

Finite Capacity Production Planning with Random Demand and Limited Information

Michael Albritton, Alexander Shapiro and Mark Spearman

School of Industrial and Systems Engineering

Georgia Institute of Technology, Atlanta, Georgia 30332-0205 USA

April 1999

Revised January 2000

This work was supported, in part, by Grant DMI-9713878 from the National Science Foundation.

Abstract

Production planning has a fundamental role in any manufacturing operation. The problem is to decide what type of, and how much, product should be produced in future time periods. The decisions should be based on many factors, including period machine capacity, profit margins, holding costs, etc. Of primary importance is the *estimate* of demand for manufacturer's products in upcoming periods.

Our focus is to address the production planning problem by including in our models the randomness that exists in our estimates for future demands. We solve the problem with two variants of Monte Carlo sampling based optimization techniques, to which we refer as "simulation based optimization" methods. The first variant assumes that we know the actual demand distribution (assumed to be continuous) with which we approximate the true optimal solution by averaging sample estimates of the corresponding expected value function. The second approach is useful when we have limited information about the demand distribution. We illustrate the robustness of the approach by comparing a three mass-point approximation of the continuous distribution to the results obtained using the continuous distribution. This second approach is particularly appealing as it results in a solution that is close to optimal while being much faster than the continuous distribution approach.

1. Introduction

We consider a multi-product (P products) make-to-stock manufacturing environment. Products are manufactured in a single stage, on one of a variety (S) of resources (tools). The products compete for these resources—since each resource has a finite capacity. Further, demand for each product is random, can have large variations, and cannot be forecasted accurately. When demand for a product exceeds the available inventory for that product, the manufacturer will lose sales. Thus, a production planner in such an environment must decide how much product to make on which resource in each time period to satisfy that time period’s—and future time periods’—random demands.

The problem we address occurs after inventory has been netted from demand (i.e., the net-demand output from a MRP module). In each period, we attempt to meet this net-demand (henceforth referred to as simply “demand”) while not violating capacity constraints. We further assume that the planning cycle is longer than the manufacturing cycle time (e.g., a month). We thus have a rolling time horizon problem. The problem can be complicated when demand in future periods exceeds capacity. Then inventory must be built ahead of time to satisfy this future demand. In such a case, the decision maker is faced with a problem for which there is little intuition. If she decides to produce too much, the manufacturer is left with an uncomfortable amount of inventory—which may face obsolescence. If too little, the manufacturer has missed a revenue opportunity.

Clearly, a deterministic schedule will be of limited use if forecast errors are large. Having dimensions of product, resource, and time period causes dynamic programming approaches to explode in size. Traditional simulation using a hill climbing approach is very slow. Moreover, the cost parameters are not known in this situation (i.e., the decision maker doesn’t know the true cost of losing a sale or even the true cost of carrying inventory).

This paper addresses these issues—and makes the following contributions:

1. Two approaches for which cost parameters do not have to be set *a priori*,
2. Accurate models of the stochastic nature of the problem,
3. Algorithms for optimizing the model which employ the demand randomness.
4. A demonstration of the effect of having (and then using) less information than a completely defined distribution function.

These results have been implemented with a local manufacturer.

Consider solving the production planning problem where future demands are known. The optimal solution for such a problem can be achieved using linear programming (LP). Here, the objective function would include the cost of holding inventory and the “cost” of missing a sale. We minimize the overall cost subject to capacity (and non-negativity) constraints. We refer to this as the deterministic production scheduling procedure (DET).

We provide two variants of a statistical sampling optimization algorithm for solving this production planning problem to which we refer as “simulation based optimization”. Both create production schedules over a T period rolling time horizon. Thus, the problem has PST decision variables. SBO is an optimization procedure that includes statistical sampling. We show that as our sample size N increases, the SBO solution approaches optimality.

Due to the amount of sampling that takes place during the procedure, the SBO method cannot practically solve problems with a large number of decision variables. The second variant uses a probability mass function to model the demand of each product in each period of the horizon. The production planning problem is then solved using a variation of the SBO procedure. This solution approaches optimality as sample size grows—much faster than that of the SBO solution working with a continuous distributions. We refer to this procedure as “discrete simulation based optimization”, or DIS SBO. This approach is particularly useful when there is insufficient information to completely specify the probability density function for demand.

We provide a variety of single-product/resource production planning cases, as well as a multi-product/resource case, to demonstrate the relative effectiveness of the procedures under a variety of conditions.

We provide a literature survey in section 2. Sections 3 and 4 develop the SBO and DIS SBO procedures, respectively, including comparisons to the original DET procedure and to each other. Section 5 addresses the procedures’ relative effectiveness when errors in choosing a demand trend occur. Section 6 summarizes the procedures offering the pros and cons of each, along with the environmental applicability of each.

2. Literature Review

Using simulation based optimization techniques to solve the production planning problem subject to random demands is a new area of research. One of the relevant works in the simulation based optimization area includes Homem-de-Mello, et al (1999), where a simulation based optimization procedure is developed in order to find optimal material release times.

We structure our literature review in the following three areas:

1. Infinitesimal perturbation analysis,
2. Deterministic demand production planning,
3. “Uncertain” demand production planning.

The first, infinitesimal perturbation analysis, is a technique that we use extensively in our gradient estimations. The deterministic and uncertain demand production planning areas survey production planning models when demand is either known or unknown, respectively.

Infinitesimal Perturbation Analysis

We will make use of perturbation analysis (Glasserman, 1991, and Ho and Cao, 1991, and references therein) in both our SBO and DIS SBO techniques. This technique allows us to estimate gradients with *one sample path*. This yields a powerful result, greatly reducing the “noise” which would exist in estimating the gradient by other means.

Togzylowski, et al. (1986) used a solution procedure that mixes Lagrangian relaxation and perturbation search techniques to solve the lot-size scheduling problem hierarchically. Hasan (1996) used perturbation analysis to determine optimal product release times.

Deterministic Demand Production Planning

As previously noted, many production planning models assume that demands are deterministic. Some models include multi-product, multi-resource, and multi-echelon environments which can be quite complex.

Nagasawa, et al (1985) addressed the problem of finding the proper length of a planning horizon based on the unit time increments (week versus month) used within the horizon and the cost structure of the manufacturer.

Homburg (1996) modeled a decentralized organization. Here, multiple departments, each with their own objectives and constraints, “compete” for company production resources. He modeled the interaction as a linear programming problem with multiple objectives.

Birewar, et al (1990) took advantage of the structure of a particular scheduling problem and graph representation to achieve detailed production schedules derived from their multi-period LP model’s solution.

Das and Subhash (1994) modeled the master production scheduling problem as a mixed integer programming problem—incorporating product families and demand priorities.

Huang and Xu (1998) transformed the multi-stage, multi-item aggregate scheduling problem into a static job-assignment problem. They used an adapted Frank Wolfe algorithm to solve their modified problem.

Qiu and Burch (1997) modeled a hierarchical production planning problem for a yarn-fiber facility.

Uncertain Demand Production Planning

There have been a number of other approaches to production planning with unknown demands. Zapfel (1996) proposed a hierarchical model that can be incorporated with MRP II to schedule production with uncertain demand. Demand is only known for the first time period in the horizon. For the later time periods in the horizon, only upper and lower bounds for demand are known.

