we propose a general probabilistic model for modeling the evolution of demand forecasts, referred to as the Martingale Model of Forecast Evolution (MMFE). We combine the MMFE with a linear programming model of production and distribution planning implemented in a rolling horizon fashion. The resulting simulation methodology is used to analyze safety stock levels for a multi-product/multi-plant production/distribution system with seasonal stochastic demand. In the context of this application we demonstrate the importance of good forecasting.

■ In this paper we make three contributions: 1) we propose a general probabilistic model for modeling the evolution of demand forecasts, referred to as the Martingale Model of Forecast Evolution (MMFE); 2) we describe an application of this model in a simulation study to analyze safety stock levels for a multi-product/multi-plant production/distribution system with seasonal stochastic demand; and 3) in the context of this application we demonstrate the importance of good forecasting. The simulation model in 2) is a combination of the MMFE with a linear programming (LP) model of production and distribution planning implemented in a rolling horizon fashion.

Case Study

We motivate the need for a general model of forecast evolution by considering the specific problem posed in the application. The operating company is a national producer and distributor of consumer grocery products. Any data in the case that could identify the company or its product lines have been disguised. The focus of the study is a product family consisting of five product lines that differ primarily in package sizes. The family has dedicated production plants in different regions of the country with a number of production lines of differing efficiencies and capabilities. It is also possible to outsource the production with another ten possible suppliers, although outsourcing arrangements are less flexible than internal production plans. The production plants feed a national distribution system with eight regional warehouses. Transshipments between warehouses can be used to redress inventory imbalances.

Demand for the product family is highly seasonal with a major selling season spanning six months and the peak selling season spanning three months. Due to sales promotions, one month in particular always exhibits significantly higher demand than adjacent months. In general, demand in any month following a sales promotion is negatively correlated with demand in the promotion month. This is due in part to major buyers batching their orders to take advantage of the promotions. Forecasts are made by region, by month, and by product line for twelve months in advance. Forecast error is high, even on the eve of the month of sale. Demand is highly correlated across product lines and, to a lesser extent, across regions.

Production capacity is limited and so inventory is produced in anticipation of sales beginning three to four months in advance of the major selling season. The inventory is sold on a first-in-first-out basis and, in any event, the product has a long shelf life. So much inventory is created that the company must rent additional warehouse space ("overflow"), at high cost, during the cyclic build-up phase of the season. In spite of high production changeover times, every product line is produced in every plant in every month. There is sufficient volume of some product lines to dedicate certain production lines to these products. The company uses a linear programming model for aggregate production and distribution planning by month. Plant-specific scheduling rules are used to disaggregate this plan into weekly production schedules.

The existing corporate safety stock policy for each region is expressed in months of supply at the beginning of each month: the region must have enough inventory to cover s months of forecast demand. The current value of s at the beginning of the study was significantly larger than 1. The safety stock factor, s, is a key input to the aggregate production planning LP, and a major driver of the cyclic build-up of inventory within the distribution system.

The performance attributes of the production-distribution system are cost and customer service. Major cost categories are production costs (proportional to standard hours of production), transportation costs (proportional to volume-miles), and inventory holding costs (including financing cost and a premium for overflow inventory). Customer service is measured by a weighted average of monthly fill rates across product lines and regions.

The purposes of the study were as follows. The major concern was to find an economical safety stock factor. The current, conservative value of s had been set during the product family introduction phase some years earlier when demand uncertainty had been extemely high

 0740-817X/94/$3.00x.00

and production capabilities were low. Although forecast errors were still high, the product family was now in a mature phase of marketing and production capabilities had expanded considerably. The company wanted a quantitative model of safety stock analysis that would permit them to explore the impact of varying s by product line, by region, and by month; and to explore the impact of adding additional production lines in different parts of the country.

In approaching this study, we found the stochastic, sequential, and multi-dimensional nature of the problem to defy an optimization-based approach, so we opted to develop a simulation model of the system and use that model to explore issues important to the company. Furthermore, since the company currently used a linear programming model for aggregate production planning, we proposed to use a variation of that model at the heart of the simulation to mimic the way in which production plans and distribution decisions were made. Note that no claim is made that the linear programming model yields optimal plans in the face of stochastic demands; only that it reasonably approximates actual planning decisions. In developing a simulation model of the system, it was clear immediately that a good model of the evolution of forecasts and demands was crucial to the overall model's accuracy. Intuitively, the production distribution system has many opportunities for substitution: low demand in one region of the country can be balanced by shipping product from that region to a region experiencing high demand; low demand for one product can be balanced by shifting production plans and using the capacity to produce a high demand product; low demand in one period can be balanced with high demand in another period by carrying inventory. However, the ability to translate this flexibility into reduced safety stock requirements depends on demand correlations. If demand among regions is highly positively correlated, then transshipments will be less effective; if demand among products is highly positively correlated, then reallocating production capacity will be less effective; and if demand from one period to the next is highly correlated, then inventory becomes less effective. Furthermore, since production and inventory plans are based on forecast demand, the correlations between changes in forecasts by region, product, and time period also become important. The forecast evolution model we propose and applied in this study captures these correlations.

We fitted the simulation forecast model to four years of historical forecasts and forecast errors to represent the behavior of the company's existing (human) forecasting system. In an interesting twist to the study, we discovered that the company had developed a detailed quantitative model for forecasting near-term demand on a weekly basis, taking into account such things as local promotions and competitors' prices. When extended to a long term forecast model for monthly demands, the model appeared to have many desirable properties. Accordingly, we simulated the performance of the production distribution system under this model as well and demonstrated the cost and customer service impact of this improved forecasting system.

Management scientists within the company conducted extensive simulation studies using this model. Based on their recommendations and ours, the company reduced the safety stock factor dramatically and accelerated the implementation schedule for the new forecasting system. One year later, the company has achieved the predicted cost reductions and still maintains the desired customer service level. While anecdotal in nature, this case study illustrates that the proposed approach is practical and that it is effective in demonstrating the cost and service impact of improving forecast accuracy.

In the next section, we review the literature that is related to this problem and observe the differences from our approach, and briefly describe the simulation model. Details of the LP production planning model are deferred to the Appendix. Then, we describe the Martingale Model of Forecast Evolution, illustrate the method with specific results from the application, and summarize some of the simulation results from the application.

Literature Review

Hausman [14] suggests modeling the evolution of forecasts as a quasi-Markovian or Markovian system. As a specific application, he suggests modeling a series of ratios of successive forecasts for the same quantity as independent lognormal variates. The independence implies the quasi-Markovian property. Hausman reports on several statistical studies of actual forecasting systems, both human and mechanical, that support the lognormal model and he suggests two rationales to explain the phenomenon. One of his rationales, that is also used in the finance literature to justify the geometric Brownian motion model of stock prices, will be used in this paper as the foundation for a more general model. Hausman suggests using the independent conditionally lognormal forecast ratio model in dynamic programming approaches to sequential decision problems. Hausman and Peterson [15] formulate a production scheduling problem for multiple products in a single capacity-constrained facility with a single selling season but with multiple production/selling periods. The forecasts for total sales in the selling season evolve over the production/selling periods according to the lognormal model. The dynamic programming state space is too large for practical computation so they propose heuristic solution techniques.

