

Effective Heuristics for Capacitated Production Planning with Multiperiod Production and Demand with Forecast Band Refinement

Philip Kaminsky

Department of IEOR, University of California, Berkeley, California 94720, kaminsky@ieor.berkeley.edu

Jayashankar M. Swaminathan

Kenan-Flagler Business School, University of North Carolina, Chapel Hill, North Carolina 27599, msj@unc.edu

In this paper we extend forecast band evolution and capacitated production modelling to the multiperiod demand case. In this model, forecasts of discrete demand for any period are modelled as bands and defined by lower and upper bounds on demand, such that future forecasts lie within the current band. We develop heuristics that utilize knowledge of demand forecast evolution to make production decisions in capacitated production planning environments. In our computational study we explore the efficiency of our heuristics, and also explore the impact of seasonality on demand and availability of information updates.

Key words: forecast evolution; production planning; discrete demand; capacity; heuristics; information updates
History: Received: March 6, 2002; accepted: November 5, 2003. This paper was with the authors 12 months for 3 revisions.

1. Introduction

As increasing competition has forced firms to operate their supply chains more efficiently, the need for improved forecasts has become clear. Indeed, firms have increasingly devoted resources to improving the quality of forecasts within the supply chain (Lee et al. 1997). These resources can be used in many ways. Subjective forecasts can be improved, for example, with improved sales force composites and more-detailed customer surveys, or by soliciting the opinions of more experts. Objective forecasts can be improved, for example, by collecting more data and developing more-detailed causal models. As a result, it is becoming more important to develop models and techniques that assist managers in determining the best places to focus their forecast improvement efforts, and that enable firms to make effective use of these forecasts in production planning. Clearly, any approach should account for the following facts about forecasts:

- Forecasts are typically wrong.
- A good forecast is more than a single number.
- The longer the forecast horizon is, the more wrong the forecast, typically, is likely to be.

In addition, in a capacitated production environment, there may not be enough production capacity to wait for the forecast of a particular event to improve as that event gets closer.

Motivated by examples in the electronics industry, in Kaminsky and Swaminathan (2001) we present a model in which forecasts are represented by a series of bands encompassing both optimistic and pessimistic forecasts that are updated each period as new information arrives. In particular, this forecast evolution model is defined by an upper and a lower bound on possible demand, which we call a band. As time moves forward, subsequent forecasts have a smaller range between the upper and lower bounds, or a smaller band (representing a better forecast), and are contained within the band defined by earlier forecasts. This models a forecasting process that gets refined over time as new information arrives. The manufacturing firm utilizing this forecast has a fixed capacity in each period and needs to decide how much to produce in each period, taking into account expected production, holding, salvage, and stockout costs. In Kaminsky and Swaminathan (2001), we explore the case with a single terminal demand

and develop effective heuristics for the problem. In this paper, we consider the finite-horizon case where demand can occur in every period, develop alternative heuristics that fit this multiperiod demand model, and computationally test their effectiveness. We also consider the impact on the system of improved forecast updates, capacity availability, and seasonality in demand.

The paper is organized as follows: In §2 we present related literature. In §3 we present the model, properties, and our heuristics. In §4, we present our computational results and insights, and we conclude in §5.

2. Related Literature

Forecast evolution models have been studied by several researchers in the past; see Kaminsky and Swaminathan (2001) and Swaminathan and Tayur (2004) for a detailed review. Hausman (1969) studies the problem in which improved forecasts are available before each decision stage, and he models the evolution of forecasts as a quasi-Markovian or Markovian system. He suggests modeling a series of ratios of successive forecasts as independent lognormal variates and presents a dynamic programming formulation. Hausman and Peterson (1972) consider a single capacity-constrained multiproduct system with terminal demand, with the forecasts for total sales following a lognormal model. They formulate the problem as a dynamic program and provide heuristics for solving the problem. Heath and Jackson (1994) introduce a martingale model for forecast evolution (MMFE) and present simulation results. Gullu (1996) considers a special case of the MMFE model and shows that in a capacitated environment the system performs better when one-period demand forecast is utilized than when it is not.

As mentioned above, in Kaminsky and Swaminathan (2001) we introduce a forecast evolution process based on bands, in which the range of possible demand is completely contained within a *forecast band*. This band, defined by upper and lower bounds, captures many of the important elements of real-world forecast revisions, and in particular, the concept of improved information over time (Buckley et al. 1995). The band model restricts the future demand forecasts using a hard limit on the lower and upper bounds. In the discrete demand case, this allows one

to simplify the capacitated production planning problem and develop effective solution approaches. In this paper, we use the same forecast band model, but we generalize this to the multiperiod case, where demand can occur in every period.

More recently, a number of papers have appeared that consider the effect of advance demand information. As opposed to the above papers, work on advanced demand information assumes that future orders, or demand, is known with certainty. Gallego and Ozer (2001, 2003) study optimal replenishment policies for single and multiechelon uncapacitated inventory systems with advance demand information. Toktay and Wein (2001) consider an MMFE model and characterize effective policies under heavy-traffic assumptions for a capacitated single server. They investigate the quality of a dynamically evolving forecast, the impact of capacity utilization, and the impact of demand correlation. Their results indicate that receiving information earlier is more valuable in high-utilization scenarios, although the incremental value of this information decreases with capacity utilization. Karaesmen et al. (2002a, b) study the effect of integrating advance order information in a make-to-stock production system with capacity constraints. They model the environment as a queueing system and explore the benefits of having information about real orders for future periods. In particular, Karaesmen et al. (2002a) consider a capacitated production-inventory system where advance orders are known exactly within a fixed forecast window. They show that the production policy is a threshold policy that depends on the vector of order information. They further demonstrate that advance order information beyond a certain leadtime (calculated in the paper) has no incremental value. The value of advance information decreases in the utilization level. Karaesmen et al. (2002b) further analyze this problem and investigate the marginal value of advance demand information and the impact of capacity.