Kelle, et al. (1994) considered randomness in demand for a single-product, single-machine line with setups in the process industry. They formulated a model that incorporates mean and standard deviation of demand in the planning horizon time periods to set production runs. Though only one product was being made, start-ups

after periods of idleness required significant setups.

Feiring and Sastri (1990) discussed a procedure for determining production levels for a product with normally distributed demand during a rolling time horizon. They did so by minimizing total costs (including production and holding costs) over the horizon subject to a deterministic equivalent constraint for service level.

Glasserman and Tayur (1995, 1996) developed estimators of derivatives in multi-echelon systems operating under base-stock policies and minimized holding and back-order costs by approximating echelon inventory by a sum of exponentials. Kira, et al (1997) used a hierarchical approach to solve the multi-period, multi-product family production scheduling problem with a finite set of possible demand realizations. Martel et al (1995) used a piecewise linear approximate of a multi-item procurement problem to reduce its form to that of the constrained multi-item deterministic lot-sizing problem with backlogging.

Ierapetritou and Pistikopoulos (1994) proposed a two-stage stochastic programming model for setting production levels for continuous products (chemicals) using upper and lower bounds on capacity, raw material availability, and demand. Similarly, Clay and Grossman (1997) solved a two-stage fixed-recourse problem by bounding RHS stochastic variables.

Kelle (1985) addressed the problem of determining safety stock levels for multiple products in a multi-stage production environment subject to random demand and failures. He formulated approximate solutions to the problem of minimizing safety stock levels in order to maintain a predetermined probability of continuous supply.

Grubbstrom and Molinder (1996) developed a method for setting safety stocks using Laplace transforms. Grubbstrom (1998) extended this method to incorporate the net present value of the annuity streams.

3. The Model

Notation

We begin with the most general model and then modify it to make it easier to obtain parameters.

The following are parameters for the problem:

- P = the number of products
- T = the number of periods
- S = the number of stations (tools)
- h_{it} = holding cost of product i in period t
- π_{it} = marginal profit of product i in period t
- c_{st} = capacity constraint for station s in period t
- a_{is} = the amount of station s used for each unit of product i

We let D_{it} be a random variable representing the demand for product i in period t . Our decision variable is x_{its} representing the amount of product i made in period

t at station s . We denote $x_{it} := \sum_{s=1}^S x_{its}$. Finally, let I_{it} be the ending inventory of product i in period t . Note the following recursive equation

$$I_{it} = [I_{i,t-1} + x_{it} - D_{it}]^+, \quad t = 1, \dots, T, \quad (3.1)$$

for any $i \leq P$, where $[a]^+ := \max\{a, 0\}$.

Since unmet demand is lost, the general profit function is

$$Z := \sum_{i=1}^P \sum_{t=1}^T \pi_{it} \min\{I_{i,t-1} + x_{it}, D_{it}\} - h_{it} I_{it}, \quad (3.2)$$

and the capacity constraints are

$$\sum_{i=1}^P a_{is} x_{its} \leq c_{st} \quad \text{for all } s \leq S \text{ and } t \leq T. \quad (3.3)$$

The demand D_{it} is not known and is considered as a *random variable* with a known distribution. We assume that D_{it} are independent and have log-normal distributions, i.e., $\log(D_{it}) \sim N(\mu_{it}, \sigma_{it}^2)$. Log-normal distributions gave a good fit to available historical data of our local manufacturer. We then address the problem of maximizing the profit subject to the capacity constraints. Since the profit Z is a function of D_{it} , it is also random. Therefore, we maximize the expected value $\mathbb{E}\{Z\}$ of the above profit (i.e. we optimize on *average*), with respect to the decision variables x_{its} and subject to the capacity constraints (3.3). This leads to the following optimization problem

$$\begin{aligned} & \max_{x \in \mathbb{R}^{PTS}} \mathbb{E}\{Z\} \\ & \text{subject to} \quad \sum_{i=1}^P a_{is} x_{its} \leq c_{st}, \quad s = 1, \dots, S, \quad t = 1, \dots, T, \\ & \quad \quad \quad x_{its} \geq 0, \quad i = 1, \dots, P, \quad t = 1, \dots, T, \quad s = 1, \dots, S, \end{aligned} \quad (3.4)$$

where $x := (x_{111}, \dots, x_{PTS})$ is the vector of the decision variables.

Let us observe that

$$\begin{aligned} \min\{I_{i,t-1} + x_{it}, D_{it}\} &= D_{it} + \min\{I_{i,t-1} + x_{it} - D_{it}, 0\} \\ &= D_{it} + I_{i,t-1} + x_{it} - D_{it} - (I_{i,t-1} + x_{it} - D_{it})^+ \\ &= D_{it} + I_{i,t-1} + x_{it} - D_{it} - I_{it} = I_{i,t-1} + x_{it} - I_{it}, \end{aligned}$$

and hence

$$\begin{aligned} Z &= \sum_{i=1}^P \sum_{t=1}^T \pi_{it} (I_{i,t-1} + x_{it} - I_{it}) - \sum_{t=1}^T h_{it} I_{it} = \\ & \sum_{i=1}^P \left[\sum_{t=1}^T \pi_{it} x_{it} + \sum_{t=1}^{T-1} (\pi_{i,t+1} - \pi_{it} - h_{it}) I_{it} + \pi_{i1} I_{i0} - (\pi_{iT} + h_{iT}) I_{iT} \right]. \end{aligned} \quad (3.5)$$

We have, by the recursive equation (3.1), that I_{it} are convex functions of the vector x of decision variables. It follows then by (3.5) that Z , and hence its expected value $\mathbb{E}\{Z\}$, is a concave function of x , provided $\pi_{i,t+1} - \pi_{it} - h_{it} \leq 0$, $t = 1, \dots, T-1$, $i = 1, \dots, P$. Since h_{it} are positive, this certainly holds if all π_{it} are equal to each other. Therefore, in that case, the optimization problem (3.4) is a *convex* programming problem, subject to linear constraints.

Obtaining Parameter Values

In working with our client, we found no managers who were comfortable with setting, a priori, the value of π_{it} or h_{it} . However, one manager was emphatic that missing sales is always bad, and thus each π_{it} should be the same for each product over the entire planning horizon. We believe the same should be true for the h_{it} values as well. Thus, we set

$$\pi_{it} = \pi \text{ and } h_{it} = h, \quad i = 1, \dots, P, \quad t = 1, \dots, T, \quad (3.6)$$

meaning that only the ratio π/h is relevant. We also have found that, in practice, managers rarely know the true values π and h . They can, however, decide on an acceptable ratio, using trade-off curves. The amount of lost sales as a function of inventory is a convex function. After solving the problem with various π/h ratios, we can build a graph of this function, as we have done in Figure 1. Using their judgment, managers can find where they would like to be on such a graph—describing a particular π/h ratio.