The Hausman model seems not to have generated the sort of research interest that we believe it deserves. The Martingale Model of Forecast Evolution, proposed here, can be seen as an extension of the Hausman model. It fits within the framework of a quasi-Markovian system but is more general than the specific model proposed by Hausman. In particular, it accomodates the simultaneous evolution of forecasts for demand in many time periods. Consequently, we are not limited to the single selling season models considered by Hausman and Peterson. Our model captures correlation in forecasts between products and between time periods. This is particularly important for the application we describe. The model can capture both times series models of prediction as well as the expertise of the human forecaster. In our case study, as in Hausman's studies, the lognormal model provides a good fit for the behavior of ratios of successive forecasts in the industrial data we studied.

A competing approach would be a Bayesian model of evolving estimates of demand distribution parameters. Scarf [22] initiated the study of Bayesian inventory models with a dynamic programming model of a single item periodic review inventory problem in which the demands in each period are independent and identically distributed random variables with an unkown distribution function. See Azoury [1] for recent extensions to the Bayesian inventory model. The Bayesian approach is applied by Bitran, Haas, and Matsuo [5] to a variation of the single selling season ("style-goods") problem in which there is a single selling period and each product family is produced in only one period prior to this. They formulate a stochastic mixed-integer programming problem to plan production and use a hierarchical approximation scheme that is easier to solve. An application of this approach to a consumer electronics company reveals that the model results in production plans that defer production of product families with initially high forecast errors to late in the season when forecast errors are smaller. Other Bayesian approaches to the style goods problem are reviewed and proposed in Bradford and Sugrue [7].

Another approach is to consider the use of time series models of forecasting demand. These models are described by Box and Jenkins [6]. For example, Johnson and Thompson [20] extend the results of Veinott [23] to show the optimality of myopic order-up-to inventory policies when demand is given by a stationary autoregressive moving-average (ARMA) process. Erkip, Hausman, and Nahmias [13] extend the approach to a depot warehouse system and derive a closed form expression for the optimal order-up-to system inventory level. Badinelli [2] argues that exponential smoothing techniques are routinely applied in practice when an examination of the pattern of autocorrelations would actually suggest an ARMA process. He demonstrates that such a mis-specification of the demand process results in a substantial inventory cost penalty under a fixed order interval, order-up-to-S type inventory policy. Exponential smoothing is developed for non-stationary processes. ARMA models are developed for stationary processes. By this interpretation, the model we develop is for a stationary process. As mentioned, the Martingale Model of Forecast Evolution can represent an ARMA process as well as human forecasting systems. We will argue that the MMFE approach is to be preferred to a direct times series approach in modeling the behavior of forecasting-production-distribution systems because of its potential to capture the impact of such factors as expert judgment.

A common approach in practice for single item problems is to assume the form of a demand distribution (e.g., Poisson or normal) and use statistical estimates of the distribution parameters in the calculation of inventory policy parameters, using for example the power approximation method of Ehrhardt [12]. Periodically, these estimates are revised based on more recent demand data and the safety stock policy parameters are recomputed. The statistical estimates commonly used would be the standard sample estimates of the mean and standard deviation. Jacobs and Wagner [18] demonstrate that when demand variability is high, exponentially smoothed estimators of the distribution mean and variance result in lower inventory costs. This is because these estimators are less sensitive to extreme values. Iyer and Schrage [17] take a different approach and suggest developing (s,S) parameters by optimizing the deterministic (s,S) inventory problem using historical demands. They demonstrate that the (s,S) parameters generated in this way outperform the parameters generated by optimizing the infinite horizon (s,S) using a statistical estimate of the long run demand rate based on the historical demands. The method also performs well when serial correlation is present in the demand process. All of these papers indicate that imperfect knowledge of the form or parameters of the demand process is an important consideration when applying inventory models.

The combined use of a sophisticated model of forecast evolution with a linear programming model of production planning in a rolling horizon is not new. Dzielinski, Baker and Manne [11] report on a simulation study in which the past history file of orders was used as input in a rolling horizon fashion to an exponential smoothing technique for forecasting orders. Orders were forecast as many periods into the future as required by the production planning technique, a linear program that considered setup costs, inventory costs, shortage costs, labor costs, and hiring and firing costs. The production decisions for the first period of the planning model were taken as the implemented decisions, in a simulation, and the simulation clock was advanced to the next period. Although they report the impact of two different levels of protection in the safety stock policy in their study, the thrust of their paper was to recommend the use of optimization-based aggregate production planning techniques. Later studies such as Lee and Khumawala [21] also use aggregate production planning models in a simulation study but the simulation is a test vehicle for examining the quality of alternative production planning heuristics.

There is considerable emphasis in the literature of production planning models on planning horizons. A planning horizon exists if extending the model plan beyond this horizon has no impact on the decisions taken in the first period of the plan. (See, for example, Bean and Smith [4].) Relatively less has been said about the overall effectiveness of rolling horizon models. Baker and Peterson [3] review this literature and, for a simplified model, study exact dynamic programming solutions. Here again, the focus is on the length of the planning horizon.

Our use of simulation combining forecast evolution models with rolling horizon, optimization-based production planning differs from these previously reported studies in several ways. One difference is the purpose of the study. The operating company under study already used linear programming production planning models and did not need to be convinced of their value. Also, because of the highly seasonal nature of the business, the choice of planning horizon was not a major issue. Every twelve month period included a month of near-zero inventory so production decisions in the current month would have negligible impact on inventory status after twelve months. What was of concern was the level of safety stock in the system and a simulation study seemed the most viable approach to adequately describing the system dynamics. The need for safety stock in the distribution system is critically related to the flexibility of the production system with regard to changes in forecast demand. Accordingly, it was important in the simulation to model the evolution of the forecasts as carefully as possible. Our forecast evolution model is therefore more detailed than in any previously reported study. The study is also distinguished from past reported studies in its magnitude. The earliest study of this type, Dzielinski, Baker and Manne, was limited to 70 constraints in the LP. Advances in computing hardware and software have made the routine use of 2000 constraint LP's in a simulation study practical.

Simulation studies are not the only type of application for the forecast evolution model that we propose. Research is underway using dynamic programming to understand the form of optimal policies under such a model and to reduce the state space requirements of computational dynamic programming approaches to the problem.

The Simulation Model

As the simulation progresses through time, it will track the evolution of inventory, production and shipment decisions, and demand. Since production and shipment decisions depend upon planned decisions for the future, the simulation will also generate a production and shipment plan as well as forecasts of future demand. The forecasts are generated by a program called SIMFORECAST. The methodology underlying this program is described in the next section. The production and shipment decisions for the current period of the simulation as well as the planned production and shipment decisions for future periods are generated by the linear programming model called SIMLP. After each period is simulated, the simulated demand observations, the inventory, production, and shipment decisions, the cost summaries, and the customer service observations for the period are stored in a file for subsequent statistical analysis.

SIMLP

SIMLP is a multi-location, multi-time-period model of the production, shipment, and inventory activities. The first period in the model, indexed by $t=1$, represents the current period of the simulation. Decisions variables for this period represent decisions made in the current period of the simulation. These decisions affect, among other things, the cost and customer service reported by the simulation for the current period. In other words, the simulation will *implement* the decisions for the current period. Decision variables for future periods in the model represent planned decisions. The simulation will not implement those decisions. When the simulation advances to the next period, it will solve a new LP and implement the decisions from the first period of the new model.

SIMLP is a variation on a standard linear programming model of production and distribution. A simplified version of the formulation is summarized in the Appendix. The actual implementation included greater detail to handle the use of overflow warehouses, and the limited flexibility associated with using certain production lines (copackers).

