A Coordinated Production Planning Model with Capacity Expansion and Inventory Management

Sampath Rajagopalan • Jayashankar M. Swaminathan
Marshall School of Business, University of Southern California, Los Angeles, California 90089
The Kenan-Flagler Business School, University of North Carolina, Chapel Hill, North Carolina 27599
srajagop@marshall.usc.edu • swaminaj@bschool.unc.edu

Motivated by a problem faced by a large manufacturer of a consumer product, we explore the interaction between production planning and capacity acquisition decisions in environments with demand growth. We study a firm producing multiple items in a multiperiod environment where demand for items is known but varies over time with a long-term growth and possible short-term fluctuations. The production equipment is characterized by significant changeover time between the production of different items. While demand growth is gradual, capacity additions are discrete. Therefore, periods immediately following a machine purchase are characterized by excess machine capacity. We develop a mathematical programming model and an effective solution approach to determine the optimal capacity acquisition, production and inventory decisions over time. Through a computational study, we show the effectiveness of the solution approach in terms of solution quality and investigate the impact of product variety, cost of capital, and other important parameters on the capacity and inventory decisions. The computational results bring out some key insights—increasing product variety may not result in excessive inventory and even a substantial increase in set-up times or holding costs may not increase the total cost over the horizon in a significant manner due to the ability to acquire additional capacity. We also provide solutions and insights to the real problem that motivated this work.

(Production Planning; Multiproduct Inventory; Capacity Expansion; Equipment Purchase; Mathematical Programming; Heuristics)

1. Introduction
This paper was motivated by a problem faced by a large manufacturing firm producing a consumer product. The firm produces many different items that are stocked in inventory for sale to customers. These items are produced on machines (or production lines) that require a significant changeover time between the production of two different items. Demand for the items is growing significantly over time as the industry is young and, as a result, the firm periodically needs to add machine capacity. The firm anticipates the demand growth to continue for the next 10 years or more. Because of the discrete nature of capacity additions as opposed to the gradual increase in demand, periods immediately following a machine purchase are characterized by excess machine capacity. The conventional approach in the operations management literature is to use this excess capacity to do more changeovers and reduce lot sizes and inventories. However, this approach ignores an important alternative use of that excess capacity in such demand growth environments, namely, benefits from postponing machine investments by building up inventory to satisfy future demand growth. Thus, in periods with...
excess capacity, the firm has two options for using the excess capacity. It could do more changeovers and thus reduce lot sizes, inventories, and inventory costs. Alternatively, it could produce in excess of demand in that period and build additional inventory that can be used to satisfy future demand increments, thus delaying the purchase of the next machine that is required to meet the growth in demand.

The benefits from the delay in capital investments in some cases may be greater than the additional inventory costs. In fact, based on the data obtained from the firm we find that these benefits could be significant. For instance, a new production line in this firm costs over $10 million and delaying its purchase by even six months can save them over $500,000 while the cost of building up additional inventory is about $200,000. Relative to this alternative, using all the excess capacity to reduce lot sizes and inventories would reduce costs by about $100,000. Thus, the option of building up inventory to postpone the machine purchase is better than the latter alternative by about $200,000. As a result, there are substantial benefits from using a model that provides management with insights on when to purchase additional capacity and how to plan production and inventory for multiple products.

In this paper, we present a multi-item, multiperiod deterministic optimization model that captures the trade-offs outlined above. In particular, we consider the objective of minimizing production, inventory, and capacity purchase costs over the planning horizon. The decision variables are (1) in which period additional capacity should be procured (binary variables) and how much; (2) what should be the production quantities and lot sizes for each product in each period in the horizon. The presence of binary decision variables (which determine when capacity should be added) and nonlinear capacity constraints makes it challenging to obtain effective solutions. We present a Lagrangian relaxation procedure to develop a lower bound. We present two heuristics: one based on the Lagrangian solution and another based on a dynamic programming approach. In our computational study, we evaluate the performance of our heuristics and provide insights into the impact of various parameters on the optimal strategy. In particular, we explore (1) when it is optimal to use excess capacity to build inventory and postpone machine purchases, (2) when it might be better to use the excess capacity to reduce lot sizes and inventories, (3) what is the effect of increased product variety, and (4) what effects do we observe with changes in factors such as holding cost, set-up time, size of capacity increment, and cost of capital. Finally, we conducted a limited computational study with the large data set obtained from the firm to test the effectiveness of our heuristics to solve large industry size problems and to understand the differences in the alternative strategies.

The rest of the paper is organized as follows. In §2 we discuss the relevant literature. In §3 we present our model and assumptions. In §4 we present a lower bound and two heuristics. In §5, we present our computational results and discuss the application that motivated this research. In §6, we discuss variants of the model presented in §3; we conclude in §7.

2. Relevant Literature
The operations management literature is replete with papers on production planning and scheduling (see Graves 1981 and Nam and Logendran 1992 for reviews). Production scheduling is concerned with the allocation of production resources while production planning is concerned with the determination of the level of production resources over time (Graves 1981). An extensive literature exists in production planning, often referred to as aggregate planning, and in production scheduling, which has further developed into the lot-sizing and machine scheduling literature for closed and open shops, respectively, as defined by Graves (1981). Nam and Logendran (1992) provide a survey of models and methodologies in the aggregate planning literature. Billington et al. (1983) and Maes Van Wassenhove (1988) provide reviews of the machine scheduling literature. We present a Lagrangian relaxation procedure to develop a lower bound. We present two heuristics: one based on the Lagrangian solution and another based on a dynamic programming approach. In our computational study, we evaluate the performance of our heuristics and provide insights into the impact of various parameters on the optimal strategy. In particular,
in the aggregate planning area include Buxey (1995) and Singhal and Singhal (1996). The aggregate planning literature generally does not consider equipment capacity issues, even though they are relevant in environments such as the one described earlier. Although some researchers have analyzed capacity expansion along with inventory and aggregate planning (Rao 1976), they restrict themselves to a single-product environment and do not consider lot-sizing issues. As a result, this stream of work does not consider the type of trade-offs outlined in the previous section, e.g., benefits and costs of postponing machine purchases by building inventory versus reducing lot sizes during periods of machine underutilization.

The scheduling function determines the allocation of resources to individual products and the lot sizes of the products to minimize inventory costs and set-up costs related to changeovers or set-ups while satisfying equipment capacity constraints. Since the production environment we consider is a closed shop (Graves 1981), the lot-sizing literature is relevant here but not the machine-scheduling literature. Starting with the standard EOQ model and its variants, this literature considers lot sizing in multiperiod environments with capacity constraints at a single stage (Maes and Van Wassenhove 1988) as well as different bill-of-material structures (Billington et al. 1983). Tempelmeier and Derstroff (1996) wrote a recent paper in this area considering a fairly general lot-sizing problem. The economic lot scheduling literature (ELSP) also considers lot sizing in a constant-demand environment with implicit capacity constraints (Elmaghraby 1978). Lopez et al. (1991) provide a survey of the research on this problem. However, the lot-sizing and ELSP literature generally treat the equipment capacity available as an exogenous parameter. Also, if set-up times are included in the model, the implicit assumption in this literature is that excess capacity in a period is used to increase the number of changeovers and thus reduce lot sizes and inventories.

The other area of research relevant to this problem is the capacity acquisition literature (Luss 1982, Rajagopalan 1998). While this literature does consider equipment capacity and when to purchase additional capacity, it generally ignores inventory and lot-sizing issues. But these issues are linked because a new machine acquired can increase capacity by, say, 15% and this excess capacity can be used in the short run to reduce lot sizes and inventories by up to 50%, as was the case in the consumer product firm motivating this study. Thus, we are not aware of any papers that consider a problem similar to the one addressed here. However, parts of our model are similar to models in the aggregate planning, lot-sizing, and capacity acquisition literature.