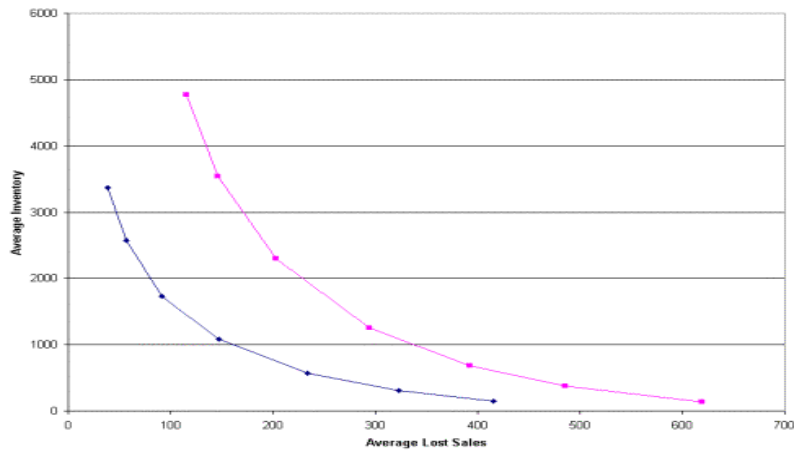


Figure 1: Average Inventory Versus Average Lost Sales (Balanced Trend, Cap=200)

We can now use the solution pertaining to the chosen ratio to schedule production.

Perturbation Analysis

We estimate values of the first order partial derivatives of the objective function with respect to the decision variables using *perturbation analysis*. For the sake of notational simplicity we drop in this section the subscript i indicating a considered

product and s indicating a considered tool. Consider the process I_t defined by the recursive equation (3.1).

Let τ_1 be the first period the process I_t hits zero, i.e., $I_t > 0$ for $t = 1, \dots, \tau_1 - 1$ and $I_{\tau_1-1} + x_{\tau_1} - D_{\tau_1} \leq 0$ and hence $I_{\tau_1} = 0$. Let τ_2 be the second time I_t hits zero, etc. Note that if $I_{\tau_1+1} = 0$, then $\tau_2 = \tau_1 + 1$, etc. Note also that if the distribution of the random vector $D := (D_1, \dots, D_T)$ is continuous and has a density (which is certainly the case if random variables D_1, \dots, D_T are independent and each has a log-normal distribution), then the event: “ $I_{t-1} + x_t - D_t = 0$ given $I_{t-1} > 0$ ” does not happen with probability one (for a given x_t).

Let $0 < \tau_1 < \dots < \tau_n \leq T$ be the sequence of hitting times and let $\tau_0 = 0$. For a given time $t \in \{1, \dots, T\}$, let $\tau_{i-1} \leq t < \tau_i$. We have then

$$\frac{\partial I_s}{\partial x_t} = \begin{cases} 1, & \text{if } t \leq s < \tau_i \text{ and } t \neq \tau_{i-1}, \\ 0, & \text{otherwise.} \end{cases} \quad (3.7)$$

It follows that

$$\frac{\partial Z}{\partial x_t} = \begin{cases} \pi - h(\tau_i - t), & \text{if } t \neq \tau_{i-1}, \\ \pi, & \text{otherwise.} \end{cases} \quad (3.8)$$

Similarly if $\tau_n < T$ and $\tau_n \leq t < T$, we have

$$\frac{\partial Z}{\partial x_t} = \begin{cases} -h(T - t + 1), & \text{if } t \neq \tau_n, \\ \pi, & \text{otherwise,} \end{cases} \quad (3.9)$$

and

$$\frac{\partial Z}{\partial x_T} = \begin{cases} -h, & \text{if } \tau_n < T, \\ \pi, & \text{if } \tau_n = T. \end{cases} \quad (3.10)$$

The above formulas allow us to estimate first order partial derivatives, of the expected value function $\mathbb{E}\{Z\}$, by the method of Monte Carlo simulation. That is, a random sample D^1, \dots, D^N , of N independent realizations of the demand vector D , is generated. The expected values of the above quantities are then estimated by the corresponding sample averages. Let us make the following observations.

The partial derivative $\partial I_s / \partial x_t$ exists as long as the event “ $I_{t-1} + x_t - D_t = 0$ given $I_{t-1} > 0$ ” does not happen. Since we assume that the demand vector D has independent components, each having a log-normal distribution, it follows that the above event does not happen with probability one, and hence $\partial I_s / \partial x_t$ exists with probability one. Moreover, it follows that I_s , and hence Z , are differentiable functions of the vector x of decision variables with probability one. Since I_s are convex functions of x , for any realization of the demand vector D , this implies that $\mathbb{E}\{I_s\}$ are differentiable functions of x and, moreover, partial derivatives can be taken inside the expected value, that is $\partial \mathbb{E}\{I_s\} / \partial x_t = \mathbb{E}\{\partial I_s / \partial x_t\}$. This follows easily from the Monotone Convergence Theorem.

By (3.7) we obtain then that

$$\frac{\partial \mathbb{E}\{I_s\}}{\partial x_t} = \begin{cases} 0, & \text{if } s < t, \\ \delta_t, & \text{if } s = t, \\ p_{st} & \text{if } s > t, \end{cases} \quad (3.11)$$

where p_{st} is the probability of the event “ $s < \tau_i$ given $\tau_{i-1} \leq t < \tau_i$ ” and δ_t is the probability that t is not a hitting time. For $s > t$, the probability p_{st} is estimated by the corresponding sample proportion \hat{p}_{st} . Note that the variance of the estimator \hat{p}_{st} is $p_{st}(1 - p_{st})/N$.

Numerical Optimization Methods

In order to solve the “true” (expected value) optimization problem (3.4) we construct an approximate problem by using Monte Carlo simulation techniques. That is, we generate a random sample D_i^1, \dots, D_i^N , of the demand vector D_i of product i , $i = 1, \dots, P$, and estimate $\mathbb{E}\{Z\}$ by the corresponding sample average \hat{Z}_N . Note that \hat{Z}_N depends on the generated sample, and hence is random. After such a sample is fixed (generated), $\hat{Z}_N = \hat{Z}_N(x)$ becomes a deterministic function of the vector x of the decision variables. Consequently we approximate problem (3.4) by the so-called “sample average approximation” problem:

$$\begin{aligned} \max_{x \in \mathbb{R}^{PTS}} \quad & \hat{Z}_N(x) \\ \text{subject to} \quad & \sum_{i=1}^P a_{is} x_{its} \leq c_{st}, \quad s = 1, \dots, S, \quad t = 1, \dots, T, \\ & x_{its} \geq 0, \quad i = 1, \dots, P, \quad t = 1, \dots, T, \quad s = 1, \dots, S. \end{aligned} \quad (3.12)$$

The above problem is “stochastic” in the sense that it depends on the generated sample. However, the moment the sample is fixed (generated) it becomes a deterministic optimization problem.

Let us make the following observations while again dropping the subscripts i and s , for simplicity. The function $\hat{Z}_N(\cdot)$ is a piecewise linear concave function. Therefore, since the constraints in the above problem are linear, in principle this problem can be formulated as a linear programming problem. Note, however, that such a linear programming problem will be subject to a large number of constraints, and it is not obvious how to write it explicitly. Let us remark that at any given point x we can calculate a gradient of $\hat{Z}_N(\cdot)$ (or rather supergradient if $\hat{Z}_N(\cdot)$ is not differentiable at x). That is, the corresponding partial derivatives $\partial \hat{Z}_N(x)/\partial x_t$, $t = 1, \dots, T$, are obtained by using sample averages derived according to formulas (3.8-10), and the corresponding gradient (or rather supergradient) is formed by $\nabla \hat{Z}_N(x) = (\partial \hat{Z}_N(x)/\partial x_1, \dots, \partial \hat{Z}_N(x)/\partial x_T)$.