The Martingale Model of Forecast Evolution

In this section, we develop a technique for modeling the results of forecasting procedures. This will show that under simple and plausible assumptions we are led to a class of models which is very general, yet very simple. The assumptions we make (or very similar assumptions) underlie most forecasting methods. In particular, for methods based only on previous demands, see the discussion in Brockwell and Davis [9, Chapter 5]. Although it might seem desirable to develop a stochastic model for these quantities based on the particular statistical method which is used to produce the forecasts, this turns out (in many cases) to be unnecessary. It is indeed fortunate that this is so, for many forecasting techniques are based on more data than past demands. For example, prices of competing goods, and marketing, advertising, and other promotional plans, and sometimes even expert judgement are used to produce the forecasts on which decisions are based. For many reasons (lack of data, the difficulty of modeling competitors' price changes, and the obvious problem of modeling expert judgement) it would be very difficult to produce and fit a model in this way, and to have confidence in the resulting model.

Notation

To make things precise we need some notation: For every pair (s,t) of times, we denote by $D_{(s,t)}$ the predictions made at time s for the amount(s) demanded at time t. (If we are considering demands for more than one type of good, each $D_{(s,t)}$ will be a vector.) For $s < t$, these are genuine predictions, while for $s \geq t$ they are "predictions of past demands" and thus are equal to the past demands. Of course, $D_{(s,s)}$ (which we sometimes write as D_s) is simply the actual demand at time s.

If each vector $D_{(s,t)}$ contains N entries, (i.e., if we are predicting demands for N goods), then for each s we construct the (infinite) vector X_s whose first N entries are $D_{(s,s)}$, whose next N entries are $D_{(s,s+1)}$, and so on. The vector X_s is simply a list of the current demands and the forecasts for all future periods. We now focus attention on the changes in the X vectors. We develop two classes of models for the behavior of these changes: the additive model and the multiplicative model. We explain the simpler of these two first, the additive model, although the multiplicative model will likely be the more useful in practice.

The Additive Model

For the additive models, we define the N-vector ϵ_{st} by

$$\epsilon_{st} = D_{(s,t)} - D_{(s-1,t)},$$

for $t \geq s$. We construct the infinite vector ϵ_s analogously to the vector X_s: $\epsilon_s = (\epsilon_{st})_{t=s}^{+\infty}$. Thus, the k-th coordinate of ϵ_s is the k-th coordinate of X_s minus the $(k+N)$-th coordinate of X_{s-1}. Clearly, each coordinate of ϵ_s represents the change from time $s-1$ to time s in the prediction of some demand occurring at or after time s. Moreover, ϵ_s is known at time s. Figure 1 illustrates: each row shows the successive forecasts and forecast changes for a given time-dated demand vector. Forecasting ceases once the actual demands are observed.

The Martingale Model of Forecast Evolution in this case produces a model in which the ϵ vectors are independent, identically distributed, multivariate normal random vectors with mean 0. The only model parameters are the variance-covariance matrix for the distribution of each ϵ vector and the initial state of the system, X_0.

Derivation of the Martingale Model of Forecast Evolution

The result in the additive model that the ϵ vectors are independent, identically distributed, multivariate normal random vectors with mean 0 follows from a sequence of assumptions each of which narrows the domain of applicability of our methodology.

The first assumption we shall make is that the information available to make predictions at time s grows as s increases. That is, at each time s, we suppose that a certain amount of information $\mathscr{I}_s$, is known and that $\mathscr{I}_s \subseteq \mathscr{I}_{s+1}$.

The second assumption requires some discussion, as it can be justified in at least two different ways. The assumption is that ϵ_{s+1} is uncorrelated with the information in $\mathscr{I}_s$, and hence is uncorrelated with all ϵ_u for $u \leq s$. Furthermore, $E[\epsilon_{s+1}] = 0$.

This second assumption will be satisfied in two important cases. In the one case, it will be satisfied if the predictions made are the conditional expectations of the variables to be predicted given the available information. If one uses the most general prediction techniques, then minimum mean-squared error predictors are conditional expected values. In this case, we assume that $\mathscr{I}_s$ is a σ-field describing the knowledge available at time s, and the prediction of any future random variable Z is its conditional expectation $E[Z \mid \mathscr{I}_s]$. In this case the successive predictions for Z form a martingale. It is this assumption that leads to the name for methodology. Hence,

Forecasting	Start with: Initial Forecasts X_0	Add: Forecast Change ϵ_1	To Get: Period 1 Forecasts X_1	Add: Forecast Change ϵ_2	To Get: Period 2 Forecasts X_2	Add: Forecast Change ϵ_3	To Get: Period 3 Forecasts X_3
D_0	$D_{(0,0)}$						
D_1	$D_{(0,1)}$	$\epsilon_{(1,1)}$	$D_{(1,1)}$				
D_2	$D_{(0,2)}$	$\epsilon_{(1,2)}$	$D_{(1,2)}$	$\epsilon_{(2,2)}$	$D_{(2,2)}$		
D_3	$D_{(0,3)}$	$\epsilon_{(1,3)}$	$D_{(1,3)}$	$\epsilon_{(2,3)}$	$D_{(2,3)}$	$\epsilon_{(3,3)}$	$D_{(3,3)}$
D_4	$D_{(0,4)}$	$\epsilon_{(1,4)}$	$D_{(1,4)}$	$\epsilon_{(2,4)}$	$D_{(2,4)}$	$\epsilon_{(3,4)}$	$D_{(3,4)}$
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.

Figure 1. Forecast evolution: additive model

$$\epsilon_{st} = D_{(s,t)} - D_{(s-1,t)} = E[D_{(t,t)} | \mathcal{I}_s] - E[D_{(t,t)} \mathcal{I}_{s-1}].$$

Assuming all random variables are square-integrable, it is easily shown that the martingale property implies that ϵ_{st} is uncorrelated with any random variable measurable with respect to $\mathcal{I}_{s-1}$ and that $E[\epsilon_{st}] = 0$.

Alternatively, the second assumption will be satisfied if we make the weaker assumption that the predictions are the minimum mean squared error linear predictions based on a set of observable random variables which grows with time and which includes the current actual observations. In this case, the set $\mathcal{I}_s$ consists of those random variables which have been observed by time s. In either case we assume that all past and present demands are included in the information available at time s. A fundamental property of the minimum mean squared error predictor is the following:

Proposition: Suppose $s_1 \leq s_2 \leq t$. If the predictions are minimum mean-squared error predictions, then the change in predictions, $D_{(s_2,t)} - D_{(s_1,t)}$, is uncorrelated with any of the random variables in $\mathcal{I}_{s_1}$ and has mean 0.

Proof: The result follows, for example, from the discussion entitled "The Prediction Equations" in Brockwell and Davis [9, p. 53]. In their notation, $\hat{X}$ is the minimum mean square error predictor of X, so let X denote $D_{(t,t)}$, let $\hat{X}_1 = D_{(s_1,t)}$, the prediction of $D_{(t,t)}$ made at time s_1, and let $\hat{X}_2 = D_{(s_2,t)}$, the prediction of $D_{(t,t)}$ made at time s_2. Let ϵ denote any one of the random variables observed by time s_1 (also observed by time s_2, since $s_1 \leq s_2$). Taking $Y = \epsilon$ in Brockwell and Davis (2.3.8) yields $E[(X - \hat{X}_1)\epsilon] = E[(X - \hat{X}_2)\epsilon] = 0$, so $E[(\hat{X}_2 - \hat{X}_1)\epsilon] = 0$. That is, the *change* in predictions is uncorrelated with any observed ϵ. Taking $\epsilon = 1$ yields $E[\hat{X}_2 - \hat{X}_1] = 0$. □

The second assumption, that changes in forecasts are mean zero and uncorrelated with past observations, describes a desirable property of a forecasting system. In fact, if the second assumption is not satisfied it will be possible to construct improved predictions (in the mean squared error sense) as a linear combination of the given predictions, the current observations, and previous changes. Replacing the given predictions by the best improved predictions ensures that this second assumption will be satisfied.