3. Basic Model

3.1. Forecast Evolution and Production Model

In this section we present the forecast evolution model. We assume that demand is discrete and the time horizon is finite, and we number time periods in

reverse order, so that when we say time t , we mean that t periods remain. Thus, period t is followed by period $t - 1$ and Period 1 is the final period. Some notation follows:

- a_t^j : the minimum possible value for demand in period j forecasted in period t ;
- w_t^j : width of the forecast for demand in period j given that we are in period t ;
- $F_t^j(a_{t+1}^j)$: probability distribution of a_t^j in the interval $[a_{t+1}^j, a_{t+1}^j + w_{t+1}^j - w_t^j]$;
- α_t^j : known reduction in width of forecast; $w_t^j = w_{t+1}^j - \alpha_t^j$;
- x_t : on-hand inventory minus back orders at the start of period t ;
- q_t : production quantity in period t ;
- C : production capacity in each period;
- π : per-unit penalty cost for demand not met;
- c : per-unit cost of production;
- h : per-unit holding cost incurred on inventory held at the end of any period;
- $s (< c)$: per-unit salvage cost for each unit remaining after demand is met;
- $I_k = (a_k^{k-1}, w_k^{k-1}, a_k^{k-2}, w_k^{k-2}, \dots, a_k^1, w_k^1)$: information available at the end of period k ;
- $V_t(x_t, I_{t+1})$: expected optimal cost incurred with t periods to go, where x_t is the inventory on hand, and I_{t+1} denotes the information available at the start of period t .

In the forecast evolution model, given that we are at the end of period t , we know that demand in period j can only take on discrete values in the interval $[a_t^j, a_t^j + w_t^j]$. Note that a_t^j refers to the lower bound of demand in period j forecasted in period t . In addition, each period the range of possible demand gets smaller in a way consistent with the process described above. That is, $w_t^j = w_{t+1}^j - \alpha_t^j$, where each $\alpha_t^j \geq 0$, $\forall t$ is such that $w_t^j \geq 0$, $\forall t > j$. The final demand a_j^j is defined in the discrete interval $[a_{j+1}^j, a_{j+1}^j + w_{j+1}^j]$ with a probability distribution $F_j^j(a_{j+1}^j)$. Furthermore, in each period, the lower bound on the forecast demand in subsequent periods is updated probabilistically. In particular, a_{t-1}^j lies in the range $[a_t^j, a_t^j + w_t^j - w_{t-1}^j]$ with the probability distribution $F_{t-1}^j(a_t^j)$. This forecast evolution process captures the above two assumptions: (1) The forecast gets tighter as one gets closer to demand since w_t^j is increasing in $t - j$, and (2) the future forecast is contained in the current range since

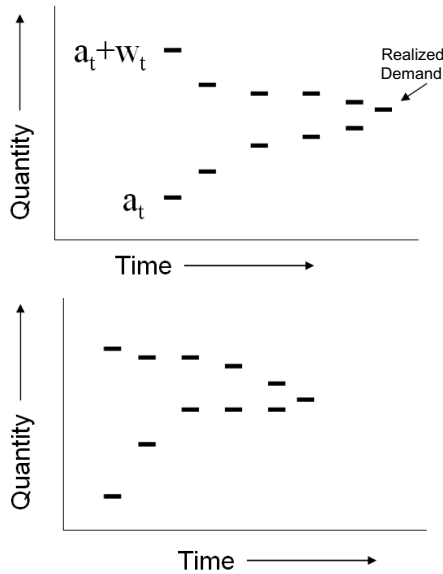
a_{t-1}^j lies in the range $[a_t^j, a_t^j + w_t^j - w_{t-1}^j]$. Note that $F_{t-1}^j(a_t^j)$ allows us to model changes in distribution over time, and α_t^j provides the ability to model the different points of time at which major or minor updates in forecast may be obtained. A firm that obtains all the forecast updates very close to the selling season will have $\alpha_t^j = 0$ for values of t farther away from j , and a high value for α_t^j for values of t very close to j . Note that we assume that the progression of w_t^j is known, whereas the progression of a_t^j is probabilistic. This models a situation in which the firm has a good understanding of when additional information will be obtained (for example, a trade show), but doesn't know a priori what that information will be; however, the firm has a model for the evolution of this information in the form of F_t .

Figure 1 illustrates two possibilities for the evolution of forecasts for demand in a single period. Each parallel pair of horizontal lines represents a lower and an upper bound on possible demand. As time advances, these bounds become tighter and are contained within the previous set of bounds. The final pair of horizontal lines in each of the pictures represents the upper and lower bounds on the distribution of demand that is realized in the next period. Although these pictures represent evolving demand for a single period, recall that in our model, in each period, demand forecasts are evolving for a series of periods simultaneously.

We assume that demand occurs every period, and that the manufacturer can produce between 0 and C units each period. We consider a T -period finite-horizon model, in which demand can occur in each of the periods. Periods are numbered backwards; T is the first period, and the last period is 1. We assume that the leadtime for production is zero, and there are no setup costs for production. The objective is to minimize the expected total cost during the finite horizon; at each period a production decision needs to be made. More specifically, the order of events in each period t is as follows:

- (1) the production decision q_t is made;
- (2) demand a_t^i is realized from the range $[a_{t+1}^i, a_{t+1}^i + w_{t+1}^i]$ with probability distribution $F_t^i(a_{t+1}^i)$;
- (3) the forecasts for each subsequent period i , a_t^i are realized from the range $[a_{t+1}^i, a_{t+1}^i + w_{t+1}^i - w_t^i]$ with probability distribution $F_{t-1}^i(a_t^i)$;

Figure 1 Sample Forecast Evolution



(4) holding and penalty costs are charged.