An approach to numerical solution of the above problem is to use a cutting plane strategy by employing an outer linearization of $\hat{Z}_N(\cdot)$. That is, at each iteration point x^ν , the new cut $h_\nu(x) := \hat{Z}_N(x^\nu) + \nabla \hat{Z}_N(x^\nu)^T(x - x^\nu)$ is added. Since the function $\hat{Z}_N(\cdot)$ is concave, we have that $\hat{Z}_N(\cdot) \leq h_\nu(\cdot)$. Based on accumulated cuts the corresponding linear programming problem is solved, a new iteration point is obtained, etc. This can be combined with the “trust-region method” by restricting the possible move at each iteration to a “box” around the current iteration point. Thus, the new value for each x_{its} must be within the region $x_{its} \pm b_{its}$.

Algorithm

The algorithm for our procedure incorporates an embedded deterministic convex programming routine for optimizing profit over a particular sample of demands. This

routine is then repeated over different samples of increasing size until an overall solution with acceptable confidence is found. The generated samples are indexed by k and the individual “steps”, taken in the algorithm within a particular sample, are indexed by r .

Optimizing a Current Sample

We begin by solving our production scheduling problem with deterministic demands given by the corresponding mean values. This deterministic problem can be formulated as an LP. With these initial values of our production variables, we generate an initial sample of demands. For each set of demand realizations in the sample, we calculate corresponding inventories (3.1), profit value (3.2), and gradient estimates (3.8-10) for each sample point based on our current production variable values. The gradient and profit estimates are averaged over the sample in order to obtain overall sample estimates for each.

By using the current gradient estimate $\nabla \hat{Z}_{N_k}$ and employing the trust region method, we adjust our production variables. Based on our new production variable levels, we repeat the process with the same demand sample until an optimality criterion over the sample is reached. We determine whether or not we are near the current sample’s optimum by performing the paired t-test between successive objective function values, which is based on the test statistic

$$t := \frac{\hat{Z}_r - \hat{Z}_{r-1}}{s_p \sqrt{\frac{2}{N_k}}}, \quad (3.13)$$

where $s_p^2 := (s_r^2 + s_{r-1}^2)/2$ and s_i^2 refers to the sample variance of the i^{th} iterate.

Resampling

Once our algorithm stops for a given sample, a choice of resampling or terminating the algorithm is made. Notice that we are at the peak of this sample’s profit curve. Thus, the sample’s average profit estimate is denoted as \hat{Z}_r , where r is the last step of the optimization algorithm for the current sample. We calculate the 95% confidence interval around our profit estimate \hat{Z}_r . If the half width of this interval, denoted ℓ_k , is less than a predetermined percentage $\alpha \hat{Z}_r$ of our profit estimate, we do not continue sampling. If the confidence region is wider than predetermined bounds, however, we must create a larger demand sample and repeat the deterministic optimization procedure. The new sample size is chosen as follows:

$$N_{k+1} := N_k \left(\frac{\ell_k}{\alpha \hat{Z}_r} \right)^2. \quad (3.14)$$

When we achieve a sample optimum, whose confidence region is within our predetermined bounds, we terminate the algorithm. The resulting production variable values, which represent the optimum values found in the last sample, represent our simulation based estimates of the corresponding optimal solution of the “true” problem.

Pseudo-Code

Let k and r denote the sample number and step number, respectively.

Initialize x_{its} (Solve LP using mean demands)

Initialize values b_{its} of the box size and sample size N_1 ;

WHILE we are still sampling

FOR $l = 1, \dots, N_k$, generate D_{it}^l

WHILE we are still optimizing a sample

Calculate I_{itr}^l 's

Calculate Z_{itsr}^l 's

Calculate τ_{itr}^l 's

Calculate each $\frac{\partial Z_r^l}{\partial x_{itsr}}$

Average the sample's partial derivative estimates

to obtain current estimates $\frac{\partial Z_r}{\partial x_{its}}$

Average the sample's individual objective function

components to obtain current estimates of each component Z_{itsr}

IF $r > 2$

Conduct a paired t-test between Z_r and Z_{r-1}

Recalculate each b_{its} as follows:

$$L_{itsr} = \frac{Z_{itsr} - Z_{its,r-1}}{\frac{\partial Z_{itsr}}{\partial x_{its}}(x_{its,r-1} - x_{its,r-2})}$$

IF $L_{itsr} < 0.25$

$$b_{itsr} = \frac{b_{its,r-1}}{4}$$

ELSE IF $L_{itsr} > 0.75$

$$b_{itsr} = 2b_{its,r-1}$$

ELSE

$$b_{itsr} = b_{its,r-1}$$

IF we have optimized this sample according to our paired t-test

Calculate confidence interval halfwidth ℓ_k
around profit function estimate

IF $\ell_k < \alpha Z_r$

STOP SAMPLING

ELSE

$$N_{k+1} = N_k \left[\frac{\ell_k}{\alpha Z_r} \right]^2$$

ELSE

$$\begin{aligned} & \max_{x \in \mathbb{R}^{PTS}} && [\nabla \hat{Z}_{N_k}(x_{itsr})]^T x_{itsr} \\ & \text{subject to} && \sum_{i=1}^P a_{is} x_{itsr} \leq c_{st} \\ & && |x_{itsr} - x_{its,r-1}| \leq b_{itsr}, \\ & && \hat{Z}_{N_k}(x_{itsr}) \leq h_\nu, \\ & && x_{itsr} \geq 0, \\ & && \nu = 1, \dots, r-1, \quad i = 1, \dots, P, \\ & && s = 1, \dots, S, \quad t = 1, \dots, T, \end{aligned}$$

Cases

Introduction to Method of Evaluation

We evaluate our model, comparing its performance to that of DET, by solving a realistic multi-product/resource problem and various single-product/resource scenarios which are designed to test the procedure under a variety of general conditions. The multi-product/resource and single-product/resource cases have 10 and 9 period horizons, respectively. We evaluate each case by producing a production schedule at time zero assuming zero initial inventory(ies).

Once production schedules are generated for each method at time 0, we obtain $\mathbb{E}\{Z\}$ for each method by simulating the respective solutions subject to random demands. We simulate the “rolling” effect by updating demands at each period (when they are realized) and adjusting subsequent periods’ productions based on the updated demands. We evaluate the effects over the first 4 periods, enabling us to see trend effects.

Multi-Product/Resource Case

We solve a realistic problem that is similar to our local manufacturer’s, with 5 product classes produced on 5 tools during a 10 month time horizon. The manufacturer does not know the actual marginal profit or holding cost for any of the products. We do know, however, that the holding cost to marginal profit ratio is the same for each product. The manufacturer believes that this ratio is close to $\frac{h}{\pi} = 1/36$ (resulting in an annual holding cost of approximately 33.3% of marginal profit).

The capacities of the machines are given as follows:

$$\begin{aligned}c_{1t} &= 300, t = 1, \dots, 10, \\c_{2t} &= 300, t = 1, \dots, 10, \\c_{3t} &= 300, t = 1, \dots, 10, \\c_{4t} &= 300, t = 1, \dots, 10, \\c_{5t} &= 200, t = 1, \dots, 10.\end{aligned}\tag{3.15}$$

Product class 1 can only be made on machine 1. Products 2 and 4 can be produced on machines 2 and 4. Product 3 can be produced on machines 3 and 5. Product 5 can only be produced on machine 5.