The third assumption is one of stationarity. We assume that the changes in the predictions (i.e., the ϵ vectors) form a stationary stochastic process. As is usual in time series modeling, "stationary" has several meanings. If one is interested only in properties which depend only on the first two moments of all variables, then "weak stationarity" (sometimes called "covariance stationarity") suffices. If one is interested in all possible properties, "strict stationarity" is needed. Under assumptions of normality, weak stationarity implies strong stationarity. In either case (considering only properties depending on first and second moments, or assuming normality and considering properties which may depend on the distributions), all important model properties can be captured by the variance-covariance matrix of the vectors.

There are many techniques for transforming a time series to make it more nearly stationary. Differencing, which is what these forecast changes represent, tends to improve stationarity and can never destroy it. Most time series texts give good discussions of techniques for modifying or decomposing a time series to obtain more nearly stationary time series, and many of these techniques are useful in our setting. In particular, the multiplicative model presented below was useful in the case study for improving stationarity.

The fourth and final assumption is that of normality (or alternatively the assumption that whatever we do with our predictions depends only on the first and second moments of these random variables).

Almost all of the existing forecasting techniques are based on assumptions stronger than those required here; thus, our model is appropriate whenever these forecasting methods are applicable.

The Multiplicative Model

The above model represents the changes due to new information as additive. In particular, the size of the changes in forecasts is unrelated to the sizes of the forecasts. In the data of the case study, however, we observed that the standard deviation of forecast error was roughly proportional to the size of the forecast. (Cf. Brown [10, p. 94] "You will be very likely to find that the standard deviation of demand is nearly proportional to the total annual usage, or to the average monthly usage." See also his Appendix C on applications of the Lognormal distribution to inventory models.) (See Hausman [14] for documented examples.) Since the forecasts were highly seasonal this meant that the stationarity assumption was violated. This observation suggested a log transformation would be appropriate to improve stationarity. If all of the demands (and hence also the forecasts) are strictly positive, a multiplicative model can be obtained by modeling the logarithms of the forecasts and demands in an additive way.

For the multiplicative model, (assuming all the data values to be non-negative) we define ν_{st} by:

$$\nu_{st} = \log(D_{(s,t)}) - \log(D_{(s-1,t)}),$$

where the logarithms are taken componentwise. We construct the infinite vector ν_s analogously to the vector X_s: $\nu_s = (\nu_{st})_{t=s}^{+\infty}$. Thus, letting $Y_s = \log(X_s)$, componentwise, the k-th coordinate of ν_s is the k-th coordinate of Y_s minus the $(k+N)$-th coordinate of Y_{s-1}. Clearly, each coordinate of ν_s represents the change from time $s-1$ to time s in the log of the prediction of some demand occurring at or after time s.

Conversely, let $R_s = \exp(\nu_s)$, taken componentwise. Then, a coordinate of R_s, represents the ratio of successive forecasts for some demand occurring at or after time s. If we assume, as before, that successive forecasts of a future demand form a martingale process, then the expected value of each future forecast is the same as the current forecast and the expected value of the ratio of these successive forecasts is 1. Furthermore, R_{s+1} is uncorrelated with R_u for $u \leq s$.

Next, instead of assuming that the forecast differences are jointly normally distributed, we assume that the components of ν_s are jointly normally distributed. It follows that the mean of each component of ν_s must equal the negative of one half of the variance of that component. (If η is $N(\mu,\sigma^2)$ then $E[\exp(\eta)] = \exp(\mu + \sigma^2/2)$. Setting $E[\exp(\eta)] = 1$ yields $\mu = -\sigma^2/2$.) It also follows that the ν's are uncorrelated. To see this, note that if η and ζ are jointly normally distributed, then we have: $\text{cov}(\exp(\eta),\exp(\zeta)) = \exp(\text{cov}(\eta,\zeta)) - 1$. Hence if $\exp(\eta)$ and $\exp(\zeta)$ are uncorrelated, then η and ζ must be uncorrelated.

In summary, for the multiplicative model, we suppose that the vectors ν are independent, identically distributed multivariate normal random vectors with the mean of each coordinate equal to the negative of one-half of its variance. Thus, once again, the model does not require (nor allow) the exogenous specification of means. The multiplicative model requires only the specification of variances and covariances and the initial state, X_0.

The Forecast Update Horizon

We now impose one further restriction: we suppose that there is some finite horizon M such that only the first MN components of each ϵ (respectively, ν) vector are non-zero. That is, forecasts for the N products are not updated until time has advanced to within M periods of the period being forecast. This is necessary to allow estimation and computation, and should, for M large, provide a reasonable approximation in a practical setting. Under this assumption, the model is completely specified by an $MN \times MN$ variance-covariance matrix and the initial state.

We stress that in each of the two models (additive and multiplicative), it is the change vectors (ϵ or ν) which are independent, and not the components of each vector. Once the variance-covariance matrix is known, the means of these vectors are determined. Thus, for each model, the only model parameters are the variance-covariance matrix for the appropriate multivariate random vectors and the initial state of the system, X_0.

Obtaining the Variance-Covariance Matrix

This variance-covariance matrix can be obtained in several ways. If the predictions are obtained from a moving average time series model, then the matrix can be computed directly. In this case, forecasts are available for all future periods, so the matrix should have infinitely many entries. For practical purposes, it would be necessary to truncate the matrix as discussed above. Causal autoregressive moving average models are equivalent to moving average models with infinitely many lags. (For this equivalence, see Brockwell and Davis [9], Def. 3.1.3, p. 83; Thm. 3.1.1, p. 85; and Remarks on p. 86. For extension to the multivariate case, see Thm. 11.1.1, p. 408). Truncating the series should provide a reasonable approximation.

If, instead, the predictions arise in a more complex way (using not only past demand data, but other data and perhaps expert opinion), past data on forecasts and demands should allow estimation of the variance-covariance matrix. From the past data, one can easily produce the sample ϵ (respectively ν) vectors. Knowing that the mean of the vector ϵ must be zero simplifies estimation of the variance-covariance matrix. For estimating the variance-covariance matrix of the ν vector (under the restriction that the means are related to the variances) we have used a simple method of moments estimator (matching the averages of the outer products of the sample ν vectors).

Simulating Forecast Evolution

To simulate multivariate normals, Bratley, Fox, and Schrage [8] suggest finding a matrix C such that CC' is the desired variance-covariance matrix. Given C, the multivariate normal is generated by $\mu + CZ$ where Z is a standard normal random vector (i.e., whose components are independent standard normal random variables).

There are many choices for finding C; we prefer to represent the variance covariance matrix as UDU' where U is a real unitary matrix (so that $U' = U^{-1}$) and D is a diagonal matrix with non-negative entries which are decreasing down the diagonal. (For existence of this representation, see Hoffman and Kunze [16], Theorem 20 and corollary, p. 266.) Thus, C can be chosen to be $UD^{1/2}$.