Let $V_t(x, I_{t+1})$ denote the optimal cost-to-go function with t periods remaining in the horizon when the on-hand inventory is x and the information available is I_{t+1} . We assume that all unmet demand is back ordered. Then,

$$V_t(x, I_{t+1}) = \min_{0 \leq q_t \leq C} \mathbf{E} \left\{ q_t c + h(x + q_t - \xi_t)^+ + \pi(\xi_t - x - q_t)^+ + V_{t-1}(x + q_t - \xi_t, I_{t-1}) \right\}$$

and

$$V_1(x, I_2) = \min_{0 \leq q_1 \leq C} \mathbf{E} \left\{ q_1 c + (h + s)(x + q_1 - \xi_1)^+ + \pi(\xi_1 - x - q_1)^+ \right\}.$$

The expectation \mathbf{E} is with respect to all possible realization of demand in the current period, and all possible values of I_t . It is clear that this problem is quite challenging. It is a nonstationary inventory problem with capacity restrictions and forecast evolution. A threshold policy is optimal if the optimal cost-to-go function is convex, which in the discrete case reduces to showing increasing first differences. We will now show that the one-period and t -period objectives have increasing first differences.

THEOREM 3.1. $V_t(x + 2, I_{t+1}) - V_t(x + 1, I_{t+1}) \geq V_t(x + 1, I_{t+1}) - V_t(x, I_{t+1}) \forall x, I_{t+1}, t = 1, \dots, T$.

See the Appendix for proof.

This allows us to conclude:

COROLLARY 3.2. *For every period t , there exists a threshold level $y_t^*(I_{t+1})$ such that it is optimal to produce in period t if $x_t \leq y_t^*(I_{t+1})$ and to not produce if $x_t > y_t^*(I_{t+1})$.*

As the state space for this problem is very large, optimally solving the dynamic programming recursion for this model is very complex and time intensive. We are thus motivated to consider a variety of heuristics for this problem in order to handle realistic-size problems.

3.2. Heuristics

We consider three heuristics for this model. In particular, the first two heuristics involve reduction of the state space by simplification of the model in various ways. The final heuristic, motivated by the heuristics in Kaminsky and Swaminathan (2001), as well as approaches first proposed by Morton and Pentico (1995), solves a series of related single-period uncapacitated problems and then uses this information to make production decisions.

H1: SOLVING THE PROBLEM IGNORING FORECAST EVOLUTION FOR SOME SMALL NUMBER OF PERIODS IN THE FUTURE. *We assume that the forecast will not evolve, and solve the P -period finite-horizon problem optimally given the current state of information I_{t+1} , where $P \ll T$. In particular, we assume that demand for period j , $j = 1, 2, \dots, t$ will be realized in the range $[a_{t+1}^j, a_{t+1}^j + w_{t+1}^j]$ and has the same distribution as the conditional distribution of demand in period j given the forecast information in period t . Note that a myopic approach is a special case of this heuristic where only one period is considered, and that salvage cost is added only if the final period is being considered.*

H2: SOLVING THE PROBLEM FOR LOWER AND UPPER BOUNDS FOR A LIMITED TIME HORIZON. *In this heuristic, we assume that the demands in future periods are deterministic, fixed at the lower bounds or upper bounds, and find the optimal production for this (deterministic) system for P' periods, where $P' < T$. Note that because this deterministic problem is easier to solve than the stochastic problem, the H2 horizon P' can be longer than the H1 horizon P . In Heuristic H2l, if we are at time t , we fix demand at time $l \leq t$ at a_l^l for all $l \leq t$, while Heuristic H2u) fixes demand at $a_l^l + w_l^l$ for all $l \leq t$. As in Heuristic H1, salvage cost is added only if the final period is being considered.*

H3: SOLVING THE NONSTATIONARY PROBLEM WITHOUT CAPACITY CONSTRAINTS AND EXTRAPOLATING. In Kaminsky and Swaminathan (2001), we solve the terminal demand version of this problem by first solving an uncapacitated single-period problem and then adapting that solution to the multiperiod capacitated production model. Conceptually, we do this “spreading the manufacturing backwards” so that the desired inventory can be manufactured in the presence of capacity constraints. In this Heuristic H3, we use a similar approach.

Of course, one difficulty in doing this is that the uncapacitated problem is challenging to solve. Fortunately, Morton and Pentico (1995) develop a heuristic which they computationally demonstrate produces very good solutions to the uncapacitated multiperiod nonstationary stochastic inventory problem. Their heuristic works as follows: Each period, the order quantity is taken to be the minimum of two order quantities. The first is the single-stage myopic order level (the newsvendor level) arrived at by ignoring future costs and only considering one-period holding cost and penalty cost, and the second is determined by solving this multiperiod problem as if it is a single-period problem with demand equal to the cumulative remaining demand and adjusted holding costs.

In our Heuristic H3, we combine this idea of “spreading back” production from Kaminsky and Swaminathan (2001) with a modified version of the Morton and Pentico (1995) heuristic.

First, we present the Morton and Pentico heuristic to determine the uncapacitated production quantity. Recall that this heuristic is based on solving a modified single-period problem. Now, let Φ_{t1} be the cumulative distribution of demand over the periods from t to 1 inclusive, assuming that the distribution in each subsequent period is conditional only on I_{t+1} (that is, without forecast evolution, as in Heuristic H1). Then the optimal production quantity for a single-period model with modified holding costs, on-hand inventory x , the cumulative demand, and maximum production equal to the total available capacity, is calculated. Specifically, the production quantity is calculated in two ways:

$$\bar{q}_{t1}^m = \arg \min_{0 \leq q \leq TC} E_{d_{t1}} \{cq + (nh - s)(q + x - d_{t1})^+ + (\pi)(d_{t1} - q - x)^-\},$$

where d_{t1} is distributed according to Φ_{t1} . Also, the one-period myopic production quantity is calculated:

$$\bar{q}_t^m = \arg \min_{0 \leq q \leq TC} E_{d_t} \{cq + (h)(q + x - d_t)^+ + (\pi)(d_t - q - x)^-\} \quad t > 1$$

or, in the final period:

$$\bar{q}_1^m = \arg \min_{0 \leq q \leq TC} E_{d_1} \{cq + (h - s)(q + x - d_1)^+ + (\pi)(d_1 - q - x)^-\},$$

where d_t is distributed according to Φ_t .