3. Model

We consider a scenario with \( M(i = 1 \ldots M) \) items and \( T(t = 1 \ldots T) \) periods. Demand for item \( i \) in period \( t \) is equal to \( d_{it} \). The period here may be a quarter or six months, depending on the environment. We assume that total demand for all items is increasing over time, i.e., \( \sum_i d_{it} < \sum_i d_{i, t+1} \), although demand for individual items may not be increasing. We assume that capacity requirements are evaluated and capacity purchase decisions made only at the beginning of each period. Let the initial capacity at the beginning of Period 1 be \( B \). Capacity can be purchased only in increments of size \( b \). Capacity here is measured in units of time available on a machine per period and not in units of output. For example, if each period is a quarter, a machine working 8 hours per day, 20 days per month will have a capacity of 480 hours per quarter. The actual output from a machine in a period depends on the processing time per unit and also changeover time and number of changeovers per period. Producing item \( i \) requires a processing time \( \alpha_i \) per unit and a set-up or changeover time \( \beta_i \); we assume that set-up times are not sequence dependent.

We assume that at most one capacity purchase is made in a period as the capacity of a single unit of equipment is large enough to satisfy the increase in demand over multiple periods. We also assume that a single resource type is used to manufacture the products and there are no components or subassemblies. These assumptions were appropriate in the manufacturing environment that motivated this study and are applicable in many other industrial environments. We later extend the model and the analysis to situations where multiple capacity increments are purchased in
a period and where capacity can be purchased in continuous increments (see §6). The cost of a capacity purchase in period $t$ is $g_{it}$, which includes initial purchase costs and fixed maintenance costs associated with the equipment, less salvage value at the end of $T$. Let $C_t$ denote the capacity available in period $t$ after any capacity purchases in that period, with $C_0 = B$.

Let $X_{it}$ denote the production of item $i$ in period $t$ and $I_{it}$ represent the “planning” inventory of item $i$ at the end of $t$, arising because $X_{it}$ may be greater than $d_{it}$. Thus, $I_{it}$ is the accumulation of stock across periods to allow postponement of capacity purchases. Recalling that a period may be of fairly long duration (say, six months), there will be more than one set-up within a period, particularly for the high-demand items. Let $Q_{it}$ denote the lot size in which an item $i$ is produced within period $t$; the corresponding number of set-ups is then $X_{it}/Q_{it}$. Note that the number of set-ups, $X_{it}/Q_{it}$, may be fractional, which is clearly an approximation. This is reasonable, however, because the inaccuracy introduced by allowing fractional set-ups is significant only for smaller values of $X_{it}/Q_{it}$; this is true only for low-demand items, which have an insignificant impact on total costs. For similar reasons, we ignore any inventory at the end of a period due to fractional set-ups. The computational results presented in §5 verify that the inaccuracy due to this approximation is negligible.

The approach outlined above ignores backorders. Allowing backorders across periods is inappropriate as the length of each period may be as long as six months. We could potentially capture backorders in each period but then we have to address issues such as sequencing production of the products within a period as backorders are a function of the sequence. Sequencing is a level of detail that is inappropriate for ease of exposition. Constraint (2) is the demand balance equation for each item $i$ in each period $t$ and Constraint (3) is the capacity constraint in each period, which takes into account both processing and set-up times. Although the set-up time ($\beta_{it}$) and the processing time ($\alpha_{it}$) are assumed to be constant during the horizon, we can easily incorporate deterministic dynamically changing product-specific values in our approach. Constraint (4) is the capacity balance equation that tracks capacity levels in each period.

For ease of exposition, we ignore the impact of finite production rates on average within-period cycle stock $Q_{it}/2$ in the objective function. However, we can
easily account for finite production rates by multiplying $Q_{it}/2$ by the constant $1 - \alpha_i d_{it}/C_i$. It is clear that ignoring this adjustment is reasonable when the number of products is large as $C_i$ will be much greater than $\alpha_i d_{it}$.

4. Solution Approach

The formulation (P) is made difficult by the non-linear term in Constraint (3) and the presence of binary variables $Y_i$. To solve the problem, we develop a lower bound using Lagrangian relaxation of Constraint (3). An upper bound is obtained from a heuristic procedure that solves a dynamic program, with subproblems of the dynamic program solved using linear programming. Further, we also propose a simple approach to determine a feasible solution using the lower bound obtained at each iteration of the Lagrangian relaxation procedure.

4.1. Lower Bound

We propose the following Lagrangian relaxation (L) of formulation (P), with Constraints (3) relaxed using the Lagrange multipliers $\lambda$. Then, for a given $\lambda = \{\lambda_t, t = 1 \ldots T\}$, we have

$$Z_L = \min \sum_{t=1}^{T} \left\{ g_t Y_t + \sum_{i=1}^{M} \left[ h_{it} (I_{it} + Q_{it}/2) + p_{it} X_{it} \right] - \lambda_t \left[ C_t - \sum_{i=1}^{M} (\alpha_i X_{it} + \beta_i X_{it}/Q_{it}) \right] \right\},$$

subject to (2), (4), and (5).

Observe that $Q_{it}$ does not appear in the Constraints (2) or (4). Therefore, the optimal $Q_{it}$ for all $i$ and $t$ is given by solving

$$\min \sum_{i=1}^{M} \sum_{t=1}^{T} \left[ h_{it} Q_{it}/2 + \lambda_t \beta_i X_{it}/Q_{it} \right] \quad \text{subject to} \quad Q_{it} \geq 0.$$ 

Thus, the optimal $Q_{it}$ is given by

$$Q_{it} = \sqrt{2\lambda_t \beta_i X_{it}/h_{it}} \quad \forall \ i, t. \quad (6)$$

Replacing $Q_{it}$ using (6) in the Lagrangian formulation (L), simplifying the square root terms and rearranging all the terms, we have

$$Z_L = \min \sum_{i=1}^{T} \left\{ g_t Y_t - \lambda_t C_t + \sum_{i=1}^{M} \left[ h_{it} I_{it} + (p_{it} + \lambda_t \alpha_t) X_{it} \right] + \sqrt{2\lambda_t \beta_t X_{it}/h_{it}} \right\},$$

subject to (2), (4), and (5).

This Lagrangean problem (L) separates into the subproblems, (L1) and (L2).

Subproblem (L1) The subproblem (L1) is given as follows.

$$Z_{L1} = \min \sum_{i=1}^{T} (g_t Y_t - \lambda_t C_t),$$

$$C_t - C_{t-1} - b Y_t = 0 \quad \forall \ t, \quad (4)$$

$$Y_t \in \{0, 1\}, C_t \geq 0 \quad \forall \ i, t.$$

Subproblem (L1) can be solved easily using the following simple approach. Expressing $C_t$ in (4) in terms of the $Y_t$ values and recalling that $C_0 = B$, we have

$$C_t = C_0 + \sum_{\tau=1}^{t} b Y_\tau = B + \sum_{\tau=1}^{t} b Y_\tau \quad \forall \ t.$$

Substituting for $C_t$ in the objective function, Subproblem (L1) becomes

$$Z_{L1} = \min \sum_{i=1}^{T} \left( g_t Y_t - \lambda_t \left( B + \sum_{\tau=1}^{t} b Y_\tau \right) \right)$$

subject to $Y_t \in \{0, 1\} \quad \forall \ t.$

The nonnegativity constraints for $C_t$ are automatically satisfied by this solution since $B$, $b$, and $Y_t$ are nonnegative. Rearranging terms so as to combine all the terms containing $Y_t$ for any $t$, we get

$$Z_{L1} = \min \sum_{i=1}^{T} \left( Y_t \left( g_t - \sum_{\tau=1}^{t} b \lambda_\tau \right) - \lambda_t B \right)$$

subject to $Y_t \in \{0, 1\} \quad \forall \ t$.