If C has been represented as above, then clearly CZ is a random weighted sum of the columns of C (with independent standard normal weights). Moreover, the columns of C are arranged in decreasing order of size. Thus, the changes in the forecasts have been represented as sums of independent random vectors whose "sizes" are decreasing. This is the "principal components representation." The first component (the first column of C) describes the "largest" sort of changes which occur; the second the second largest; and so on. These components can give some idea of the type of information which is revealed from period to period. As a simple example, if all of the entries of the first column of C have the same sign, this means that the largest component of the new information revealed is that all demands (present and future) will tend to be be larger, or all will tend to be smaller, than was previously predicted. A later section includes an examination of the first principal component of the variance-covanance matrix for two applications of this analysis.

Examples

As a simple example, consider an additive model for the demand for a single product. For the simplest case, suppose that $M=1$; i.e., that only the first component of the vector can be non-zero. Suppose further that the initial state vector is of the form $(x_1,\bar{x},\bar{x},\bar{x},\ldots)'$; i.e., today's demand is x_1, and all predicted future demands are $\bar{x}$. The model is then characterized by the 1×1 variance-covariance matrix whose only entry is the variance of x_1. This model corresponds to the case of independent demands with no forecasting updates until the actual demand is observed.

A more interesting class of models results for the single product model with $M=2$. Suppose that the initial state is $(x_1,x_2,\bar{x},\bar{x},\bar{x},\ldots)'$. Then the model is characterized by the 2×2 matrix, Σ:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}$$

where $\sigma_{12}=\sigma_{21}$. Letting $(X_s)_j$ denote the j-th element of X_s, the state vector at time s, we have (for $s \geq 2$),

$$(X_s)_1 = \bar{x} + (\epsilon_s)_1 + (\epsilon_{s-1})_2$$

$$(X_s)_2 = \bar{x} + (\epsilon_s)_2$$

$$(X_s)_j = \bar{x} \text{ for } j \geq 3$$

Notice that the two ϵ values appearing in the first equation are independent random variables since they are components of ϵ-vectors observed at different times, which are independent. However, the first two *components* of the X vector are typically not independent, since they both depend on the vector ϵ_s. In fact, their covariance will be precisely σ_{12}. It follows easily that the covariance of $(X_s)_1$ and $(X_{s+1})_1$ is also σ_{12}. Thus if $\sigma_{12}=0$, we still have independent demands, but some of the demand variation is predictable one time period in advance. In this special case, the fraction of the variability which is predicted is exactly $\sigma_2^2/(\sigma_1^2+\sigma_2^2)$.

Why Not Use Time Series Models?

This example illustrates why the Martingale Model of Forecast Evolution could be more useful in a simulation of a production system than a more direct time series approach. As an extreme case, suppose that $\sigma_1^2=0$. Then the demands are still independent, but are entirely *known one period in advance*. In this case a time series analysis of the demands would seem to support the simple models which assume independent (and unpredictable) demands when in fact the production system could rely on very accurate forecasts. Time series models that are based only on previous demands cannot capture the potential existence of very accurate forecasts that are based on more information than past demands.

We anticipate that a proponent of a time series approach to modeling forecast evolution within a simulation study could respond to the above example by pointing out that the predictions could be improved by basing the forecasts on more data than just past demands. One can imagine that human forecasters have notebooks of competitors' prices, weather forecasts, and other data that improve the accuracy of their forecasts. If these notebooks were available to the time series modeler than perhaps the time series model accuracy could rival that of the human forecaster. We agree. However, to *simulate* the behavior of such an extended time series model, we would have to simulate the behavior of these notebooks over time and we do not believe that a simulation model of notebook behavior would have greater credibility than the model we propose. Besides, if one of the notebooks contained the human forecaster's best prediction based on all available data and such predictions really did minimize mean squared error, then no other notebooks would be required: the extended time series model should select this prediction. We are still left with the problem of simulating the evolution of the human forecaster's predictions.

Correlations Reveal the Nature of New Information

In the last example, if σ_{12} is not zero, its sign tells something about the type of new information which is learned at each period. If σ_{12} is positive, then new information has a tendency to result in either an increase in both the actual demand (over that predicted) and in the prediction for the following demand, or a reduction in both of these. If, on the other hand, σ_{12} is negative, then new information tends to affect the "timing" of the demand: when the actual demand turns out to be less than predicted, there is a tendency for the prediction for the next period to rise (and vice versa). This latter situation can arise if forecasters are confident about the total demand being predicted but are unsure about the timing. For example, customers may be under contract to purchase certain amounts or the marketing department may be able to stimulate sales through subsequent promotions to achieve a total sales objective. Hence, if demands are lower than expected in one period, it is natural to forecast that the deficiency will be made up in the next period.

Application of the MMFE

As mentioned earlier, the study considered two methods of forecasting in use within the operating company. The first method, called the Traditional Method, consisted of forecasts generated by specialists within the marketing department. We analyzed four years of monthly demand and forecasts according to this method and estimated the 80×80 variance-covariance matrix for the multiplicative model of forecast evolution. The dimension 80 comes from the product of 5 products, 8 locations, and a two month forecast horizon. Using a weighted average of the variances in that matrix, we summarize in Table 1 the percentage of forecast variability that is resolved as the system evolves from two

Table 1. Percentage of Total Forecast Variability Resolved by Forecast Period, Traditional Method	
In Month of Sale	1 Month Out
64	36

periods out to 1 month out and from 1 month out to the month of sale. As is apparent from the table, a considerable fraction, 64%, of this variability is not resolved until the month of sale. This can be expected to severely limit the ability of the production system to satisfy demand without considerable safety stock.

To illustrate the demand and forecast correlations captured by the variance covariance matrix, we reproduce in Table 2 the first principal component of that matrix, scaled by the corresponding eigenvalue. Recall that in the simulation this principal component is multiplied by a standard normal random variate and that it would represent the largest sort of change to be commonly observed in the log of the forecast vector. Observe that the sign of the change in the current period tends to be the opposite of the sign of change in first period. That is, it seems that an important phenomenon of the Traditional Method is that the change in forecasts one month out are negatively correlated with observed forecast error (from one month out to the month of sale). This suggests that the forecasting specialists are confident about the total demand over a two month period so that if demand in the first month turns out to be higher or lower than expected, then the forecast for the subsequent period is adjusted downwards or upwards accordingly.

The operating company had in place a second method of forecasting demand for a portion of the total business that appeared to give very good results. This was a detailed quantitative model of customer behavior that included promotions and price considerations. Furthermore, it was able to provide forecasts for up to four months into the future. We refer to this method as the Statistical Method. In order to simulate the likely behavior of this system when extended to the entire business, the Statistical Method was simulated using two years of detailed historical data (three other years of non-overlapping data were used to calibrate the Statistical Method). From the simulated forecasts and actual demands for that two year period, we estimated the 160×160 variance covariance matrix of the logarithm of the forecast vector. The dimension of 160 is the result of the product of 5 products, 8 locations, and a forecast horizon of 4 months. The total variability of both the Traditional Method and the Statistical Method is the same. The major difference between the systems lies in when the reduction in variability occurs. Table 3 summarizes the successive reductions in variability over the course of the four months. Comparing Table 3 with Table 1 reveals that the Statistical Method of forecasting offers a great advantage over the Traditional Method. From the three month out forecast to the actual sale accounts for roughly the same fraction of variability as from the one month forecast to the actual sale in the Traditional method and the nearer term forecasts