Finally, Morton and Pentico (in their EOS heuristic) take the uncapacitated production quantity to be:

$$\bar{q}_t = \min\{\bar{q}_{t1}^m, \bar{q}_t^m\}.$$

Of course, in our problem, production capacity plays a role. Indeed, the production schedule suggested by the Morton and Pentico heuristic will typically be infeasible for the capacitated problem. We thus need to move some of the production earlier where necessary to account for these capacity constraints. We detail this approach, Heuristic H3, below.

Suppose we are in period t . We implement the following procedure to determine the order-up-to level in this period. First, we calculate production quantity \bar{q}_i , $i = t, t - 1, \dots, 1$ as described above for the uncapacitated case, using the appropriate demand distributions and convolutions, assuming that demand does not continue to evolve, but is instead distributed in future periods based on the current bands. Note that we know the initial inventory in period t , and assume an initial inventory equal to zero in all other periods. Next, we adjust these production levels to account for the capacity constraint using the following procedure:

(1) Calculate \bar{q}_i , $i = t, t - 1, \dots, 1$ as described above.

(2) New levels, denoted q_j , are determined recursively (from 1 to t) using the following formulation:

$$q_j = \min \left\{ C, \bar{q}_j + \sum_{i=1}^{j-1} (\bar{q}_i - q_i) \right\}, \quad j = 1, 2, \dots, t - 1$$

$$q_t = \min \left\{ C, \bar{y}_j + \sum_{i=1}^{j-1} (\bar{q}_i - q_i) - x_t \right\}.$$

Note that although this approach calculates a series of q_i values, these have to be recalculated for each t because the distribution and starting inventory change as information is updated.

4. Computational Study

We present the results of computational studies, in which we employ simulation to test our heuristics under different settings, and explore the benefits of improving forecasts earlier in the time horizon. We use the following parameters for our computational study:

Time Horizon. We are interested for the most part in relatively short finite time horizons. To test the effectiveness of our heuristics, we compare heuristic solutions to the optimal solution. For these tests we use a four-period time horizon, the largest time horizon for which our problem can practically be solved to optimality. For additional tests in which we use only our heuristics, we utilize a 24-period time horizon. In addition, Heuristics H1 and H2 can effectively be used for longer time horizons, so we test these two heuristics over 64-period time horizons.

Cost Parameters. We use production cost $c = 50$; penalty $\pi = 75, 150, 250$; salvage $s = 0, 5, 10, 15, 25, 45$; and holding cost $h = 0, 2, 4, 8, 12, 16, 24$ in the study to determine the effect of relatively high and low penalty, salvage, and holding costs.

Demand. We assume that demand in period t is uniformly distributed in the interval $[a_t^i, w_t^i]$. We set initial demand and forecast updates so that final demand has a width of five. We set the initial width equal to a quantity no larger than 11, and less if there is not enough time to reduce the width to five. Thus, for four-period time horizons, we set the *initial* w values as follows:

$$w_5^4 = 5, \quad w_5^3 = 6, \quad w_5^2 = 7, \quad w_5^1 = 8.$$

Similarly, for 24-period horizons, we set *initial* w values as follows:

$$w_{25}^{24} = 5, \quad w_{25}^{23} = 6, \quad w_{25}^{22} = 7, \quad w_{20}^{21} = 8, \\ w_{25}^{20} = 9, \quad w_{25}^{19} = 10,$$

and for the remaining values

$$w_{25}^i = 11, \quad i = 1, 2, \dots, 18.$$

To reduce the bands to the appropriate value in time, we set α s as follows. For the four-period horizon:

$$\alpha_l^t = 1, \quad t = 1, 2, 3, \quad l = t + 1, \dots, 4$$

with remaining α s set to 0.

For the 24-period horizon:

$$\alpha_l^t = 1, \quad t = 1, 2, \dots, 18, \quad l = t + 1, t + 2, \dots, t + 6 \\ \alpha_l^t = 1, \quad t = 19, 20, \dots, 23, \quad l = t + 1, \dots, t + (24 - t)$$

with remaining α s set to 0.

All that remains is to set the initial a^t values. Recall that the a^t values represent a lower bound on demand, and thus can be set to simulate stationary demand as well as a variety of seasonal demand variation. For the four-period time horizon, we consider the following three sets of initial conditions:

- the *base* case: $a_5^t = 4; t = 1, 2, \dots, 24$.
- the *moderately seasonal* case: $a_5^t = 5; t = 1, 3$; and $a_5^t = 3; t = 2, 4$.
- the *seasonal* case: $a_5^t = 8; t = 1, 3$; and $a_5^t = 0; t = 2, 4$.

In addition, to control for starting and ending effects, we also consider offset versions of the moderately seasonal and seasonal cases, in which high and low periods are swapped:

- the *offset moderately seasonal* case: $a_5^t = 5; t = 2, 4$; and $a_5^t = 3; t = 1, 3$.
- the *offset seasonal* case: $a_5^t = 8; t = 2, 4$; and $a_5^t = 0; t = 1, 3$.