While Subproblem (L1) can be solved easily using inspection, based on our computational experiments we found that it may not provide tight lower bounds.
So, we add surrogate constraints that are superfluous in the original problem but result in a tighter relaxation when added to the Subproblem (L1). The surrogate constraints ensure that the cumulative production capacity available in any set of periods \([1 \ldots t]\) where \(t = 1 \ldots T\) is greater than or equal to the cumulative production capacity required in those periods, which depends on the demands, processing times, and set-up times. Recall that \(C_t\) is the capacity level in period \(t\) after any capacity purchases in \(t\); \(C_t = B + \sum_{1 \leq s \leq t} bY_s\). Let \(P_t\) denote the cumulative production capacity available in periods 1 through \(t\). Then,

\[
P_t = \sum_{1 \leq k \leq t} C_k = \sum_{1 \leq k \leq t} \left(\sum_{1 \leq s \leq t} bY_s + B\right) = \sum_{1 \leq k \leq t} ((t - k + 1)bY_k + B)
\]

Let \(D_t\) be the cumulative production capacity required in Periods 1 through \(t\). The capacity required comprises of time required for processing and set-ups. However, the number of set-ups is itself a decision variable. So we simplify by requiring that there be at least one set-up of each product in a set of periods (in which there is demand for that product). This ensures that we do not eliminate any feasible solutions even though the surrogate constraint may not be tight. Then, the minimum cumulative capacity required in periods \([1 \ldots t]\) is given by:

\[
D_t = \sum_{i=1}^{M} \left(\beta_i + \sum_{1 \leq s \leq t} \alpha_i d_{it}\right)
\]

The subproblem (L1) with the surrogate constraints is then given by:

\[
Z_{L1} = \min \sum_{t=1}^{T} \left( Y_t \left( g_t - \sum_{t=1}^{T} \lambda t \right) - \lambda_i B \right), \quad \sum_{1 \leq k \leq t} ((t - k + 1)bY_k + B) \geq D_t, \quad \forall \ t; \quad \sum_{1 \leq k \leq T} bY_k \leq C_t, \quad \forall \ t; \quad Y_t \in [0, 1] \quad \forall \ t.
\]

At first glance, this appears to be a binary, multidimensional knapsack problem. However, we can exploit the structure of this Subproblem and solve it using a shortest path approach. In particular, we can reformulate (L1) as a dynamic program with state variables \(C_t\) and \(P_t\). The recursive equations are:

\[
F_t(C_t, P_t) = \begin{cases} 
\infty & \text{if } P_t < D_t, \\
0 & \text{if } P_t \geq D_t.
\end{cases}
\]

\[
F_t(C_t, P_t) = \begin{cases} 
\infty & \text{if } P_t < D_t, \\
\min \left\{ F_{t+1}(C_t, P_t + C_t) \right\} & \text{if } P_t \geq D_t,
\end{cases}
\]

In the above recursion, the first term within the minimization is the one where \(Y_t = 0\) and the second case is the one where \(Y_t = 1\). While the state variables \(C_t\), and \(P_t\) may appear large, only values that are multiples of \(b\) have to be considered where \(b\) is the capacity increment purchased. Also, the maximum value of the state variable \(C_t\) is equal to \(D_t\) and we can subtract the production capacity associated with the initial capacity from the capacity requirement for all periods and this will reduce \(D_t\) and the state space. Finally, there are only two alternatives at each stage. So this dynamic program can be solved efficiently using a shortest path algorithm wherein some of the paths are not feasible.

**Subproblem (L2).** The Subproblem (L2) is given by

\[
Z_{L2} = \min \sum_{t=1}^{T} \sum_{i=1}^{M} \left( h_{ilt} I_{ilt} + (p_{ilt} + \lambda_i \alpha_i) X_{ilt} + \sqrt{2\lambda_i \beta_i X_{ilt} h_{ilt}} \right), \quad X_{ilt} - I_{ilt} + I_{ilt-1} = d_{ilt} \quad \forall \ i, t,
\]

\[
X_{ilt} \geq 0, I_{ilt} \geq 0 \quad \forall \ i, t.
\]

Subproblem L2 splits into \(M\) subproblems corresponding to the \(M\) items, each of which is a single-item, uncapacitated lot-sizing problem that can be solved using efficient procedures such as the Wagner-Whitin algorithm (Wagner and Whitin 1958) or more efficient procedures developed recently (see van Hoesel and Wagelmans 1996). This is clear from the fact that we have an inventory balance equation (2) for each item and concave costs in the objective function.
Lagrange Multipliers. The Lagrange multipliers were updated using a subgradient approach wherein the subgradients for $\lambda_i$ are computed as follows. Let $X^*_i(k), Q^*_i(k)$ be the optimal solution to the Lagrangian problem in iteration $k$. The subgradient for $\lambda_i$ is

$$
\delta_i = C_i - \sum_{t=1}^{M}(\alpha_i X^*_i(k) + \beta_i X^*_i(k)/Q^*_i(k)),
$$

and the multipliers are updated as follows starting in Iteration 2

$$
\lambda^{k+1}_i = \lambda^k_i + \gamma_i \delta_i, \quad \text{where}
$$

$$
\gamma_i = \epsilon(Z^{UB} - Z_{LB})/\sum_{t=1}^{T} \delta_i^2,
$$

where $Z^{UB}$ and $Z_{LB}$ are the best upper and lower bounds at the end of iteration $k$. The initial values of $\lambda_i$ (in Iteration 1) are chosen as follows: $\lambda_0 = 0$ and $\lambda_i = (\lambda_{i-1} + \max_i(h_{i-1}/\alpha_i))/2$. Note that when $\lambda_i > (\lambda_{i-1} + \max_i(h_{i-1}/\alpha_i))$, the implied cost of producing in period $t$ is greater than the cost of producing in period $(t-1)$ and carrying inventory to period $t$, so the solution to the Lagrangian relaxation will be such that production always occurs in period $t$. Further, if this condition is true for a large number of periods then we end up with a lumpy initial solution. To avoid such solutions we start with $\lambda_i$ values that result in production being spread across periods. We start with an initial $\epsilon$ value of 2, and halve the value once in 10 iterations. We terminate the procedure after 300 iterations if the lower bound is not equal to the upper bound.

4.2. Heuristic Procedures

We present two heuristic procedures for obtaining an upper bound or a feasible solution to the problem. First, we describe the initial upper bound obtained before solving the Lagrangian relaxation discussed earlier. Then, we present a procedure for computing a feasible solution using the lower bound obtained at any iteration of the Lagrangian relaxation procedure.

4.2.1. DP-LP Heuristic. The basic idea of the heuristic procedure for determining the initial feasible solution, referred to as DP-LP, is as follows. We search for only those feasible solutions that have the following property: Capacity is purchased in a period only if there is no incoming inventory in that period. While it is not generally true that this is the case for an optimal solution to the problem, such a property is true for an optimal solution to simpler versions of this problem with one product and no lot sizing (Rao 1976).

Let $j$ and $k(j > j)$ be two periods such that capacity is purchased in periods $j$ and $(k+1)$ but not in any period between them. Then, the set of periods $(j \ldots k)$ is referred to as a capacity block. We then assume that the ending inventory in period $k$ is equal to zero. Then a recursive equation that expresses this approximation to the original problem is given by

$$
F(j, C_j) = \min_{k \geq j} \{g_j + K(j, k, C_j) + F(k + 1, C_j + b)\},
$$

where $F(j, C_j)$ is the cost of the best policy from period $j$ through $T$, given that we started period $j$ with capacity level $C_j$, which includes the capacity purchased in that period. The parameter $g_j$ is the cost of purchasing the capacity increment $b$ in period $j$ as defined earlier. $K(j, k, C_j)$ is the cost of an optimal production and inventory policy in periods $j$ through $k$, given that the capacity in periods $j$ through $k$ is equal to $C_j$. However, it is not easy to compute the optimal $K(j, k, C_j)$ as it is the solution to a multi-item lot-sizing problem with set-up times. We use a heuristic approach that first ignores the lot-sizing decisions and assumes that there is one set-up of each product in a period. Once the costs $K(j, k, C_j)$ are known for all $j$ and $k$, then we can solve the recursive equations to determine the set of periods in which to purchase capacity at the lowest cost.