Table 2. First Principal Component of VCV Matrix for Traditional Method

		Location							
Period	Product	1	2	3	4	5	6	7	8
0	1	−.5511	.0086	.0549	−.4902	−.1735	−.0286	.3602	−.6934
0	2	−.3526	−.3450	−.0854	−.2994	−.3982	−.2608	−.0907	−.0183
0	3	−.4483	−.1447	−.1369	−.2256	−.2584	−.1792	.0699	.0687
0	4	−.6352	−.2848	−.1855	−.3547	−.6595	−.2726	−.2890	.0758
0	5	−.4719	.5393	.2631	−.1407	−.0932	.2805	.1289	.4920
1	1	.4343	.5623	.2880	.1736	−.0027	.1454	.2605	.7386
1	2	.2981	.3231	.1257	.2454	.2418	.2621	.0654	.0285
1	3	.4571	.4149	.1714	.2560	.2768	.1838	.0835	.0387
1	4	.2555	.1171	.1395	.0787	.1844	.1178	−.0326	−.0960
1	5	.2048	.0651	.0000	−.0045	.0421	.0000	.0474	.0000

Table 3. Percentage of Total Forecast Variability Resolved by Forecast Period, Statistical Method			
In month of Sale	1 Month Out	2 Months Out	3 Months Out
7	18	30	44

are even better. This should permit more effective management of the production and distribution system. In the next section we summarize the degree to which this advantage can translate into improved customer service and lower costs.

To illustrate the demand and forecast correlations captured by the variance covariance matrix of the Statistical Method, we reproduce in Table 4 the first principal component of that matrix, scaled by the corresponding eigenvalue. The average absolute value of entries is increasing by period (0.149, 0.238, 0.296, and 0.400 for periods 0, 1, 2, and 3, respectively). Taking the view that the magnitude of entries in the principal component is indicative of the amount of information that is gained (or uncertainty resolved), this is consistent with the observations concerning Table 3. Although not as pronounced as in Table 2, Table 4 does exhibit a general trend towards sign reversals as the forecast period increases: there are 27 negative numbers in periods 0 and 1 and 58 negative numbers in periods 2 and 3. This suggests the phenomenon that lower than expected (respectively, higher than expected) sales in the current month result in downward (resp. upward) adjustments of forecasts for sales in the next month and upward (resp. downward) adjustments in forecasts for sales in the following two months.

Simulation Results

The simulation methodology, the rolling horizon LP with the MMFE, was tested with data from the operating company for a variety of safety stock factor levels. Figures 2 and 3 summarize the results of these simulations by relating the minimum average fill rate (minimum across products and locations and average over simulation runs) and annual cost reductions, respectively, to the relative safety stock factor level. The figures compare the impact of the two forecasting methods described in the previous section, the Traditional Method and the Statistical Method. Each plotted point

Table 4. First Principal Component of VCV Matrix for Statistical Method

		Location							
Period	Product	1	2	3	4	5	6	7	8
0	1	.1616	.1254	−.1043	.1974	.0373	.0416	.1342	.1776
0	2	−.0650	−.0552	.3584	−.0248	.3673	.1770	.0677	−.1579
0	3	.2146	.0507	−.0115	.3478	.0323	.1288	.2140	.2724
0	4	.0688	−.0566	.2144	.0845	.3662	.1700	.1718	−.1038
0	5	.1025	.0141	.1348	.3569	.3547	.0684	.0863	.0802
1	1	.3143	.0621	−.2724	.2532	−.3628	−.3628	.2268	.1606
1	2	−.1466	.5185	.0789	.2807	.1212	−.3917	−.3019	−.1001
1	3	.2589	.0056	−.0824	.1076	−.3237	−.1962	.2455	.0121
1	4	−.0087	.5195	.2011	.3984	.1073	−.4212	−.1456	−.1140
1	5	−.1250	−.2462	.1520	.1297	.1173	−.7305	−.1789	−.7194
2	1	−.0880	−.0336	.5566	−.1703	.5851	.5965	−.2469	−.2557
2	2	−.2385	−.1575	−.1836	.0515	−.1122	−.5729	−.2546	−.2745
2	3	−.0999	.2656	.3814	−.1636	.6767	.3992	−.1412	.0293
2	4	−.3108	−.0727	−.0952	−.0771	−.0352	−.5335	−.3089	−.5078
2	5	−.4701	−.4420	.1318	−.0971	−.1176	−.7775	−.5727	−.7525
3	1	−.2636	−.5465	−.4960	−.3411	−.4126	−.4594	−.5531	−.5784
3	2	.5144	−.3061	−.3454	−.2544	−.2015	.6555	.4665	.4505
3	3	−.2339	−.3087	−.3579	−.3369	−.2875	−.3382	−.3243	−.3118
3	4	.6342	−.1717	−.1985	−.1899	−.0624	.7127	.6639	.7549
3	5	.4911	−1.0359	−.6303	−.2990	−.2929	.0325	.2605	.2125

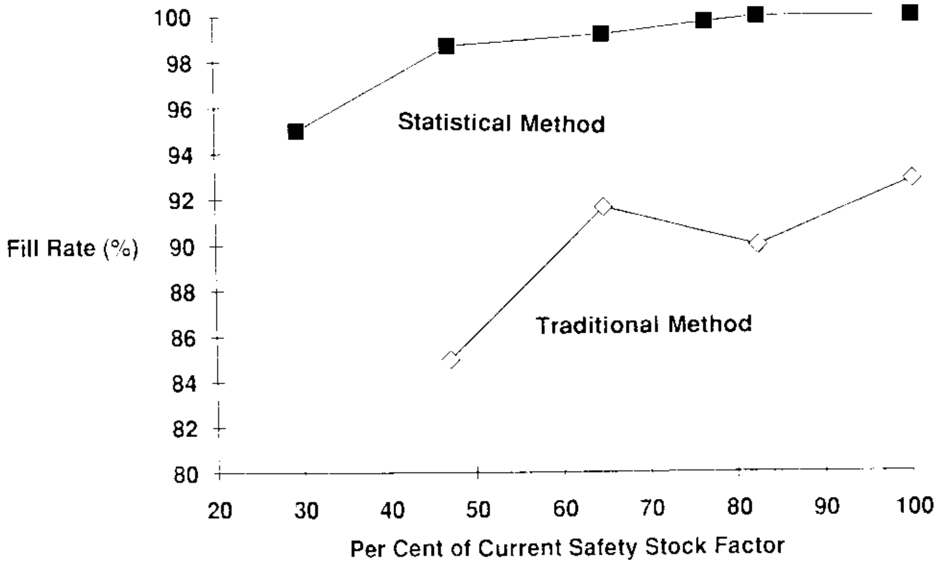

Figure 2. Impact of safety stock factor on fill rate for two forecasting methods

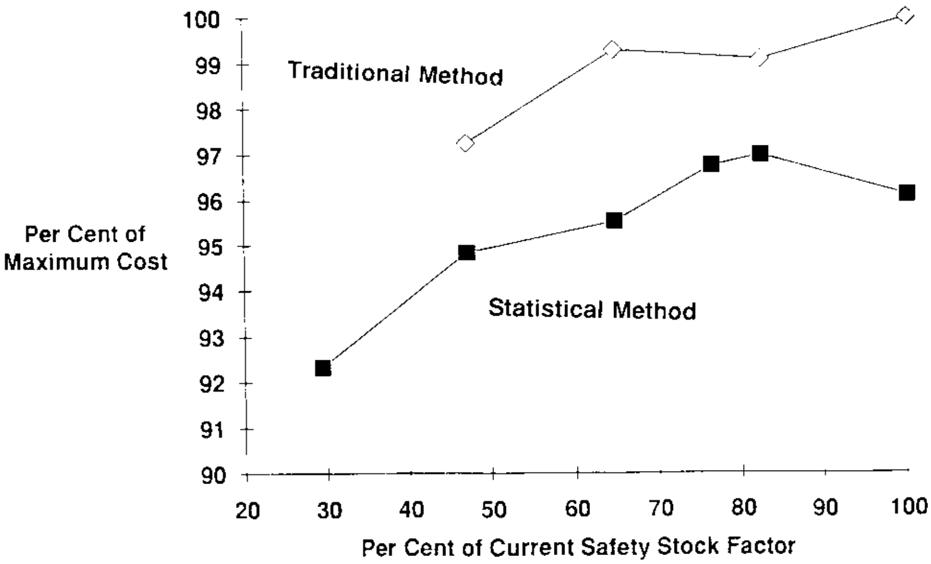

Figure 3. Impact of safety stock factor on average annual cost for two forecasting methods