Similarly, for the 24-period time horizon, we consider the following five sets of initial conditions:

- the *base* case: $a_{25}^t = 4; t = 1, 2, \dots, 24$.
- the *moderately seasonal* case: $a_{25}^t = 5; t = 1, 2, 3, 4, 9, 10, 11, 12, 17, 18, 19, 20$; and $a_{25}^t = 3; t = 5, 6, 7, 8, 13, 14, 15, 16, 21, 22, 23, 24$.
- the *offset moderately seasonal* case: $a_{25}^t = 3; t = 1, 2, 3, 4, 9, 10, 11, 12, 17, 18, 19, 20$; and $a_{25}^t = 5; t = 5, 6, 7, 8, 13, 14, 15, 16, 21, 22, 23, 24$.
- the *seasonal* case: $a_{25}^t = 8; t = 1, 2, 3, 4, 9, 10, 11, 12, 17, 18, 19, 20$; and $a_{25}^t = 0; t = 5, 6, 7, 8, 13, 14, 15, 16, 21, 22, 23, 24$.
- the *offset seasonal* case: $a_{25}^t = 0; t = 1, 2, 3, 4, 9, 10, 11, 12, 17, 18, 19, 20$; and $a_{25}^t = 8; t = 5, 6, 7, 8, 13, 14, 15, 16, 21, 22, 23, 24$.

Each period, the lower bound a^t evolves as described in §3.1. In particular, a_{t-1}^i lies in the range $[a_t^i, a_t^i + w_t^i - w_{t-1}^i]$. Furthermore, for the purpose of this computation study, we assume that a_{t-1}^i can take each of the integer values in this range with equal probability.

For the 64-period problem, we set initial w values, α values, and initial a^t values analogously. In particular, seasonal and offset seasonal a^t values start and cycle in the same way.

Capacity. We examine three capacity cases, *high*: $C = 13$, *medium*: $C = 9$, *low*: $C = 7$. To understand these capacity values, compare them to the $a_i + w_i$ values for the different scenarios described above.

4.1. Performance of the Heuristics

We tested the effectiveness of the heuristics across the four-period and 24-period time horizons. For the four-period horizon, we compared heuristics and the optimal solution for all combinations of the parameters listed above. Each parameter combination was tested through 10 simulation replications. For a given replication, the same set of randomly generated demand and forecast update data was used in each of the heuristics to facilitate comparison. Note that the optimal solution involves minimizing the total *expected cost to go* at each stage, so for a given simulation realization, it is possible for one of our heuristics (or some other solution) to dominate what we are calling the optimal solution.

In the simulation runs, at each period the ordering decision was made based on the algorithm, and then random demand and forecast updates were realized. In addition, we compared the four heuristics (without the optimal solution) for all of the combinations of the parameters listed above for a 24-period time horizon, using the same simulation-based approach. In all cases, the horizon for Heuristic H1 was three periods. For the four-period problems, Heuristics H2l, H2u, and H3 were solved for all periods. For the 24-period problems, Heuristics H2l and H2u were calculated using a six-period horizon.

We find that for the four-period problems (2,520 in all), H1 shows the best performance, with average objective values of 0.009% over the optimal. H2u follows with an average objective value of 0.24% over the optimal. H3 produces results that are 15.19% away from the optimal. H2l performs rather poorly, averaging 62.09% worse than the optimal solution (see Table 1). It is not surprising that for a four-period problem solving the three-period horizon optimally, (H1) is likely to be very close to the overall optimal solution. On the other hand, by producing to meet the

Table 1 Average Error and Running Times for H1, H2l, H2u, and H3 for the Four-Period Problems

	H1	H2l	H2u	H3
Average error (%)	0.009	62.09	0.24	15.19
Average time	0.06	0.002	0.002	0.002
Percent best (%)	64	0	26	10

lower bound in the demand, H2l can be quite ineffective because the penalty cost is much higher than the holding cost. Using a similar reasoning, H2u performs quite well because it leans towards producing more. However, each of the three dominant heuristics did perform best at least some of the time (see the last line of Table 1). In general, H2u performs best (and beats only H1) when penalty costs are relatively low. Similarly, H3 performs best when penalty costs are relatively high. We also find that H2l, H2u, and H3 run much faster than H1 (see Table 1, where times are seconds on a Pentium III-800 computer).

For the problem with 24 periods, computing the optimal strategy was unrealistic in terms of time and memory constraints. As a result, we compare the average cost incurred by the heuristics across the 7,560 problem instances that we tested. We find that H2u and H1 perform significantly better than H3 and H2l. All the heuristics run within a reasonable amount of time, even for these longer-horizon problems (see Table 2). This suggests that Heuristics H2u and H1 are efficient and effective even when the number of periods is high. When we compare these two heuristics, we see that their performance is quite similar, with H2u having a slightly better average performance in the 24-period case. As in the four-period case, three of the four heuristics are best some of the time, as reflected in the last line of Table 2. Also as before, H1 performs best when penalty costs are relatively high, particularly when holding costs are relatively low. Similarly, H3 performs best when penalty costs are relatively high.

Table 2 Average Simulated Costs and Running Times for H1, H2l, H2u, and H3 for the 24-Period Problems

	H1	H2l	H2u	H3
Average costs	40,162	51,854	36,617	50,512
Average time	1.46	0.12	0.12	1.26
Percent best (%)	42	0	50	8

4.2. Managerial Insights

In this section we describe additional managerial insights that we obtained through an extensive computational study aimed at understanding the effect of forecast updates, seasonal fluctuation, and firm capacity. For the four-period trials, we use costs resulting from solving the dynamic program to optimality. For the 24-period trials, we apply each of the heuristics, and use the best cost. As we observed above, three of the four heuristics work best at least some of the time. By using the “best heuristic” for each trial, we attempt to minimize the impact of a particular heuristic on the insights listed below. For the 64-period trials, we use two replications rather than 10, and to control for initialization bias, we let the system “warm up” for 16 periods, or two complete cycles, before we begin to collect the 64 periods of data. We do not apply Heuristic H3 for this problem, as the convolutions are too time consuming to calculate. Note that although these heuristics can be used for longer time horizons than we can solve optimally, we did not intend or design them for an extremely long-term or infinite-horizon setting.

We summarize our insights below. We have selected specific examples from our data to illustrate our points, but each of these points can be observed throughout our data set.