In a capacity block $(j, k)$ the capacity in each period is equal to $C_j$, the demand of each product in each period is known and also by our assumptions, the beginning and ending inventory in each block is zero for all the products. Therefore, if we ignore lot sizing, then the problem of computing $K(j, k, C_j)$ can be formulated as

$$
K(j, k, C_j) = L(j, k) + \sum_{i=1}^{M} \sum_{l=i}^{k} h_{il} Q_{il}/2,
$$

where $L(j, k)$ is a linear program with the objective of minimizing inventory and production costs within
the capacity block subject to demand and capacity constraints.

\[
L(j, k) = \min \sum_{i=j}^{k} \sum_{t=1}^{M} h_{it} I_{it} + p_{it} X_{it},
\]

subject to

\[
X_{it} - I_{it} + I_{it-1} = d_{it} \quad \forall \ i, j \leq t \leq k,
\]

\[
\sum_{i=1}^{M} (\alpha_i X_{it} + \beta_i) \leq C_i, \quad j \leq t \leq k,
\]

\[
I_{jt-1} = 0, I_{tk} = 0 \quad \forall \ i,
\]

\[
X_{it} \geq 0, I_{it} \geq 0 \quad \forall \ i, j \leq t \leq k.
\]

Next, we check if \( Q_{it} \leq X_{it} \); otherwise we set \( Q_{it} \) equal to \( X_{it} \). We then check if the capacity constraints are satisfied, i.e.,

\[
\sum_{i=1}^{M} (\alpha_i X_{it} + \beta_i X_{it}/Q_{it}) \leq C_i \quad j \leq t \leq k.
\]

If the constraint is violated for any particular period \( t \), then we sequentially (indexed by \( i \)) set the value of \( Q_{it} = X_{it} \), one product at a time and continue until we have a feasible solution. Then,

\[
K(j, k, C_j) = L(j, k) + \sum_{i=1}^{M} \sum_{t=1}^{k} h_{it} Q_{it}/2.
\]

4.2.2. Lagrangian Upper Bound. At any iteration, the solution to the Lagrangian subproblem (L1) with the surrogate constraints provides a set of \( Y_i \) values that can provide a feasible solution. This in turn determines the capacity available in each period. Then we compute \( X_{it} \) and \( Q_{it} \) using the following greedy allocation procedure (see Appendix A for the detailed procedure). We go from Period 1 to \( T \) and within each period, we satisfy demand for the products, sorted in order of decreasing holding costs, from any remaining capacity in that or prior periods. The period chosen to produce and satisfy that demand is based on the marginal cost of capacity in that period (which comes from the Lagrange multiplier for the capacity constraint) and the cost of carrying inventory from the period of production to the period of demand. Note that this procedure is not guaranteed to produce a feasible solution since it allocates demand in a greedy manner going from Period 1 to \( T \). So, there may be a period with unsatisfied demand that could have been satisfied from capacity allocated to set-ups in a prior period.

Once the \( X_{it} \) values are obtained as above, then the feasible values of \( Q_{it} \) are obtained by first checking if \( Q_{it} > X_{it} \) and then correcting appropriately as in the procedure for computing the DP-LP upper bound. Then the capacity constraint in each period is checked and the \( Q_{it} \) values are modified as in the procedure for computing the DP-LP upper bound.

5. Computational Study

The computational study was performed with three objectives in mind: (1) to evaluate the performance
of the solution approach under various conditions; (2) to explore the impact of problem parameters on the inventory and capacity decisions; (3) to test the applicability of the solution procedure for industry size problems. We first describe the experimental set-up and then discuss the performance of the solution approaches. Subsequently we provide insights on the effect of problem parameters and finally discuss the industry application.

Problem Parameters. We consider two sets of problems in the primary study which are used to evaluate the performance of the heuristic relative to the lower bound. Each set of problems has 10 products and 16 periods, with each period assumed to be six months. One set (called symmetric) had identical demand for all the products in a period and in the second set (called asymmetric), the products have different demands with the highest- to lowest-demand values in any period having a ratio of about 10:1. We also solve some additional problems to study the impact of product variety, set-up times, and holding costs. Further, we also test the impact of the approximation of fractional set-ups for these problems.

To determine the parameters to use in the study, we started with data at the firm that motivated this study. We then introduced sufficient variation in the data sets to evaluate the solution approach under different conditions and to study the impact of problem parameters on capacity and inventory decisions. The primary computational study uses the following sets of values for the various parameters in the model.

- Demand. \( d_{i,t} = d_{i,t-1} \times (1 + 0.1 \times \text{rand}()) \) where \( \text{rand}() \) is a random number between 0 and 1. So, demand grows by 0%-10% each period (average 5%) with \( d_{i,0} = 500 \) for all \( i \) in the symmetric case and \( d_{i,0} = 500 \times i \) for \( i = 1 \ldots 10 \) in the asymmetric case.
- Unit processing time. \( \alpha_i = 10 \forall i, t \).
- Set-up time. Three different values were considered in the primary study, \( \beta_i = 150, 300, 450 \) and were assumed to be identical for all products. The ratios of set-up time to processing time therefore varies from 15 to 45. In the industry size problems, we considered a higher set-up to processing time ratio of 120.
- Unit inventory carrying cost per period. \( h_i = 0.5 \times 10 \times (k + c) \) where \( c \) is the annual cost of capital, \( k \) represents the annual non-capital related carrying charges expressed as a percentage of the item cost and 10 is the cost per unit, or cost of goods sold per unit, for all items and the factor 0.5 is present because each period is six months. In the primary study, we set \( k = 0.1 \).
  - Cost of capital \( c \). Three different values were considered: 0.1, 0.2, and 0.3.
  - Initial capacity. \( B = m \times (\sum (\alpha \times d_{i,0} + 10 \times 10 \times \beta_i)) \). The initial capacity at the beginning of Period 1 is equal to a multiple \( (m) \) of the total time required for processing all the demand in the first period plus the time required for 10 set-ups of each product (note that there are 10 products). Thus, \( m \) controls the level of initial capacity and whether capacity has to be purchased in the initial periods. We consider three values for \( m: 0.9, 1.2, \) and 1.5. When \( m = 0.9 \), initial capacity is low enough that capacity will have to be purchased early, while \( m = 1.5 \) implies a scenario where initial capacity is high enough that capacity may never be purchased, depending on other parameters. The multiple values for \( m \) allow for varying numbers of set-ups and ratio of time spent in set-ups to time spent processing.
  - Capacity size \( b \). Three different values were considered—B/6, B/3, and B/2, where B is the initial capacity. Since there were currently six machines at the firm, we considered the value of B/6 as the capacity for one machine. Also, given the average 5% increase in demand, the B/6 case represents a scenario where a new machine is sufficient to meet the demand growth for 3–4 periods. The other values B/3 and B/2 were used to model higher-capacity machines.
  - Capacity purchase cost. \( g_i = g_{i-1} \times (1 - 0.5 \times c) \) where \( c \) is the cost of capital and \( g_0 = b \times f \) where \( b \) is the capacity size and \( f \) is the capacity purchase cost per unit time on the machine that can take four different values in the study—0.75, 1.5, 3, 4.5. So, if \( f = 4.5 \), the implication is that the cost of purchasing one unit of capacity on the machine is 4.5, where one unit of capacity is expressed in units of time available on the machine per period. Given the processing time per unit of 10, the cost of purchasing capacity to make a unit of output per period in this case will be 4.5 \times 10 = 45. While a cost of 4.5 per unit of capacity may appear small, it represents the cost that enables production of one tenth additional unit of a product over a six-month period.
Thus, we solved 324 ($3 \times 3 \times 3 \times 3 \times 4$) problems in each of the symmetric and asymmetric cases for a total of 648 problems in the primary study. Since the unit production cost $p_{it}$ was assumed to be the same in all periods, we have excluded the production costs while reporting our results.

5.1. Performance of Heuristics

Table 1 provides results on the performance of the heuristics across the 648 problems. Duality gap, the % difference between the heuristic solution value and the lower bound solution value, was used to evaluate the heuristics. The DP-LP heuristic performed on average better than the Lagrangian heuristic. While the average duality gap of the DP-LP heuristic across the 648 problems was 2.55%, the average duality gap of the Lagrangian heuristic was 3.04%. More significant, in only 11 out of 324 symmetric problems did the DP-LP heuristic have a duality gap greater than 5%; the Lagrangian heuristic had 26 cases. In the asymmetric case, the numbers were 65 and 75, respectively, for the DP-LP and Lagrangian heuristics.

Table 2 provides details about the performance of the DP-LP heuristic as a function of the problem parameters. From the table, it is clear that the performance of the heuristic deteriorates with an increase in the unit capacity purchase cost ($f$), and an increase in the capacity size ($b$). The performance of the heuristic improves with increase in initial capacity multiple ($m$) in the symmetric case, with increase in cost of capital ($c$) in the asymmetric case and with increase in set-up time ($\beta$) under both symmetric and asymmetric cases.

We also evaluated the speed and accuracy of our methods on 32 large industry size problems (with 100 products and 16 time periods). We were able to solve these large problems in a reasonable amount of time (around 30 minutes on average on a Sun ULTRA-1 workstation) and the duality gaps were quite small (0.21%–0.3%) on average.