The demand distributions for each product follow log-normal distributions with the following means and standard deviations:

$$\begin{aligned}u_1 = \sigma_1 &= 200, \quad t = 1, \dots, 10, \\u_2 = \sigma_2 &= 250, \quad t = 1, \dots, 10, \\u_3 = \sigma_3 &= 275, \quad t = 1, \dots, 10, \\u_4 = \sigma_4 &= 150, \quad t = 1, \dots, 10, \\u_5 = \sigma_5 &= 75, \quad t = 1, \dots, 10.\end{aligned}\tag{3.16}$$

Each product consumes the same amount of each resource.

There are infinitely many optimal solutions to the deterministic problem—so long as the total production over all machines for each product in each period equals

the expected demand for that product. One optimal solution for the deterministic problem is as follows:

$$\begin{aligned}
x_{1,t,1} &= 200, & t &= 1, \dots, 10, \\
x_{2,t,2} &= 125, & t &= 1, \dots, 10, \\
x_{2,t,4} &= 125, & t &= 1, \dots, 10, \\
\\
x_{3,t,3} &= 275, & t &= 1, \dots, 10, \\
x_{3,t,5} &= 0, & t &= 1, \dots, 10, \\
\\
x_{4,t,2} &= 75, & t &= 1, \dots, 10, \\
x_{4,t,4} &= 75, & t &= 1, \dots, 10, \\
\\
x_{5,t,5} &= 75, & t &= 1, \dots, 10.
\end{aligned} \tag{3.17}$$

Notice that total facility monthly capacity is 1400, while the total deterministic demand is 950. By using an LP (the deterministic method, DET) to schedule production, only 950 total units are produced, and no machine is fully utilized—regardless of the inventory holding expense to marginal profit relationship.

This problem was then solved using the SBO method. The SBO solution allowed for a higher machine utilization and improved upon the DET solution—with an average profit increase of approximately 8.9%.

Single-Product/Resource Cases

We test our algorithm over various scenarios of interest. In order to observe general scenarios, we simplify our problem to observe the single-product/resource case. All demands follow log-normal distributions, and the time horizon is nine periods for each case. The four parameters we change are as follows:

1. Capacity,
2. Demand trend,
3. Variability,
4. π/h ratio.

We tested the cases with all periods' capacities equaling 100 and 200. Note that u_t denotes the demand mean for period t . The following are the tested demand trends:

1. $u = [100, 100, 100, 100, 100, 100, 100, 100, 100]$,
2. $u = [60, 70, 80, 90, 100, 110, 120, 130, 140]$,
3. $u = [140, 130, 120, 110, 100, 90, 80, 70, 60]$.

For each demand trend, two levels of variability were tested:

1. Moderate Variability— $CV = 1$,
2. High Variability— $CV = 2$.

where CV refers to the coefficient of variation, $CV = \sigma/\mu$.

For all combinations of capacity, demand trend, and variability, $\pi/h = 1, 2, 4, 8, 16, 32$, and 64 were tested.

In the analysis of these scenarios, demand trend and capacity combinations make up the six overall cases. Within each of these cases, CV and π/h are altered yielding 14 sub cases. For each case, we are able to observe the trade-off of average inventory versus average lost sales. We can conclude our comparison of the two production scheduling methods by comparing the average profits generated by our simulation based optimization method (SBO) to those of the linear programming solution with deterministic demand estimates (DET).

The average profit differences between SBO and DET depend heavily upon the capacity. By design, we have chosen two values of capacity that test “boundary” conditions. When capacity equals 100 in each period, it is equal to the average of the demand means over the time horizon for each demand trend—modeling a highly capacity constrained environment. When capacity equals 200 in each period, we have a relatively large amount of “extra” capacity. Intuitively, the former provides little room for our SBO procedure to outperform the DET procedure, unlike the latter. When viewing the average profit differences between SBO and DET, as functions of π/h , two general shapes are seen.

For the more tightly constrained capacity cases, the average profit difference between SBO and DET decreases with increasing π/h to a certain point after which both methods achieve the same average profit. Here, the SBO production schedule converges to that of the DET method, as Figure 2 depicts.

Now, consider the less tightly constrained capacity. Again, as π/h increases, the average profit difference between SBO and DET decreases to zero until, at a certain π/h value around eight, the two procedures generate a similar production schedule. At this point, however, unused capacity still exists. Thus, as π/h continues to increase, the performance of SBO begins to again dominate that of DET, as Figure 3 depicts.

The decreasing demand trend case where capacity equals 100 differs. Unlike the increasing and balanced trend cases, full resource utilization in all periods can never be achieved under DET—for any π/h value. Thus, even though capacity is tightly constrained, the average profit difference between SBO and DET has the high capacity shape, with the two methods achieving the same (statistically) average profit at only one π/h value (at $\pi/h = 8$ for $CV = 1$ and $\pi/h = 16$ for $CV = 2$). At all other values of π/h , SBO achieves greater average profits than DET.

Discussion

We have developed a method by which optimal production schedules can be generated considering demand randomness. The method has been generally developed for the multi-product/resource case.

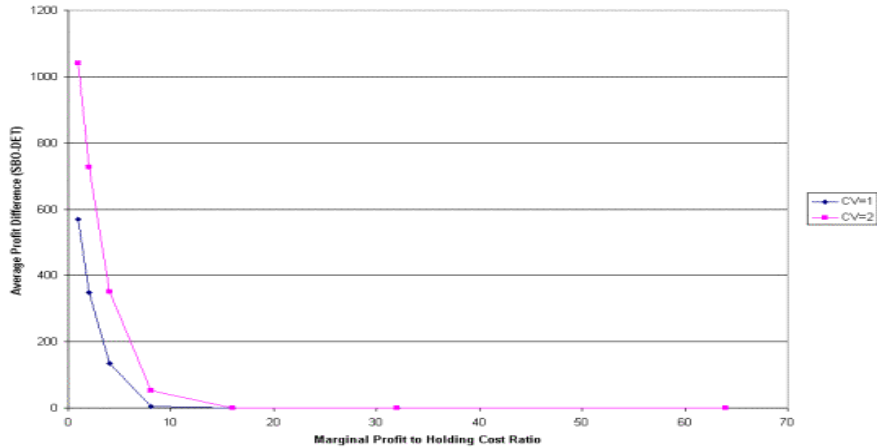


Figure 2: Average Profit Difference Between SBO and DET (Balanced Trend, Cap=100)

We solve a realistic multi-product/resource problem as well as various single-product/resource cases of interest to demonstrate the effectiveness of the method and its dominance over the deterministic counterpart.

The algorithm’s executable time is a considerable restriction. For instance, the realistic case we solved for the local manufacturer involved 5 product classes produced on 5 resources over a 10 month time horizon—thus, 250 decision variables. In each optimization step within a sample, 250 values of profit contribution and partial derivatives for each replicate had to be calculated for the LP to be solved. With sufficient values of N_k to achieve solutions with appropriate confidence, this relatively small number of product classes takes a considerable amount of time to solve.