4.2.1. Effect of Cost and Capacity. Not surprisingly, total costs increase with increase in penalty and holding costs, and costs decrease with increase in capacity. Further, we find that the cost decreases in a convex fashion with increase in capacity (see Table 3). This implies that moving from a highly capacity-constrained situation to a medium-level capacity is much more beneficial than having a similar increase in capacity to move from medium level to high level of capacity.

Table 3 Effect of Capacity and Information Updates on the Average Simulated Costs when $h = 15$, $p = 150$, $c = 50$, and $s = 24$ for the 24-Period Problem Under Base Seasonality

	Capacity = 7	Capacity = 9	Capacity = 13
Early update	88,085	17,791	12,319
Intermediate update	90,200	18,677	12,335
Late update	94,205	20,543	12,427

4.2.2. Effect of Information Updates. We modeled the information updates in the forecast process by changing the width of the forecast band from period to period. We consider *early-information*, *intermediate-information*, and *late-information* models. In the *early* (*intermediate*, *late*)-*information* case the band width reduces more per period in the first few (middle, last few) periods, respectively, although in all cases the bands decrease from their initial width to a width of five. For example, for *early information*, all α s except for the final six before the demand period are zero. The final six α s are represented by the vector (3 2 1 0 0 0), so that $a_{t+6}^t = 3$, $a_{t+5}^t = 3$, $a_{t+4}^t = 2$, etc. In other words, the accuracy of the forecast (as measured by the width of the forecast band) does not improve until six periods before final demand. At that point, the forecast improves rapidly at first and does not improve later than four periods before the demand is realized. Similarly, the *intermediate* and *late-information* cases are represented by (1 1 1 1 1 1) and (0 0 0 1 2 3), respectively. In the *intermediate-information* case, forecasts do not improve until six periods before demand, and then either improve monotonically. In the *late-information* case, forecasts do not improve until three periods before demand is realized, and even then they only improve one period at the most before demand is realized. Presumably the firm has to invest in better forecasting systems to obtain the refined updates earlier in the horizon. We do not consider such costs in our analysis, but focus only on the benefits.

We observe that obtaining information updates earlier is almost always beneficial (see Table 3). When the capacity is high, the marginal benefits of earlier information updates are minimal. When the capacity is medium, early and intermediate updates are both useful, although an intermediate update can capture most of the benefits. When the capacity is low, both an early and an intermediate update change the costs in a significant manner. We see similar results for other instances and heuristics. Intuitively, obtaining earlier updates enables the firm to plan production more effectively. However, this is crucial only when the firm has limited capacity. When the firm has a large capacity base, it has the ability to adequately respond to abrupt changes in the demand process, and hence the value of earlier information is minimal.

Our insight is similar in spirit to the one presented by Toktay and Wein (2001), where they indicate that receiving information earlier is more valuable in high-utilization scenarios, although the incremental value of this information decreases with capacity utilization under heavy-traffic approximations. Similarly, Karaesmen et al. (2002a, b) show that for a capacitated production-inventory system where advance orders are known *exactly* within a fixed forecast window, advance order information beyond a certain leadtime has no incremental value. The above insights (ours and those of other researchers) indicate that firms need to take into account their capacity conditions before deciding on the type of information updates to obtain for their forecasts. Significant investment is typically required to obtain information updates, and earlier updates may cost more and may not be as beneficial, depending on the capacity conditions.

4.2.3. Effect of Seasonal Demand. As described earlier, we created five types of seasonal demand—*base*, *intermediate seasonal*, *offset intermediate seasonal*, *high seasonal*, and *offset high seasonal*. The effect of seasonality (or alternating high and low demands) on the costs is interesting. There are two factors that influence the outcome—first is the level of capacity, and the second is if the initial periods are a high-demand season or a low-demand season. We find that, as expected, seasonality increases costs (for all capacities) when the first period is a high season (offset intermediate seasonal and offset high seasonal cases). However, when the first period is a low season (intermediate seasonal and high seasonal), then seasonality could in fact be beneficial when capacity is low (see Capacity = 7 in Table 4). When capacity is medium (Capacity = 9 in Table 4), the costs first decrease and then increase with seasonality. Having seasonality and a low season in the first period

Table 4 Effect of Capacity and Seasonality on the Average Simulated Costs when $h = 10, p = 75, c = 50$, and $s = 12$ for the Four-Period Problem with Intermediate Information Update

	Capacity = 7	Capacity = 9	Capacity = 13
Base seasonal	1,914	1,557	1,539
Intermediate seasonal	1,831	1,544	1,539
High seasonal	1,691	1,574	1,559
Offset intermediate seasonal	1,995	1,600	1,542
Offset high seasonal	2,375	1,794	1,580

Table 5 Effect of Capacity and Seasonality on the Average Simulated Costs when $h = 10, p = 75, c = 50$, and $s = 12$ for the 24-Period Problem with Intermediate Information Update

	Capacity = 7	Capacity = 9	Capacity = 13
Base seasonal	48,242	14,448	11,592
Intermediate seasonal	45,434	14,209	11,603
High seasonal	40,736	15,496	12,109
Offset intermediate seasonal	51,407	15,184	11,594
Offset high seasonal	62,025	21,697	11,940

provides an opportunity to produce in anticipation for the high demand in the next period, thereby avoiding high stockout costs. This is particularly useful when capacity is low. We see similar results for other instances. Table 5 shows the insight for the best heuristic solution for the same problem with 24 periods in the horizon. These results stress the importance in seasonal environments of planning for high-demand seasons ahead of time. As expected, over a longer horizon, and when a warm-up period is used in the simulation, the impact of the first season is diminished. In Table 6, for example, we see that over a longer horizon the observed patterns remain the same, but the difference in overall objective between when the first period is high demand or low demand is reduced.