One of the approximations in our model and solution approach relates to the fact that the number of set-ups ($X_{it}/Q_{it}$) is allowed to be fractional, a deviation from reality. We wanted to check how adapting our solution to the real environment (by rounding down the number of set-ups to integer values) changed the total costs and lot-size-related inventory costs. To do that, we took the solution from the upper bound and then increased $Q_{it}$ so that the number of set-ups are rounded down to the nearest integer. Based on 84 sample problems with seven different values for the number of products (ranging from $N = 10$ to $N = 100$) and 12 different sets of problem parameters we find that the error in total cost

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Duality Gaps for the Two Heuristics Across 646 problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem Type</td>
<td>Performance Measure</td>
</tr>
<tr>
<td>Symmetric</td>
<td>Average Duality Gap</td>
</tr>
<tr>
<td>No. of Problems with &gt;5% Gap</td>
<td>26</td>
</tr>
<tr>
<td>Asymmetric</td>
<td>Average Duality Gap</td>
</tr>
<tr>
<td>No. of Problems with &gt;5% Gap</td>
<td>75</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Duality Gap for DP-LP Under Different Problem Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem Parameter Values</td>
<td>Symmetric</td>
</tr>
<tr>
<td>Unit Capacity</td>
<td>$f = 0.75$</td>
</tr>
<tr>
<td>Purchase Cost ($f$)</td>
<td>$f = 1.5$</td>
</tr>
<tr>
<td></td>
<td>$f = 3$</td>
</tr>
<tr>
<td></td>
<td>$f = 4.5$</td>
</tr>
<tr>
<td>Cost of Capital ($c$)</td>
<td>$c = 0.1$</td>
</tr>
<tr>
<td></td>
<td>$c = 0.2$</td>
</tr>
<tr>
<td></td>
<td>$c = 0.3$</td>
</tr>
<tr>
<td>Set-up Time ($\beta$)</td>
<td>$\beta = 150$</td>
</tr>
<tr>
<td></td>
<td>$\beta = 300$</td>
</tr>
<tr>
<td></td>
<td>$\beta = 450$</td>
</tr>
<tr>
<td>Capacity Size ($b$)</td>
<td>$b = B/6$</td>
</tr>
<tr>
<td></td>
<td>$b = B/3$</td>
</tr>
<tr>
<td></td>
<td>$b = B/2$</td>
</tr>
<tr>
<td>Initial Capacity</td>
<td>$m = 0.9$</td>
</tr>
<tr>
<td>Multiple ($m$)</td>
<td>$m = 1.2$</td>
</tr>
<tr>
<td></td>
<td>$m = 1.5$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Increase in Costs with Integral Number of Set-ups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Products</td>
<td>Error in Cycle Stock Costs</td>
</tr>
<tr>
<td>10</td>
<td>11.3%</td>
</tr>
<tr>
<td>15</td>
<td>11.1%</td>
</tr>
<tr>
<td>20</td>
<td>7.9%</td>
</tr>
<tr>
<td>25</td>
<td>7.5%</td>
</tr>
<tr>
<td>30</td>
<td>7.2%</td>
</tr>
<tr>
<td>50</td>
<td>4.2%</td>
</tr>
<tr>
<td>100</td>
<td>4.1%</td>
</tr>
<tr>
<td>Average</td>
<td>7.6%</td>
</tr>
</tbody>
</table>
is quite small, especially when the number of products is large (see Table 3). In fact, for industry size problems with 100 products, the error in total cost is only 0.3% while the error in cycle stock cost is 4.1%. The errors decrease with an increase in the number of products because the impact of an error in any one item is lower on total costs. This is especially true because the error is most significant for the low-demand items. The results of the study help justify the approximation made in the paper.

5.2. Impact of Parameters on Decisions and Costs

We performed an analysis of the behavior of total costs, capacity purchase costs, and inventory carrying costs and capacity purchase and inventory decisions for different values of the cost of capital \( c \), capacity size \( b \), initial capacity multiple \( m \), holding cost \( h \), set-up time \( \beta \), number of products \( M \) and capacity purchase cost multiple \( f \). This analysis was based on the best heuristic solution for a problem. The insights based on this analysis are valuable but need not be true under all conditions for two reasons. First, the heuristic solution may be suboptimal and so the optimal solution may show different behavior.

Second, the insights are specific to the set of problem parameters used in our study and may not apply for an alternative set of parameters. However, with that caveat, we do find several interesting insights. To keep the exposition concise, we have not reported the results for all sets of parameters in the tables but only a representative set. However, similar patterns were observed in the other cases as well.

**Cost of Capital.** Interestingly, both total cost and capacity purchase cost decline with an increase in the cost of capital \( c \) and more capacity is purchased (i.e., in more periods), especially in later periods (see Table 4). This is primarily because an increase in the cost of capital makes capacity purchases in later periods cheaper. At the same time, an increase in the cost of capital increases unit inventory carrying costs. So, it is better to buy more capacity in later periods to satisfy demand rather than carry inventory from earlier periods. We do find that inventory carried declines in all cases, due to a decrease in both \( I_n \) and \( Q_n \). Total inventory costs also decline in most cases; however, total inventory costs increase in some cases with an increase in the cost of capital because unit inventory carrying cost is higher.
Table 6 Effect of Set-up Time (β) on Costs and Capacity Decisions

<table>
<thead>
<tr>
<th>Set-up Time (β)</th>
<th>Purchase Periods</th>
<th>Purchase Cost</th>
<th>Inventory Cost</th>
<th>Cycle Stock</th>
<th>Total Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>150</td>
<td>1,6,7,10,11,13,14</td>
<td>166,766</td>
<td>5,961</td>
<td>8,456</td>
<td>181,185</td>
</tr>
<tr>
<td>210</td>
<td>1,6,7,10,11,13,14</td>
<td>166,766</td>
<td>5,961</td>
<td>11,670</td>
<td>184,398</td>
</tr>
<tr>
<td>270</td>
<td>1,6,7,10,11,13,14</td>
<td>166,766</td>
<td>5,961</td>
<td>14,883</td>
<td>187,612</td>
</tr>
<tr>
<td>330</td>
<td>1,5,8,9,11,13,14</td>
<td>170,922</td>
<td>4,711</td>
<td>15,113</td>
<td>190,747</td>
</tr>
<tr>
<td>390</td>
<td>1,5,7,9,11,13,14</td>
<td>174,677</td>
<td>4,154</td>
<td>14,384</td>
<td>193,215</td>
</tr>
</tbody>
</table>

Capacity Size. An increase in the capacity size (b) generally results in an increase in the capacity costs and total costs (see Table 5). Increase in capacity size also results in capacity being purchased in fewer periods. The total capacity costs may sometimes decline with an increase in capacity size. In such cases, more inventories may be carried and inventory carrying cost increase. However, this may be offset by the decrease in capacity costs due to fewer capacity purchases, resulting in a decline in total costs when the cost of capital is low (see the case where c = 0.1 in Table 5).

Set-up Time. Our parameters are determined in a manner that the initial capacity increases when set-up time is increased. To isolate the impact of set-up time, we solved several problems (symmetric case) where the set-up time β was varied but all the other parameters were kept constant including initial capacity (see Table 6). An increase in set-up time does not have a major impact on total costs. For instance, even when set-up time is doubled, total cost increases only by about 5%. (Recall that the total cost excludes production costs; including this cost will only decrease this percentage.) This is because even though total capacity purchased may not increase, capacity is purchased a little earlier. This increases the capacity available in each period and so inventory carried from one period to the next declines. Lot sizes may go up and so inventory carried within a period increases. Thus, purchasing capacity earlier offsets the effect of a substantial increase in set-up times on total costs.

Holding Cost. The holding cost changes with a change in cost of capital but so does the capacity purchase cost in each period. To isolate the impact of holding cost, we solved several problems where the noncapital related carrying charge k was varied (from 0.1 to 0.5) but all other parameters were kept constant (see Table 7). Again, we find that doubling the unit holding cost results only in a 5% increase in the total cost because purchasing capacity earlier in a few periods offsets the impact of an increase in unit inventory holding cost. This increases the capacity available in each period and reduces the inventory carried between periods.