represents the average across between ten and twenty simulated years; that is, the rolling horizon simulation was terminated after at most twenty consecutive years. The X_0 vector was identical for all simulation runs and consisted of the forecast for the 1990-91 fiscal year repeated 20 times. The initial random number seed was identical for all runs using the same forecasting model. Differences between minimum and maximum points for a given forecasting method in both figures are significant at the 5% confidence level. The runs are too short, however, to establish conclusively the significance of the difference, either cost or fill rate, between the different forecasting methods.

These preliminary simulations suggested that annual cost could be reduced by several million dollars annually if the safety stock factor were reduced and that, provided the new Statistical Method of forecasting was implemented and used in a timely manner to plan production, there would be little adverse impact on customer service by reducing safety stock. Other experiments were run that investigated the impact of increasing capacity (adding another production line). These runs suggested that it was far more important to increase forecasting accuracy than to increase capacity, at least for the year under consideration. These results were sufficiently intriguing to the company that management scientists within the company made extensive simulation runs using the methodology. These runs were all greater than 100 simulated years each with some runs exceeding 240 simulated years to ensure greater statistical significance. The internal studies confirmed the above conclusions and verified that the forecasting methods yielded significantly different results. We are not at liberty to reveal the details of these studies. However, as a result of these internal studies, the company made a commitment to implement the new forecasting method for the entire business and, simultaneously, to reduce substantially the safety stock factor for the 1990-91 year. After one year, the company reports that it achieved the predicted cost savings and that customer service did not suffer. Obviously, this experience of one year proves nothing, especially since the extended simulations did reveal rare years in which the simulated business experienced high costs and low customer service levels at all safety stock levels. However, the example does show that the methodology was effective in assisting management to adopt a new strategy for production management.

Conclusion

Production managers in many industries are aware that forecast error is a major factor determining production and distribution costs. In discussions with higher level management, however, they are typically unable to quantify the impact. This paper considers one approach to subjecting the impact of forecast error on cost and customer service to quantitative analysis. The Martingale Model of Forecast Evolution is proposed as a plausible model for the evolution of forecasts. We have demonstrated the practicability of this model in a large scale simulation study for an operating company. The study was effective in leading the company to implement improved forecasting techniques. Applied research in this area is now being conducted with a company that requires a more detailed production simulation model and with another company for whom component lead time is a greater concern than assembly capacity. Theoretical research is focusing on the form of the optimal inventory policy under non-trivial demand models and on state-space reduction techniques to implement the MMFE in computational dynamic programming.

Acknowledgements

The simulation study reported in this paper benefitted from the interaction with managers and management scientists within the participating operating company. We thank them for their support and professional contributions to this study. Their identity remains confidential under the terms of the study. This research was supported in part by NSF grants DMC-8509131 and DDM-8819542.

REFERENCES

[1] Azoury, K. S., "Bayes Solution to Dynamic Inventory Models Under Unkown Demand Distribution," *Management Science*, 31, No. 9, 1150-1160 (1985).

[2] Badinelli, R. D., "The Inventory Costs of Common Mis-Specification of Demand-Forecasting Models," *International Journal of Production Research*, 28, No. 12, 2321-2340 (1990).

[3] Baker, K. R., and Peterson, D. W., "An Analytic Framework for Evaluating Rolling Schedules," *Management Science*, 25, No. 4, 341-351 (1979).

[4] Bean, J., and Smith, R., "Conditions for the Existence of Planning Horizons," *Mathematics of Operations Research*, 9, 391-401 (1984).

[5] Bitran, G. B., Haas, E. A., and Matsuo, H., "Production Planning of Style Goods with High Setup Costs and Forecast Revisions," *Operations Research*, 34, No. 2, 226-236 (1986).

[6] Box, G. E. P., and Jenkins, G. M., *Time Series Analysis Forecasting and Control*, Holden-Day; San Francisco (1970).

[7] Bradford, J. W., and Sugrue, P. K., "A Bayesian Approach to the Two-period Style-goods Inventory Problem with Single Replenishment and Heterogeneous Poisson Demands," *J. Opl. Res. Soc.*, 41, No. 3, 211-218 (1990).

[8] Bratley, P., Fox, B. L., and Schrage, L. E., *A Guide to Simulation*, Springer Verlag: New York (1983).

[9] Brockwell, P. J., and Davis, R. A., *Time Series: Theory and Methods*, Springer Verlag: New York (1987).

[10] Brown, R. G., *Statistical Forecasting for Inventory Control*, McGraw-Hill: New York (1959).

[11] Dzielinski, D. P., Baker, C. T., and Manne, A. S., "Simulation Tests of Lot Size Programming," *Management Science*, 9, No. 2, 229-258 (1963).

[12] Ehrhardt, R., "The Power Approximation for Computing (s,S) Inventory Policies," *Management Science*, 25, No. 4, 777-786 (1979).

[13] Erkip, N., Hausman, W. H., and Nahmias, S., "Optimal Centralized Ordering Policies," *Management Science*, 36, No. 3, 381-392 (1990).

[14] Hausman, W. H., "Sequential Decision Problems: A Model to Exploit Existing Forecasters," *Management Science*, 16, No. 2, B93-B111 (1969).

[15] Hausman, W. H., and Peterson, R., "Multiproduct Production Scheduling for Style Goods with Limited Capacity, Forecast Revisions, and Terminal Delivery," *Management Science*, 18, No. 7, 370-383 (1972).

[16] Hoffman, K., and Kunze, R., *Linear Algebra*, Prentice-Hall, Englewood Cliffs, New Jersey (1961).

[17] Iyer, A. V., and Schrage, L. E., "Analysis of the Deterministic (s,S) Inventory Problem," to appear in *Management Science*.

[18] Jacobs, R. A., and Wagner, H. M., "Reducing Inventory System Costs by Using Robust Demand Estimators," *Management Science*, 35, No. 7, 771-787 (1989).

[19] Jacobs, R. A., and Mosier, C., "A Revision of the Power Approximation for Computing (s,S) Policies," *Management Science*, 30, 618-622 (1984).

[20] Johnson, G. D., and Thompson, H. E., "Optimality of Myopic Inventory Policies for Certain Dependent Demand Processes," *Management Science*, 21, No. 11, 1303-1307 (1975).

[21] Lee, W. B., and Khumawala, B. M., "Simulation Testing of Aggregate Production Planning Models in an Implementation Methodology," *Management Science*, 20, No. 6, 903-911 (1974).

[22] Scarf, H., "Bayes Solutions of the Statistical Inventory Problem," *Annals of Mathematical Statistics*, 30, No. 2, 490-508 (1959).