5. Conclusions

We present a model that captures forecast evolution information while making production decisions in a capacitated environment. We consider a forecast evolution model that is defined by a band (with upper and lower bounds) that captures the uncertainty in the forecast. As time moves forward, we assume that the next forecast has a smaller width (representing a better forecast) and that the new band can lie anywhere inside the old band. This models a forecasting

Table 6 Effect of Capacity and Seasonality on the Average Simulated Costs when $h = 8, p = 75, c = 50$, and $s = 10$ for the 64-Period Problem with Intermediate Information Update

	Capacity = 7	Capacity = 9	Capacity = 13
Base seasonal	549,988	94,183	31,454
Intermediate seasonal	540,388	85,289	31,471
High seasonal	521,188	95,627	32,394
Offset intermediate seasonal	559,588	103,452	31,513
Offset high seasonal	588,388	129,520	32,803

process that gets refined over time as new information arrives. We consider a manufacturing firm utilizing this forecast that has a fixed capacity in each period and needs to decide in which periods it should produce in order to meet demand in each period, taking into account expected production, holding, salvage, and stockout costs. We prove the existence of inventory threshold levels in each period below which the firm should produce and above which it should not. We develop several heuristics, based on either reducing the size of the dynamic programming state space, or on solving a series of uncapacitated problems and then modifying the solutions for the capacitated case. Our computational study shows that several of these heuristics are very effective.

We empirically observe that early information updates in the forecasting process generally lead to decreased costs, although this is most dramatic when capacity is tight. We also observe that in this model, in general increases in seasonality lead to increases in total costs. However, when the initial season is a low-demand season, then seasonality could in fact be beneficial when capacity is low. This is particularly true for relatively short-lifecycle products. Finally, we find that the marginal benefit of earlier information is lower when capacity is high.

There are certain limitations in the model we present in this paper. First, we assume that forecasts get better, or at least are as good, in future periods. Although this assumption is reasonable under most conditions, there may be situations where this condition may be violated. Second, we present a model where only one product is being produced in a capacitated setting. A multiproduct system would be more appropriate in a general setting; it will be more complex to handle, however.

Acknowledgments

The authors would like to thank the Editor and referees for their valuable feedback and comments on earlier drafts of this paper. The research of the first author was supported in part by NSF Career Award DMI-0092854 and NSF Grant DMI-0200439, and the research of the second author was supported in part by NSF Career Award DMI-9984252.

Appendix

LEMMA 1. Let $L(y, I) = cy + \mathbf{E}_\xi\{(h+s)(y-\xi)^+ + (\pi)(\xi-y)^+\}$. Then $L(y+2, I) - L(y+1, I) \geq L(y+1, I) - L(y, I)$.

PROOF. Let $\hat{L}(y, \xi) = cy + \{(h+s)(y-\xi)^+ + (\pi)(\xi-y)^+\}$. When $\xi \geq y+2$, then $\hat{L}(y+2, \xi) - \hat{L}(y+1, \xi) = \hat{L}(y+1, \xi) - \hat{L}(y, \xi) = c - \pi$. When $\xi \leq y$, then $\hat{L}(y+2, \xi) - \hat{L}(y+1, \xi) = \hat{L}(y+1, \xi) - \hat{L}(y, \xi) = c + (h+s)$. When $\xi = y+1$, then $\hat{L}(y+2, \xi) - \hat{L}(y+1, \xi) = c + h + s$ and $\hat{L}(y+1, \xi) - \hat{L}(y, \xi) = c - \pi$. Therefore,

$$(\hat{L}(y+2, \xi) - \hat{L}(y+1, \xi)) - (\hat{L}(y+1, \xi) - \hat{L}(y, \xi)) \geq 0 \quad \forall \xi.$$

Thus, $L(y+2, I) - L(y+1, I) \geq L(y+1, I) - L(y, I)$. \square

PROOF OF THEOREM 3.1. We will show this by induction. First we will show that this is true for $t=1$.

$$\begin{aligned} V_1(x, I_2) &= \min_{0 \leq q_t \leq C} \mathbf{E}\{q_t c + (h+s)(x+q_t-\xi_t)^+ + \pi(\xi_t-x-q_t)^+\} \\ &= -cx + \min_{x \leq y \leq x+C} \mathbf{E}\{cy + (h+s)(y-\xi_t)^+ + \pi(\xi_t-y)^+\} \\ &= -cx + \min_{x \leq y \leq x+C} L(y, I_2). \end{aligned}$$

Let $g = \arg \min L(y, I_2)$. There are five cases for possible values of g .

Case 1. $g \leq x$. In this case, $V_1(x+2, I_2) - V_1(x+1, I_2) = -c + L(x+2, I_2) - L(x+1, I_2)$ and $V_1(x+1, I_2) - V_1(x, I_2) = -c + L(x+1, I_2) - L(x, I_2)$. From Lemma 1, we know that $V_1(x+2, I_2) - V_1(x+1, I_2) \geq V_1(x+1, I_2) - V_1(x, I_2)$.

Case 2. $g = x+1$. In this case, $V_1(x+2, I_2) - V_1(x+1, I_2) = -c + L(x+2, I_2) - L(x+1, I_2)$ and $V_1(x+1, I_2) - V_1(x, I_2) = -c$. Since $g = x+1$ is the minimal point and L has increasing first differences, we know that $L(x+2, I_2) \geq L(x+1, I_2)$; therefore, $V_1(x+2, I_2) - V_1(x+1, I_2) \geq V_1(x+1, I_2) - V_1(x, I_2)$.

Case 3. $g = x+C+1$. In this case, $V_1(x+2, I_2) - V_1(x+1, I_2) = -c$ and $V_1(x+1, I_2) - V_1(x, I_2) = -c + L(x+C+1, I_2) - L(x+C, I_2)$. Since $g = x+C+1$ is the minimal point and L has increasing first differences, we know that $L(x+C, I_2) \geq L(x+C+1, I_2)$; therefore, $V_1(x+2, I_2) - V_1(x+1, I_2) \geq V_1(x+1, I_2) - V_1(x, I_2)$.