Capacity Purchase Cost. An increase in the capacity purchase cost multiple (f), which implies higher unit capacity costs, results in either less capacity being purchased, i.e., capacity is purchased in a fewer number of periods, or in capacity being purchased later (see Table 8). Despite this, total capacity costs increase. Also, when less capacity is purchased, more demand is satisfied by carrying inventory from earlier periods (increase in Ii) and there are fewer set-ups in each period resulting in larger batch sizes (Qi). So, inventory-carrying costs are also higher and thus total costs are also higher with an increase in the capacity purchase cost multiple (f).

Ratio of Capacity Purchase Cost to Holding Cost. We also evaluated the impact of different ratios of unit capacity purchase cost to unit holding cost (f/h) on capacity, inventory, and lot-sizing decisions. h was

Table 7 Effect of Holding Cost Factor (k) on Costs and Capacity Decisions

<table>
<thead>
<tr>
<th>Holding Cost Factor (k)</th>
<th>Purchase Periods</th>
<th>Capacity Purchase Cost</th>
<th>Inventory Cost</th>
<th>Cycle Stock</th>
<th>Total Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>1,6,7,10,11,13,14</td>
<td>166,766</td>
<td>5,961</td>
<td>8,456</td>
<td>181,185</td>
</tr>
<tr>
<td>0.2</td>
<td>1,6,7,10,11,13,14</td>
<td>166,766</td>
<td>7,452</td>
<td>10,570</td>
<td>184,789</td>
</tr>
<tr>
<td>0.3</td>
<td>1,5,8,9,11,13,14</td>
<td>170,922</td>
<td>7,067</td>
<td>10,304</td>
<td>188,294</td>
</tr>
<tr>
<td>0.4</td>
<td>1,5,8,9,11,13,14</td>
<td>170,922</td>
<td>8,245</td>
<td>12,022</td>
<td>191,189</td>
</tr>
<tr>
<td>0.5</td>
<td>1,5,7,9,11,13,14</td>
<td>174,677</td>
<td>8,309</td>
<td>11,064</td>
<td>194,050</td>
</tr>
</tbody>
</table>
tions, we considered a case where the number of products increases from 5 to 25, in steps of 5. We considered the case where the number of products is 20. We find that an increase in the number of products results in one or more of the following. (1) Less capacity is purchased; this is not surprising because, as the number of products increases, buying capacity is more expensive than carrying inventory. (2) Capacity is generally purchased a little later in many cases because buying later reduces the total capacity cost due to the impact of the cost of capital. (3) We also find that the "planning" inventory increases as the number of products increases. This seems reasonable due to the lower relative inventory holding cost implied by a higher product variety. However, there are two other factors that work. Since less capacity is purchased, more inventory is likely to be carried across periods. On the other hand, as capacity is purchased later, there is less scope for producing early using excess capacity and carrying inventory across periods. It appears that the net effect of all these factors is an increase in the inventory carried across periods. (4) Finally, we find that lot sizes and corresponding cycle stock costs are higher. As the number of products increases, it is not optimal to use capacity to perform many setups and reduce lot sizes as the cost of capacity is high relative to the cost of carrying inventory.

**Product Variety.** We analyzed the impact of an increase in product variety on total costs over time by solving problems where the number of products increases from 5 to 25, in steps of 5. We considered two possible cases here. First, we considered a case where total demand increases proportionally with an increase in variety. Not surprisingly, we found that total costs over time increased proportionally as variety and demand increased. Second, we considered the case where variety increases but total demand does not increase, which is true in mature industries where firms increase variety to retain market share. The results in this case, reported in Table 10, are more interesting. Doubling variety increased total cost only by about 10%–25% and even a five-fold increase in variety increased total cost only by about 50% and cycle stocks only doubled. This is primarily because as variety increases dramatic increases in batch sizes and cycle stocks are restrained by purchasing capacity earlier. As a result, more capacity becomes available which can be used to do more set-ups and thus keep batch sizes smaller or to build up "planning" inventory for later periods. This inventory build-up helps in decreasing the amount of production in later periods thus providing capacity for more set-ups. It is interesting to note that increase in product variety need not always result in higher inventories and costs because of the effect of capacity purchases. Thus, we see a significant value in coordinating capacity and demand.

**Table 8** Effect of Unit Capacity Purchase Cost \((f)\) on Costs and Capacity Decisions

<table>
<thead>
<tr>
<th>Unit Purchase Cost ((f))</th>
<th>Periods</th>
<th>Capacity Purchase Cost</th>
<th>Inventory Cost ((I_E))</th>
<th>Cycle Stock Cost ((Q_f))</th>
<th>Total Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.75</td>
<td>5, 6, 10, 12, 14</td>
<td>9,963</td>
<td>54</td>
<td>2,266</td>
<td>14,196</td>
</tr>
<tr>
<td>1.5</td>
<td>6, 8, 10, 13, 15</td>
<td>18,148</td>
<td>139</td>
<td>5,622</td>
<td>23,910</td>
</tr>
<tr>
<td>3.0</td>
<td>7, 8, 11, 13, 15</td>
<td>33,333</td>
<td>702</td>
<td>7,166</td>
<td>41,202</td>
</tr>
<tr>
<td>4.5</td>
<td>7, 8, 12, 13</td>
<td>58,928</td>
<td>870</td>
<td>13,620</td>
<td>72,052</td>
</tr>
</tbody>
</table>

**Table 9** Effect of the Ratio of Unit Capacity Purchase Cost \((f)\) and the Holding Cost \((h)\) on Costs and Capacity Decisions

<table>
<thead>
<tr>
<th>Capacity Ratio</th>
<th>Periods</th>
<th>Capacity Purchase Cost</th>
<th>Inventory Cost ((I_E))</th>
<th>Cycle Stock Cost ((Q_f))</th>
<th>Total Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1, 5, 8, 10, 12, 13, 15</td>
<td>81,683</td>
<td>0</td>
<td>12,513</td>
<td>94,193</td>
</tr>
<tr>
<td>10</td>
<td>1, 5, 8, 10, 12, 14, 16</td>
<td>155,807</td>
<td>247</td>
<td>17,533</td>
<td>173,589</td>
</tr>
<tr>
<td>15</td>
<td>1, 7, 8, 10, 12, 14</td>
<td>218,543</td>
<td>1,860</td>
<td>26,575</td>
<td>246,980</td>
</tr>
<tr>
<td>20</td>
<td>1, 7, 8, 11, 12, 14</td>
<td>286,826</td>
<td>2,152</td>
<td>30,486</td>
<td>319,465</td>
</tr>
<tr>
<td>30</td>
<td>1, 7, 8, 11, 12, 14</td>
<td>430,239</td>
<td>2,152</td>
<td>30,486</td>
<td>462,879</td>
</tr>
</tbody>
</table>

**Table 10** Effect of Increasing Product Variety on Costs and Capacity Decisions

<table>
<thead>
<tr>
<th>Number of Products ((M))</th>
<th>Periods</th>
<th>Capacity Purchase Cost</th>
<th>Inventory Cost ((I_E))</th>
<th>Cycle Stock Cost ((Q_f))</th>
<th>Total Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>8, 10, 13</td>
<td>10,937</td>
<td>485</td>
<td>4,231</td>
<td>15,654</td>
</tr>
<tr>
<td>10</td>
<td>8, 9, 12, 14</td>
<td>13,880</td>
<td>549</td>
<td>5,101</td>
<td>19,531</td>
</tr>
<tr>
<td>15</td>
<td>7, 9, 12, 14</td>
<td>14,771</td>
<td>72</td>
<td>4,937</td>
<td>19,731</td>
</tr>
<tr>
<td>20</td>
<td>7, 9, 12, 14</td>
<td>14,771</td>
<td>70</td>
<td>6,392</td>
<td>21,236</td>
</tr>
<tr>
<td>25</td>
<td>7, 9, 12, 14</td>
<td>14,771</td>
<td>406</td>
<td>8,058</td>
<td>23,236</td>
</tr>
</tbody>
</table>
inventory decisions, especially in environments where variety increases over time.