In reality, the manufacturer produces several hundred products which have been grouped into five families. Thus, demand distributions, resource consumption, etc., was aggregated for each product class. Once solutions have been obtained for the production schedule of each product class, separating the production within a product family must be made by production coordinators at the manufacturer. Obviously, such aggregation before the solution and disaggregation after the solution has a negative impact on performance.

The current procedure can be modified to increase speed by using less information than a completely specified probability distribution. Such a change would mean that convergence to “true” optimality would be sacrificed for the sake of speed. Increasing speed, however, would allow for more product family classes to

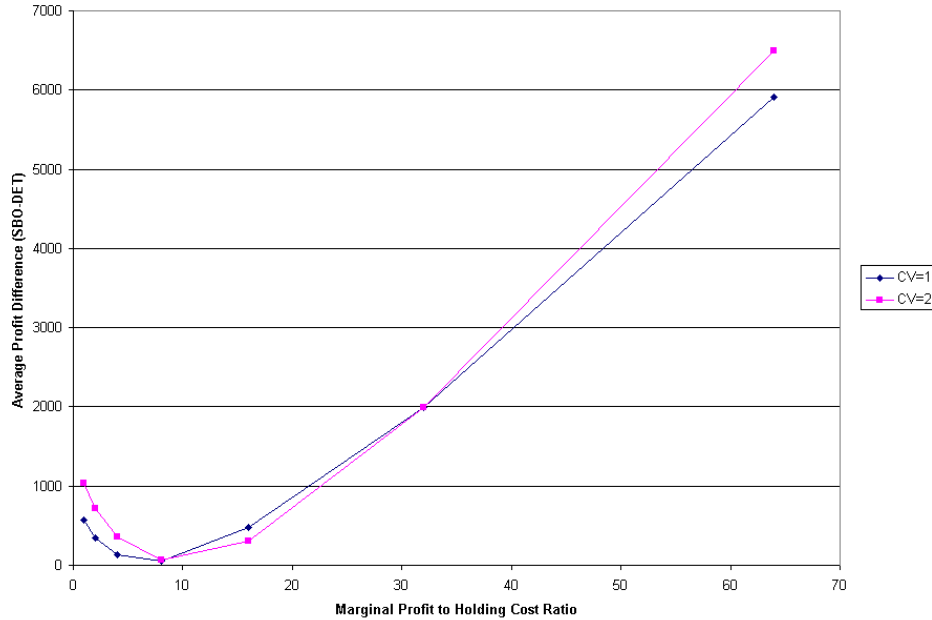


Figure 3: Average Profit Difference Between SBO and DET (Balanced Trend, Cap=200)

be considered—which may be more beneficial in the long-run.

Furthermore, it is not clear that using less information is always less appropriate. There are few instances (this one in particular) in which we have enough information to completely specify a probability distribution. Thus, using a probability distribution based on, say, the first two moments, is really less information than using a discrete distribution with three points.

The next section describes the algorithm for the discrete version of the simulation based optimization of production quantities.

4. Discrete Simulation Based Optimization

Introduction

We now propose a different approach where we use a discrete distribution to model demand. The SBO procedure then becomes simpler and faster to solve—with a good chance that, even with moderate sample sizes, the true optimal solution will be exactly reached. We refer to this method as discrete simulation based optimization (DIS SBO).

There are three important reasons for considering this new discrete demand distribution method:

1. The SBO procedure requires a relatively large amount of computational time,
2. The true distribution may actually be discrete.
3. We have insufficient data to completely specify a continuous distribution.

Method

In the discrete case, the expected value function is a piecewise linear convex function and, therefore, is not everywhere differentiable. Nonetheless, equations (3.7)-(3.10) can be used for calculation of a subgradient of the sample approximation and subsequent solution of the obtained optimization problem.

To observe the effect of discretizing the demand distribution, we will choose three characteristic points from the continuous distributions used in the previous section. Recall that we assume that demands have log-normal distributions, hence $\log(D_{it}) \sim N(\mu_{it}, \sigma_{it})$. We approximate these lognormal distributions by discrete distributions taking three values where each are equally probable. The values should be such that the mean and standard deviation of the discrete distribution are the same as the underlying log-normal distribution.

For the log-normal distribution with known mean and variance, the normal distribution's parameters, μ and σ , can easily be found by the following relationships (Law and Kelton, 1991):

$$\begin{aligned}\mu &= \ln\left(\frac{\text{mean}^2}{\text{variance} + \text{mean}^2}\right) \\ \sigma &= \sqrt{\ln\left(\frac{\text{variance} + \text{mean}^2}{\text{mean}^2}\right)}\end{aligned}\tag{4.1}$$

We set one of the discrete points, x , in our distribution to μ . If we choose the other two values such that one is $x - a$ and the other is $x + a$, the resulting PMF will have a mean of μ . Further, we can select a such that the resulting standard deviation of the PMF equals σ by setting:

$$a = \sqrt{1.5\sigma^2}\tag{4.2}$$

Our final *low*, *med*, and *high* values of our PMF are e^{x-a} , e^x , e^{x+a} , respectively.

Once these values are chosen, the solution procedure is similar to the original SBO procedure in section 3. Here, however, using a trust region is not necessary. Initial sample sizes are small, and subsequent and final sample sizes are much smaller, in practice, than the sample sizes of the original SBO procedure.

Cases

Table 1 lists the PMF conversion values for the single-product/resource case and multi-product/resource cases.

Multi-Product/Resource Case

The SBO solution, which approaches optimality, yielded an 8.9% increase in average profit over the initial deterministic solution. Our DIS SBO procedure's

Table 1: PDF to 3-Level PMF Conversion

mean	std dev	μ	σ	a	$x - a$	x	$x + a$	<i>low</i>	<i>medium</i>	<i>high</i>
60	60	3.7	0.8	1.0	2.7	3.7	4.8	15.3	42.4	117.6
70	70	3.9	0.8	1.0	2.9	3.9	4.9	17.9	49.5	137.2
80	80	4.0	0.8	1.0	3.0	4.0	5.1	20.4	56.6	156.8
90	90	4.2	0.8	1.0	3.1	4.2	5.2	23.0	63.6	176.4
100	100	4.3	0.8	1.0	3.2	4.3	5.3	25.5	70.7	196.0
110	110	4.4	0.8	1.0	3.3	4.4	5.4	28.1	77.8	215.6
120	120	4.4	0.8	1.0	3.4	4.4	5.5	30.6	84.9	235.2
130	130	4.5	0.8	1.0	3.5	4.5	5.5	33.2	91.9	254.8
140	140	4.6	0.8	1.0	3.6	4.6	5.6	35.7	99.0	274.4
60	120	3.3	1.3	1.6	1.7	3.3	4.8	5.7	26.8	126.9
70	140	3.4	1.3	1.6	1.9	3.4	5.0	6.6	31.3	148.0
80	160	3.6	1.3	1.6	2.0	3.6	5.1	7.6	35.8	169.2
90	180	3.7	1.3	1.6	2.1	3.7	5.2	8.5	40.2	190.3
100	200	3.8	1.3	1.6	2.2	3.8	5.4	9.5	44.7	211.5
110	220	3.9	1.3	1.6	2.3	3.9	5.4	10.4	49.2	232.6
120	240	4.0	1.3	1.6	2.4	4.0	5.5	11.3	53.7	253.8
130	260	4.1	1.3	1.6	2.5	4.1	5.6	12.3	58.1	274.9
140	280	4.1	1.3	1.6	2.6	4.1	5.7	13.2	62.6	296.1
200	200	5.0	0.8	1.0	3.9	5.0	6.0	51.0	141.4	392.1
250	250	5.2	0.8	1.0	4.2	5.2	6.2	63.8	176.8	490.1
275	275	5.3	0.8	1.0	4.3	5.3	6.3	70.1	194.5	539.1
150	150	4.7	0.8	1.0	3.6	4.7	5.7	38.3	106.1	294.0
75	75	4.0	0.8	1.0	3.0	4.0	5.0	19.1	53.0	147.0

solution improvement over DET was 8.4% (94% of the total possible improvement (seen using SBO) over DET).