Case 4. $g \geq x+C+2$. In this case, $V_1(x+2, I_2) - V_1(x+1, I_2) = -c + L(x+C+2, I_2) - L(x+C+1, I_2)$ and $V_1(x+1, I_2) - V_1(x, I_2) = -c + L(x+C+1, I_2) - L(x+C, I_2)$. From Lemma 1, therefore, $V_1(x+2, I_2) - V_1(x+1, I_2) \geq V_1(x+1, I_2) - V_1(x, I_2)$.

Case 5. $x+2 \leq g \leq x+C$. In this case, $V_1(x+2, I_2) - V_1(x+1, I_2) = -c$ and $V_1(x+1, I_2) - V_1(x, I_2) = -c$; therefore, $V_1(x+2, I_2) - V_1(x+1, I_2) \geq V_1(x+1, I_2) - V_1(x, I_2)$. Thus, $V_1(x+2, I_2) - V_1(x+1, I_2) \geq V_1(x+1, I_2) - V_1(x, I_2) \quad \forall x, I_2$.

We have shown that with one period, objective function $V_1(x, I_1)$ has increasing differences, and therefore a threshold policy is optimal. Next, we show that this is true of any $V_t(x, I_t)$. Let us assume that it is true for $t-1$, and we will prove it for t . Let $G(y, I) = cy + \mathbf{E}_\xi\{(h)(y-\xi)^+ + (\pi)(\xi-y)^+\}$.

$$\begin{aligned} V_t(x, I_{t+1}) &= \min_{0 \leq q_t \leq C} \mathbf{E}\{q_t c + h(x+q_t-\xi_t)^+ + \pi(\xi_t-x-q_t)^+ \\ &\quad + V_{t-1}(x+q_t-\xi_t, I_t)\} \\ &= -cx + \min_{x \leq y \leq x+C} \{G(y, I_{t+1}) + \mathbf{E}V_t(y-\xi_t, I_t)\}. \end{aligned}$$

Note that $G(y, I)$ is identical to $L(y, I)$ except that the holding cost is changed from $h + s$ to h . It can be easily shown that $G(y + 2, I) - G(y + 1, I) \geq G(y + 1, I) - G(y, I)$ similar to Lemma 1. $V_{t-1}(y - \xi_t, I_{t-1})$ had increasing differences (by the induction hypothesis), so $T(y) = G(y, I_{t+1}) + EV_t(y - \xi_t, I_t)$ has increasing differences. Let $g = \arg \min T(y)$. There are five cases, for each possible value of g —(i) $g \leq x$; (ii) $g = x + 1$; (iii) $g = x + C + 1$; (iv) $g \geq x + C + 2$; (v) $x + 2 \leq g \leq x + C$. Using arguments similar to the above proof of V_1 and the fact that T has increasing differences, one can show that V_t has increasing differences.

$V_t(x + 2, I_{t+1}) - V_t(x + 1, I_{t+1}) \geq V_t(x + 1, I_{t+1}) - V_t(x, I_{t+1})$
 $\forall x, I_{t+1}$ follows by induction. \square

References

- Buckley, S., R. Jayaraman, G. Lin, N. Nayak, R. Srinivasan, J. M. Swaminathan, M. Taubenblatt. 1995. Responding to demand uncertainty through supplier flexibility and safety stocks. *INFORMS*, Los Angeles, CA, (April).
- Gallego, G., O. Ozer. 2001. Integrating replenishment decisions with advance order information. *Management Sci.* **47** 1344–1360.
- Gallego, G., O. Ozer. 2003. Optimal replenishment policies for multiechelon inventory problems under advance demand information. *Manufacturing Service Oper. Management* **5** 157–175.
- Gullu, R. 1996. On the value of information in dynamic production/inventory problems under forecast evolution. *Naval Res. Logist.* **43** 289–303.
- Hausman, W. H. 1969. Sequential decision problems: A model to exploit existing forecasters. *Management Sci.* **16**(2) B93–B110.
- Hausman, W. H., R. Peterson. 1972. Multiproduct production scheduling for style goods with limited capacity, forecast revisions and terminal delivery. *Management Sci.* **18**(7) 370–383.
- Heath, D. C., P. L. Jackson. 1994. Modeling the evolution of demand forecasts with application to safety stock analysis in production/distribution systems. *IIE Trans.* **26**(3) 17–30.
- Kaminsky, P., J. M. Swaminathan. 2001. Utilizing forecast band refinement for capacitated production planning. *Manufacturing Service Oper. Management* **3**(1) 68–81.
- Karaesman, F., J. A. Buzzacott, Y. Dallery. 2002a. Integrating advance order information in make-to-stock production systems. *IIE Trans.* **34** 649–662.
- Karaesman, F., G. Liberopoulos, Y. Dallery. 2002b. The value of advance demand information in production/inventory systems. Working paper, Koc University, Istanbul, Turkey.
- Lee, H. L., V. Padmanabhan, J. Whang. 1997. Information distortion in a supply chain: The bull whip effect. *Management Sci.* **43**(4) 546–548.
- Morton, T. E., D. W. Pentico. 1995. The finite horizon nonstationary stochastic inventory problem: Near-myopic bounds, heuristics, testing. *Management Sci.* **41**(2) 334–343.
- Swaminathan, J. M., S. R. Tayur. 2004. Tactical planning models for supply chain management. Steve Graves, Ton de Kok, eds. *OR/MS Handbook on Supply Chain Management: Design, Coordination and Operation*, Vol. II. Elsevier Publishers, Amsterdam, The Netherlands, 423–456.
- Toktay, L. B., L. M. Wein. 2001. Analysis of a forecasting production inventory system with stationary demand. *Management Sci.* **47**(9) 1268–1281.