5.3. An Illustrative Application
We now illustrate the application of our model to the problem faced by the manufacturer of consumer products mentioned at the beginning of this paper. The firm produced 208 products with demands ranging from 10 to 1,460 units per day in a plant located in Ireland. All the 208 products belonged to the same product group and were produced on a machine (more accurately a production line) with a changeover required between the production of different products due to mold changes. This was a capital-intensive process and purchasing and installing a production line was estimated to cost $10.2 million. The initial capacity in the plant was six machines (or production lines) with each machine requiring a changeover time of 30 minutes and a processing time of 0.25 minutes per unit independent of the product. Thus, the maximum production capacity of a machine was about 2.1 million units per year, assuming a 24-hour operation and ignoring changeover times. The average cost of goods sold per unit was $9 and the inventory carrying cost was taken to be $2.70 per unit per year (30% of the unit cost). The company provided the above information and we also collected detailed information about demand forecasts for five years. We then extrapolated the demand forecasts for another three years for a total of eight years or 16 six-month periods. We ignored production costs, as they were not expected to change over time.

The firm was currently assessing whether to buy another production line now or to buy it later. There appeared to be sufficient capacity to meet demand for the next few years without increasing lot sizes substantially. Current capacity was sufficiently high so that they could perform on average five changeovers for each product every month. Of course, by optimizing the use of setups among the products, they could achieve low cycle stocks. So, the finance group was against buying any capacity and felt that it may be possible to not buy any production line at all over the next five or more years. This would imply no large capital investments. However, the production-planning group that was in charge of capacity planning had a different perspective. They felt that they should buy a machine now to build sufficient capacity as they had experienced serious capacity problems in the past when demand had grown and it was their large inventories that had saved them. Recently, there was pressure from top management to keep inventories lean. We tested both strategies using the model by setting \( Y_1 = 1 \) to represent the production planning group’s perspective and setting \( Y_t = 0 \) for all \( t \) to represent the finance group’s perspective. We compared these strategies with the best solution of the two heuristics.

The differences between the alternative strategies proposed by the firm’s employees that sound quite reasonable illustrate the value of the modeling approach. The production-planning group’s approach (called Strategy-1 in Table 11) involved purchasing capacity in the first period and carrying some planning inventory between periods. The finance group’s strategy (called Strategy-3 in Table 11) involved purchasing no capacity at all but required carrying substantial planning inventory from the initial periods to later ones while at the same time carrying very little cycle stock. While the strategy of not buying any capacity was indeed feasible, the cost of this strategy was almost 40% higher than Strategy-1 because of the high levels of inventories. The strategy proposed by the heuristic solution (called Strategy-2 in Table 11), whose cost was 1.5%, or $2 million, lower than Strategy-1, delayed the purchase of capacity to

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Purchase Periods</th>
<th>Purchase Cost</th>
<th>Inventory Cost ( I_t )</th>
<th>Cycle Stock Cost ( Q_t )</th>
<th>Total Cost (Excluding Production Cost)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategy-1</td>
<td>1</td>
<td>10,200,000</td>
<td>135,563,843</td>
<td>7,796,002</td>
<td>153,559,846</td>
</tr>
<tr>
<td>Strategy-2</td>
<td>6.7</td>
<td>17,749,672</td>
<td>127,374,878</td>
<td>6,088,538</td>
<td>151,213,088</td>
</tr>
<tr>
<td>Strategy-3</td>
<td>—</td>
<td>0</td>
<td>199,458,294</td>
<td>290,102</td>
<td>199,748,396</td>
</tr>
</tbody>
</table>
the sixth period (third year). The initial excess capacity was used to build up inventory that allowed the purchase of the machine to be delayed until Period 6. In turn, this lowered the discounted cost of purchase and so more capacity was purchased in Period 7. The excess capacity allowed a production plan that carried lower inventory in later periods and at the same time had a lower amount of cycle stock than Strategy-1.

Our model and solution approach not only illustrated a third and more efficient strategy for capacity and inventory planning but also illustrated the significant difference in performance (up to 40% in costs) that could arise from two equally reasonable strategies, thereby clearly demonstrating the need for a model and efficient solution methodology for this problem.

6. Variants

In this section we discuss two variants of the basic model: (1) continuous capacity purchase; and (2) multiple unit capacity purchase. These variants are motivated by settings where capacity purchases may have the above characteristics. We show how the bounds and heuristics derived in §3 extend to these settings.

6.1. Continuous Capacity Increments

In this setting, we assume that capacity can be bought in any amount. However, the cost of purchasing capacity consists of a fixed and a variable cost. Let \( z_t \) be the capacity added in period \( t \) and \( y_t \) represent the 0-1 decision variable that indicates if any capacity was purchased in period \( t \). The variable cost of capacity per unit is given by \( f_t \) and the fixed cost is given by \( g_t \). Then the problem (V1) can be formulated as follows.

\[
Z_{V1} = \min \sum_{t=1}^{T} \left( g_t y_t + f_t z_t \right) + \sum_{i=1}^{M} \left( h_i (I_{it} + Q_{it}/2) + p_i X_{it} \right)
\]

subject to

\[
X_{it} - I_{it} + I_{it-1} = d_{it} \quad \forall \, i, t,
\]

\[
\sum_{i=1}^{M} (a_i X_{it} + \beta_i X_{it}/Q_{it}) \leq C_t \quad \forall \, t,
\]

\[
C_t - C_{t-1} - z_t = 0 \quad \forall \, t,
\]

\[
z_t \leq L y_t \quad \forall \, t,
\]

\[
y_t \in \{0, 1\}, \quad z_t \geq 0, \quad X_{it} \geq 0,
\]

\[
I_{it} \geq 0, \quad Q_{it} \geq 0, \quad C_t \geq 0 \quad \forall \, i, t,
\]

where \( L \) is a large number.

**Lower Bound.** A Lagrangian relaxation of the above formulation relaxing the capacity constraint using multipliers \( \lambda_t \) will result in the subproblems (L1-V1) and (L2-V1) as in §4.1. Subproblem (L2-V1) is identical to subproblem (L2) so we will not discuss it further. Subproblem (L1-V1) is given as follows.

\[
Z_{L1-V1} = \min \sum_{t=1}^{T} \left( g_t y_t + f_t z_t - \lambda_t C_t \right)
\]

\[
C_t - C_{t-1} - z_t = 0 \quad \forall \, t
\]

\[
z_t \leq L y_t \quad \forall \, t
\]

\[
y_t \in \{0, 1\}, \quad z_t \geq 0 \quad \forall \, t
\]

As in §4.1, the problem can be reformulated as follows:

\[
Z_{L1-V1} = \min \sum_{t=1}^{T} \left( g_t y_t + z_t \left( f_t - \sum_{\tau=t}^{T} \lambda_{\tau} \right) - \lambda_t B \right)
\]

subject to \( z_t \leq L y_t \) \( \forall \, t \)

In the above formulation if \( \{ f_t - \sum_{\tau=t}^{T} \lambda_{\tau} \} < 0 \) for any \( t \) then it is optimal to set \( z_t = \infty \) and the lower bound will not be of much use. As a result, we need to ensure that the Lagrange multipliers are chosen so that \( \{ f_t - \sum_{\tau=t}^{T} \lambda_{\tau} \} \geq 0 \) \( \forall \, t \). This can be made sure by recursively setting upper bounds for \( \lambda_t \) as follows,

\[
\lambda_T < f_T
\]

\[
\lambda_{T-i} < f_{T-i} - \sum_{j=0}^{i-1} \lambda_{T-j}
\]

Surrogate constraints similar to those described in §4.1 are added to the formulation to get better bounds. The resulting problem is similar to the subproblem (L1) with surrogate constraints considered in §4.1 but differs in that in each period we not only have to decide whether capacity is purchased or not but also how much. So, we can use a similar shortest path approach. The resulting dynamic program is simplified to an extent in that there are only a limited set of choices for capacity addition in each period. For example, if a capacity purchase is made in period \( i \) then possible values for the capacity addition are such that one purchases enough to satisfy the requirements from \( i \) to \( j \) where \( j = i + 1, \ldots, T \). In addition, the amount to be purchased can also be computed
The main steps of the heuristic are as follows. Let $D_t$ be the incremental demand for product $i$ in period $t$ so, $i_{d_{it}} = d_{it}$ for $t = 1$ and $i_{d_{it}} = d_{it} - d_{it-1}$ for $1 < t \leq T$

Step 1. The solution to the Lagrangean subproblem (L1-V1) provides a set of capacity purchase periods. Among these capacity purchase periods, identify for each $i_{d_{it}}$ the best period to purchase that capacity while taking into account both the holding cost (due to earlier production) as well as the variable capacity purchase cost. For example, if $id_{i3} = 20$, there is an increase of demand of 20 units in the third period which needs to be satisfied in all periods starting in Period 3 (since demand is increasing over time). Now, capacity for this demand increment can be bought in Period 1, Period 2, or Period 3 (at the latest). Let us assume there are 10 periods in the horizon. Therefore, total capacity requirement for this increment is $20 \ast 8 = 160$ units. If capacity is purchased in Period 1 then this capacity is available in every period so one can buy 16 units of capacity and keep producing 16 units every period to satisfy the demand. In the process, one ends up carrying inventory to meet this incremental demand in Periods 1 through 10. On the other hand, one could purchase 20 units of capacity in Period 3 and end up not carrying any inventory at all. The heuristic selects the best prior period to purchase capacity based on inventory and variable capacity purchase costs.