Single-Product/Resource Cases

The DIS SBO provides for average profits (under the original demand PDF's), for $\pi/h \geq 8$, that approach those produced by the SBO procedure for balanced and decreasing demand trends. For increasing demand trends, DIS SBO average profits are considerably less than SBO, however, they achieve over 50% of the possible improvement over DET (based on SBO). Graphs 4 and 5 depict the methods' performances.

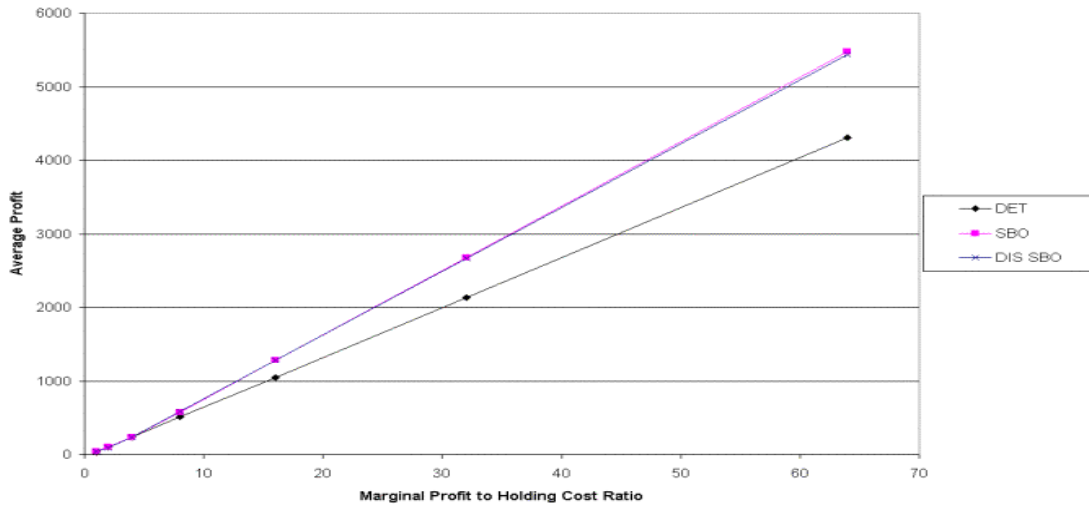


Figure 4: Average Profit (Balanced Trend, Cap=200, CV=1)

Discussion

Consider the true demands to be random variables with log-normal distributions. We have developed a procedure which uses estimates of such distributions in the crudest fashion—3 discrete values, skewed (though not equal to the log-normal skewness), with mean and standard deviation following the underlying normal distribution. Besides these general similarities, the discrete distribution is quite different than its continuous counterpart. Yet, the quality of the solutions are almost identical, except when the demand trend is increasing and capacity is high. Even in this case, DIS SBO significantly outperforms DET. Thus, we can reasonably conjecture that our procedures are not highly sensitive to the demand distributions.

5. Production Planning Procedure Sensitivity to Errors In Predicting Demand Trends

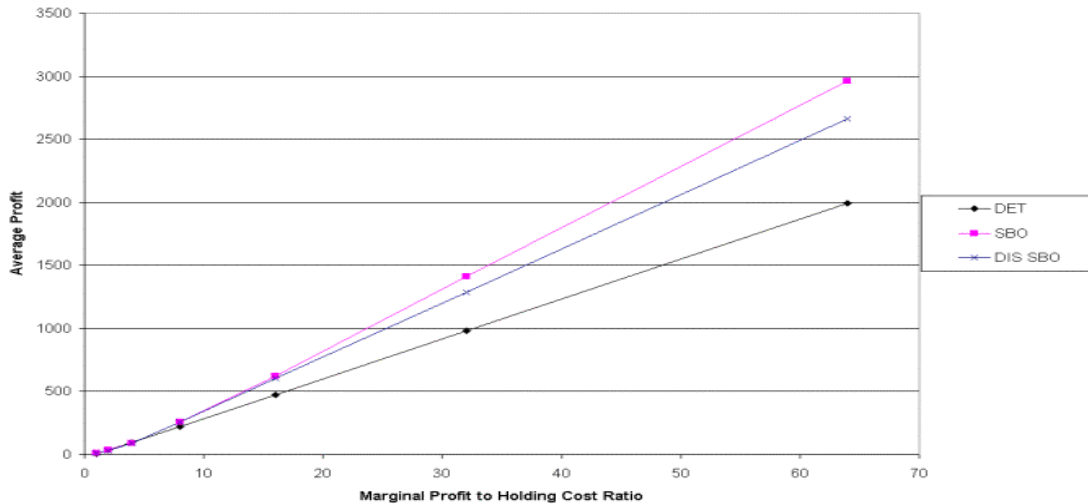


Figure 5: Average Profit (Increasing Trend, Cap=200, CV=2)

We have shown the effectiveness of the two production scheduling techniques (SBO and DIS SBO) for various capacities, variability levels, demand trends, and π/h ratios. We have compared each of the procedures' average profits to each other and the DET procedure. In doing so, we assume that though the demands are random, we know the trends and distributions. It is interesting to view the solution effectiveness when an important planning assumption fails to be observed.

If the parameters altered, the demand trend is the easiest to misjudge. We consider the increasing, balanced, and decreasing demand trends. These three basic trends are commonly found in practice. The increasing trend is associated with newer products whose demands are “ramping up” in their initial production and introduction into the market. Balanced trends are associated with products after their initial ramp-up, when they have achieved a steady, stable demand. The decreasing trend is associated with products at the end of their life-cycles—usually being phased-out while new, increasing-trend products are taking their places.

The actual time-table for the life of a product is hard to judge. Further, management might intend to ramp up demand for a new product but not achieve their goals. Thus, errors in the demand trend over the horizon should be addressed.

We denote two types of errors: primary and secondary. Though any combination of errors may occur (with three trends there are six possible scenarios for error), errors of demand trends which “border” each other in time are of the primary focus—the primary errors. These include the following:

1. Planned Increasing Trend—Observed Balanced Trend,
2. Planned Balanced Trend—Observed Increasing Trend,
3. Planned Balanced Trend—Observed Decreasing Trend,
4. Planned Decreasing Trend—Observed Balanced Trend.

The rarer (hopefully) secondary errors include:

1. Planned Increasing Trend—Observed Decreasing Trend,
2. Planned Decreasing Trend—Observed Increasing Trend.

Recall that for our single-product/resource cases, we have 4 capacity/variability combinations. Within each combination, we observe the average profit in the rolling horizon when a particular error in predicting the demand trend takes place. We compare the solutions for each production planning method and compare to the original sets of solutions where the correct demand trend prediction took place.