[23] Veinott, A. F., Jr., "Optimal Policy for a Multi-Product, Dynamic Non-Stationary Inventory Problem," *Management Science*, 12, No. 3, 206-222 (1965).

Appendix I: LP Formulation

SIMLP Formulation

Variable Definitions:

P_{ijt} = the production, in cases, of product i on line j in period t;

O_{jt} = the overtime on line j in period t; and

X_{ihlt} = the transshipment, in cases, of product i from DC h to DC l in period t;

I_{ilt} = the inventory, in cases, of product i at DC l at the end of period t;

B_{ilt} = the backorders, in cases, of product i at DC l at the end of period t;

S_{ilt} = the shortfall below the coverage target, in cases, of product i at DC l at the end of period t;

K = total production cost (regular time and overtime), shipping cost, and inventory holding cost.

Index Functions:

a_{jl} = 1 if DC l is supplied by line j; 0 otherwise;

f_t = 1 if $t = 1$; 0 otherwise (first period indicator);

Sets:

DC = set of DCs;

PD = set of products;

LN = set of production lines;

PL = set of production facilities (a subset of DC);

RT = set of allowed DC transshipment possibilities (pairs of origin DC—destination DC);

TM = set of time periods = $\{1,2,\ldots,\text{Horizon}\}$

Coefficient and Constant Definitions:

p_{ij} = production rate in cases/hr. of product i produced on line j;

r_{jt} = regular time hours available on line j in period t;

o_j = overtime hours available on line j in current period;

d_{ilt} = forecast demand, in cases, for product i at DC 1 in period t (actual demand for period $t=1$);

w_{il} = initial inventory less backorders, in cases, of product i at DC l in period 0;

q_{ilt} = minimum inventory requirements, in cases, for product i in DC l at end of period t;

Cost Coefficients:

c_{ij} = variable cost to produce one case of product i on line j and ship to local DC;

x_{ihl} = variable cost to ship one case of product i from DC h to DC l, including freight rates and in and out handling costs;

h_{il} = variable cost to hold stock of product i in DC l for one period, including cost of capital and storage cost;

v_j = variable cost of running line j on overtime for one hour in any period;

s_i = penalty cost for shortage of product i in period 0;

b_i = penalty cost for backorder of product i in period 0;

Objective Function:

$$\text{Min } K = \Sigma_{\{i \in PD\}}\Sigma_{\{j \in LN\}}\Sigma_{\{t \in TM\}}c_{ij}P_{ijt} + \Sigma_{\{j \in LN\}}v_jO_j + \Sigma_{\{i \in PD\}}\Sigma_{\{(h,l) \in RT\}}\Sigma_{\{t \in TM\}}x_{ihl}X_{ihlt} + \Sigma_{\{i \in PD\}}\Sigma_{\{l \in DC\}}\Sigma_{\{t \in TM\}}(h_{il}I_{ilt} + b_iB_{ilt} + s_iS_{ilt})$$

Capacity Constraints:

$$\Sigma_{\{i \in PD\}}p_{ij}^{-1}P_{ijt} - f_tO_j \leq r_{jt} \quad \text{for } j \in \text{LN},\ t \in \text{TM}; \quad (1)$$

Overtime Limits:

$$O_j \leq o_j \quad \text{for } j \in \text{LN}; \quad (2)$$

Material Balance Equations:

$$I_{ilt} - B_{ilt} = I_{il,t-1} - B_{il,t-1} + \Sigma_{\{j \in N\}}a_{jl}P_{ijt} + \Sigma_{\{(h,l) \in RT\}}X_{ihlt} - \Sigma_{\{(l,h) \in RT\}}X_{ilht} - d_{ilt} \text{ for } i \in \text{PD},\ l \in \text{DC},\ t \in \text{TM}; \quad (3a)$$

$$I_{il0} - B_{il0} = w_{il} \quad \text{for } i \in \text{PD},\ l \in \text{DC}; \quad (3b)$$

Coverage Constraints:

$$I_{ilt} - B_{ilt} + S_{ilt} \geq q_{ilt} \text{ for } i \in \text{PD},\ l \in \text{DC},\ t \in \text{TM}; \quad (4)$$

Backorder Limits:

$$B_{ilt} - B_{ilt-1} \leq d_{ilt} \text{ for } i \in \text{PD},\ l \in \text{DC},\ t \in \text{TM}; \quad (5)$$

Nonnegativity Constraints:

$$P_{ijt},\ O_j,\ X_{ihlt},\ I_{ilt},\ B_{ilt},\ S_{ilt} \geq 0$$

for $i \in$ PD, $j \in$ LN, $h,l \in$ RT, $l \in$ DC, $t \in$ TM;

and

$$I_{il0},\ B_{il0} \geq 0 \text{ for } i \in \text{PD},\ l \in \text{DC}. \quad (6)$$

Observations Concerning SIMLP

Shipments of stock of any product are allowed in any period between DCs. Shipments between particular pairs of DCs can be disallowed.

Each production line consists of only one component. Raw material availability is ignored.

Overtime is allowed on each line in the current period only. There is no planned overtime (overtime in future periods).

It is possible, since production in any period is limited, that the demand for a particular product in a particular DC in the current period cannot be satisfied, even with transhipments from other DCs. Shortages would result. Shortages in one period must be satisfied by planned production, or shortages, in the next period, or be left unsatisfied at the end of the horizon. In practice, shortages for some products are more critical than for other products. Since the LP solution is extreme, it may concentrate all the shortages in one product. We can discourage shortages in some products by making the penalty costs for shortages in those sizes to be significantly higher than those in other sizes. However, no other attempt is made in the model to balance shortages across products.

There is no penalty for underutilization of a line.

Coverage restrictions are typically stated in terms of months of supply. These are translated into bounds on inventory.

In an extreme point solution to SIMLP, we will have $I_{il0} \cdot B_{il0} = 0$ for all $i \in$ PD and $l \in$ DC, because the corresponding columns in the matrix of coefficients are linearly dependent. This is not true in general for the variables I_{ilt} and B_{ilt} because the inventory variables, I_{ilt}, do not appear in the backorder limits, constraints (5). Hence, care must be taken in analyzing the simulation history file to interpret variables and to compute costs in a reasonable manner. The set of backorder limits are necessary because otherwise there are examples in which the solution exhibits $B_{ilt} > d_{ilt}$ and $X_{ilkt} > 0$; that is, some backorders would be "shipped" to other locations.

Peter L. Jackson received his undergraduate degree in Economics with Mathematics from the University of Western Ontario; his M.Sc. in Statistics, and his Ph.D. in Operations Research from Stanford University. He is responsible for the Master of Engineering Program within the School of Operations Research and Industrial Engineering. He teaches courses in manufacturing system design, managerial accounting, economic analysis of engineering systems, manufacturing logistics, and optimization. He is the recipient of the student-voted I.I.E. award for Excellence in Teaching (1986, 1987, 1990) and the Dean's Prize for Innovation in Teaching (1986 and 1991). He has served as a consultant to the Chicago Pneumatic Tool Company, General Foods, and the Quaker Oats Company. He has been an associate editor of *Operations Research*. He conducts research in production planning and scheduling, inventory management, and graphical modeling systems for manufacturing logistics.

David Heath was awarded his Ph.D. in mathematics in 1969 and joined the School of Mathematics at the University of Minnesota; in 1975 he moved to the School of Operations Research and Industrial Engineering at Cornell. His research includes work on the foundations of statistics, the control of random processes, mathematical finance, and manufacturing systems. He has served as a consultant to DuPont, IBM, and Quaker Oats, as well as to several firms in the finance industry.