Step 2. Compute total capacity to be purchased in each purchase period, based on the demand increment assignments made in Step 1 and taking into account only the processing time requirements. Then compute the $X_{it}$ and $I_{it}$ values based on the decisions in Step 1.

Step 3. Compute $Q_{it}$ values using the Lagrange multipliers as

$$Q_{it} = \sqrt{2\lambda_i \beta_i X_{it} / h_{it}} \quad \forall i, t$$

We check if $Q_{it} \leq X_{it}$, otherwise we set $Q_{it}$ equal to $X_{it}$.

Step 4. Determine the additional capacity to be purchased in the closest prior feasible period to meet the set-up time requirements implied by the $Q_{it}$ values computed in Step 3. Excess capacity in any period, created by these additional capacity purchases, is completely utilized by revising,

$$Q_{it} = \frac{\sum_{i=1}^{M} \sqrt{X_{it} h_{ii} \beta_i}}{h_{ii} (C_i - \sum_{i=1}^{M} \alpha_i X_{it})^2}.$$
We conducted a computational analysis to check the effectiveness of this heuristic. We found that the average duality gap across 81 problems in the asymmetric case was 6.58% showing that the above heuristic is actually quite effective for the continuous case. Further, we found that the gap increased with an increase in unit capacity cost as well as the initial capacity available and was relatively insensitive to changes in holding cost (see Table 12).

### 6.2. Capacity Addition in Multiples of Unit Capacity Size

In this variant, capacity can be added in multiples of the unit capacity $b$ in any period. This implies that in the problem formulation $P$ (discussed in §3), $Y_t \in \{0, 1\}$ is replaced by $Y_t \in I$ where $I$ refers to the set of nonnegative integers.

**Lower Bound.** We can obtain the lower bound by relaxing the capacity constraint. Subproblem (L2) is unchanged (as compared to §4.1) and Subproblem (L1) is changed to reflect that $Y_t$ is a nonnegative integer rather than being $0–1$. Again surrogate constraints can be added to tighten the formulation and a dynamic programming approach can be extended to this case as well. In particular, we will have the following recursive equations.

$$F_t(C_t, P_t) = \infty \quad \text{if } P_t < D_t,$$

$$F_t(C_t, P_t) = \min_{k=0}^{\infty} \left\{ F_{t+1}(C_t + kb, P_t + C_t + kb) + k_t \left( g_t - \sum_{t=1}^{T} b\lambda_t \right) \right\} \quad \text{if } P_t \geq D_t,$$

where $k_t$ is the number of units of equipment purchased in $t$. Note that the only difference between the above formulation and the one in §4.1 is that we have to consider more possible alternatives in the dynamic programming recursion here. In the above formulation, we have used an upper limit $L$ for the value of $k_t$. This limit $L$ can be based on the maximum total demand for capacity in periods $t$ through $T$ and can be easily computed. Clearly, the computational effort for the lower bound is higher in this case.

**Upper Bound.** The DP-LP and Lagrangean heuristics similar to those in §4.2 are applicable for this variant as well.

### 7. Conclusions

In this paper we explore the interaction between production-planning decisions and capacity acquisition decisions in environments with demand growth. This work was motivated by our interactions with a large firm in the consumer products industry. We present a model that considers a firm producing multiple items in a multiperiod framework. Demand for these items is known but varies over time with long-term growth and possible short-term fluctuations (say, due to seasonality). The production equipment is characterized by significant changeover time between the production of consecutive items. While demand growth is gradual, capacity additions are typically discrete. Therefore, periods immediately following a machine purchase are characterized by excess machine capacity. A production-planning perspective in the conventional operations management literature would suggest using excess capacity in a period to do more equipment changeovers and thus reduce inventories. On the other hand, a longer-term capacity acquisition perspective may suggest building additional inventory with that excess capacity to meet demand growth in future periods and thus delay the purchase of additional capacity in the future. This particular trade-off is faced by managers in several firms in the growth stage but there is a lack of models in the literature that address these concerns.
The mathematical programming model and solution approach presented in this paper can be considered as initial work in this area of integrating capacity acquisition and production planning. The integral nature of capacity acquisition and the nonlinearity in lot sizing constraints make this problem difficult to solve to optimality. We develop effective lower bounds (using Lagrangean relaxation) and efficient heuristics (which are less than 5% away from the lower bound). Through a computational study, we show the effectiveness of the solution approach in terms of solution quality and investigate the impact of product variety, cost of capital, and other important parameters on the capacity and inventory decisions. The computational results drive home some key insights such as—increasing product variety may not result in excessive inventory and even a substantial increase in set-up times or holding costs may not increase the total cost over the horizon in a significant manner. These results highlight the fact that incorporating the ability to change capacity in a production planning environment leads to novel managerial insights. Finally, utilizing the real data from the firm, we tested the performance of our heuristic and found that they are less than 0.5% away from the lower bound. Further, we found that three reasonable strategies could lead to very different cost performance (up to 40% difference).

Acknowledgments
The authors wish to thank Luk Van Wassenhove and three anonymous referees for their insightful comments and suggestions which helped us in improving the content and presentation of this paper. They also thank the participants at the INFORMS Conference in Salt Lake City for their comments. The research of the second author was supported in part by the NSF CAREER Award #9984252.

Appendix. Heuristic Allocation of Production Quantities in Different Periods
For \( t = 1 \ldots T \) (period)
- Let \( RC_t = C_t \) where \( C_t \) is the capacity available in period \( t \) and \( RC_t \) is the remaining capacity in any period yet to be allocated to produce and meet demand.
- For \( i = 1 \ldots M \) (going from product with highest to lowest holding cost) Let \( RD_i = d_i \) (remaining demand is set equal to actual demand) where \( RD_i \) stands for the remaining demand yet to be satisfied from production in \( t \) or a prior period.

— While \( (RD_i > 0) \)
  - Let \( t'(i) \) be the period with the lowest marginal cost for producing and meeting product \( i \) demand in period \( t \). That is \( t'(i) = \arg \min_{k \leq t} k (RC_t - (1 - z_k) \beta_i) / \alpha_i \) where \( \lambda_i \) is the capacity (lagrange) multiplier in period \( k \) for period \( (1 \ldots t) \), given the current production levels \( X_k \) and \( (t - k) h_i \) is the cost of carrying inventory of a unit of product \( i \) from period \( k \) to \( t \). If no such \( k \) exists, there is no feasible allocation, otherwise increase production in period \( t'(i) \) by the amount \( q_{it} \) equal to \( \min(RD_i, (RC_t - (1 - z_k) \beta_i) / \alpha_i) \) where \( z_k \) is \( 1(0) \) denotes if product \( i \) has (not) been produced so far in period \( k \). Then update all the values as follows.
  - \( X_k = X_k + q_{it} \)
  - \( RD_i = RD_i - q_{it} \)
  - \( RC_t = RC_t - z_k \beta_i - \alpha_i q_{it} \)
  - Update \( z_k \)
  - \( \lambda_i = \frac{\left( \sum_{t=1}^{M} \sqrt{X_t h_t \beta^2} \right)^2}{2(C_t - \sum_{t=1}^{M} \alpha_i X_t)^2} \)

Note that \( \lambda_i \) is identical to \( \Delta_t \), the Lagrange multiplier computed in §4.2.

References


Accepted by Luk Van Wassenhove; received January 24, 2000. This paper was with the authors 8 months for 1 revision.