Because the three procedures' production levels converge (except for DET with decreasing demand trend) for the capacity of 100 cases with increasing π/h , their profits with primary and secondary errors are the same (just as when the correct demand trend was chosen).

For the capacity of 200 cases, we see that, when demand trends are correctly predicted, the DIS SBO solution approaches the optimal SBO solution.

DIS SBO performed equally as well as SBO for the balanced and decreasing demand trends. Again, just as when the demand trend was correctly predicted, DIS SBO performed well under the SBO procedure when an increasing demand trend was planned for and a balanced or decreasing trend was observed.

6. Conclusions and Recommendations

We have addressed a fundamental problem of nearly all make-to-stock manufacturers—how to do production planning when future demands are random, and maybe highly variable. We have proposed two methods for solving this problem.

The first method, simulation based optimization, allows us to solve the production planning problem to achieve optimal average profits—assuming we know the distributions which describe the products' demands.

Our second method, the discrete simulation based optimization method, performs SBO after the continuous demand distributions have been estimated by discrete distributions taking three different values. If the demands are discrete, the DIS SBO solutions will approach optimality—much faster than the SBO solutions will approach optimality under continuous distributions. If, however, the demand distributions are continuous, this approximation will allow us to greatly reduce the original SBO procedure's computational time, while benefiting from good solutions. The DIS SBO solution quality approached that of SBO for all cases—except ones which included increasing demand trends.

It is important to note the similarity between the SBO and DIS SBO production values generated by each procedure. In practice, planners might not know the actual distributions that describe demands. The SBO and DIS SBO probability distributions are skewed and have similar means and variances, however, they are very different distributions. Since the actual production schedules under our scenarios were shown to be very similar, we can conclude that our methods are not highly sensitive to distribution “misjudgments.”

We also show that our methods are not highly sensitive to errors in predicting demand trends.

Though we have considered the “lost sales” environment, an extension to this research would be to alter the formulations to consider backorders. Though profits will differ under otherwise identical scenarios, the relative goodness of the procedures should be the same.

Considering multi-level and multi-stage environments would also be of interest and may be a subject of future research.

Bibliography

- [1] BIREWAR, DEEPAK B., and IGNACIO E. GROSSMANN. 1990. Simultaneous Production Planning and Scheduling in Multiproduct Batch Plants. *Industrial & Engineering Chemistry Research* 29(4):570-80.
- [2] CLAY, R. L., and I. E. GROSSMANN. 1997. A Disaggregation Algorithm for the Optimization of Stochastic Planning Models. *Computer Chemical Engineering* 21(7):751-74.
- [3] DAS, SANCHOY K., and SUBHASH C. SARIN. 1994. Integrated Approach to Solving the Master Aggregate Scheduling Problem. *International Journal of Production Economics* 34(2):167-78.
- [4] FEIRING, BRUCE R., and TEP SASTRI. 1990. Improving Production Planning By Utilizing Stochastic Programming. *Computers and Industrial Engineering* 19(1-4):53-56.
- [5] GLASSERMAN, PAUL, 1991. *Gradient Estimation Via Perturbation Analysis*, Kluwer Academic Publishers, Boston, MA.
- [6] GLASSERMAN, PAUL, and SRIDHAR TAYUR. 1996. A Simple Approximation for a Multistage Capacitated Production-Inventory System. *Naval Research Logistics* 43:41-58.
- [7] GLASSERMAN, PAUL, and SRIDHAR TAYUR. 1995. Sensitivity Analysis for Base-stock Levels in Multiechelon Production-inventory Systems. *Management Science* 41(2):263-81.
- [8] GRUBBSTROM, ROBERT W. 1998. Net present value approach to safety stocks in planned production. *International Journal of Production Economics* 56-57:213-29.
- [9] GRUBBSTROM, ROBERT W., and ANDERS MOLINDER. 1996. Safety production plans in MRP-systems using transform methodology. *International Journal of Production Economics* 46-47:297-309.

- [10] HASAN, NIRMAL C. 1996. Transient Analysis of Stochastic Production Environments—Job Completion and Optimal Release Time Determination. Ph.D. diss., Northwestern University, Chicago.
- [11] HO, Y.C. and X.R. CAO, 1991. *Perturbation Analysis of Discrete Event Dynamic Systems*, Kluwer Academic Publishers, Boston, MA.
- [12] HOMBURG, CARSTEN. 1996. Production Planning with Multiple Objectives in Decentralized Organizations. *International Journal of Production Economics* 56-57:243-52.
- [13] HOMEN-DE-MELLO, T., A. SHAPIRO, and M. SPEARMAN. 1999. Finding Optimal Material Release Times Using Simulation Based Optimization. *Management Science* 45:86-102.
- [14] HUANG, H. J., and G. XU. 1998. Aggregate Scheduling and Network Solving of Multi-stage and Multi-item Manufacturing Systems. *European Journal of Operational Research* 105(1):52-65.
- [15] IERAPETRITOU, M. G., and E. N. PISTIKOPOULOS. 1994. Long Range Planning Under Uncertainty. *ICHEME Symposium Series* 133:427-34.
- [16] KELLE, PETER. 1985. Safety Stock Planning In a Multi-stage Production-Inventory System. *Engineering Costs and Production Economics* 9:231-37.
- [17] KELLE, P., G. CLENDENEN, and P. DARDEAU. 1994. Economic Lot Scheduling Heuristic for Random Demands. *International Journal of Production Economics* 35(1-3):337-42.
- [18] KIRA, D., M. KUSY, and I. RAKITA. 1997. A Stochastic Linear Programming Approach To Hierarchical Production Planning. *Journal of the Operational Research Society* 48:207-11.
- [19] LAW, AVERILL M., and W. DAVID KELTON. 1991. *Simulation Modeling and Analysis*. New York: McGraw-Hill, Inc.
- [20] MARTEL, ALAIN, MOUSTAPHA DIABY, and FAYEZ BOCTOR. 1995. Multiple Items Procurement Under Stochastic Nonstationary Demands. *European Journal of Operational Research* 87:74-92.
- [21] NAGASAWA, HIROYUKI, NORIYUKI NISHIYAMA, and KATSUNDO HITOMI. 1985. Decision Analysis for Determining the Optimum Planning Horizon in Aggregate Production Planning. Part 2: Difference Between Planning Horizons in Weekly and in Monthly Schedulings. *International Journal of Production Research* 23(2):423-36.
- [22] QIU, M. M., and E. E. BURCH. 1997. Hierarchical Production Planning and Scheduling in a Multi-product, Multi-machine Environment. *International Journal of Production Research* 35(11):3023-42.
- [23] TARDIF, VALERIE, and MARK L. SPEARMAN. 1997. Diagnostic Scheduling in Finite-capacity Production Environments. *Computers & Industrial Engineering* 32(4):867-78.
- [24] TOCZYLOWSKI, E., K. NOWOSAD, H. JAGDEV, and K. S. HINDI. 1986. Aggregate Lot-size Scheduling for a Class of Two-stage Batch Manufacturing Systems. *Large Scale Systems* 11(2):165-75.
- [25] ZAPFEL, GUNTHER. 1996. Production Planning in the Case of Uncertain Individual Demand Extension for an MRP II Concept. *International Journal of Production Economics* 46-47:153-64.