

Production and distribution lot sizing in a two stage supply chain

PHILIP KAMINSKY¹ and DAVID SIMCHI-LEVI²

¹*Department of Industrial Engineering and Operations Research, University of California, Berkeley, CA 94720, USA*
E-mail: kaminsky@ieor.berkeley.edu

²*Professor of Engineering Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
E-mail: dslevi@mit.edu

Received April 2001 and accepted March 2002

We develop a two stage model of a manufacturing supply chain. This two stage production transportation model features capacitated production in two stages, and a fixed cost (or concave cost) for transporting the product between the stages. We prove several properties of this model, which we call the Two Stage Production Distribution Problem (2SPDP) model. By placing “non-speculative” assumptions on production and transportation, we show that our model reduces to a related model, with one capacitated production stage with linear production cost, and transportation between two inventory locations with non-linear transportation cost. Finally, we present polynomial algorithms for this model under several different transportation cost structures and capacity assumptions.

1. Introduction

In many manufacturing supply chains, products are manufactured in a series of production facilities, each of which adds additional value to the product. This is often the case if a manufacturer uses an existing series of facilities to manufacture a new product. Each of the existing facilities has the capability to perform a subset of manufacturing steps required by the new product.

For example, in the pharmaceutical industry, manufacturing facilities are expensive to build, and pharmaceuticals have a limited profitable life span (since after patent protection expires, generic manufacturers can manufacture the same product). Thus, production must be ramped up and down rapidly, and furthermore, modern pharmaceuticals are extremely potent, so it often does not take an entire year to manufacture a year’s worth of demand. For these reasons, multi-purpose plants, which can perform several different manufacturing steps for many different products, are typically built. Once a network of these plants is constructed, new products are manufactured sequentially at several different plants, depending on the particular processes required for manufacture.

In many multi-stage manufacturing supply chains, transportation related costs are a significant portion of final product costs. In particular, over the last 20 years, much effort has been devoted to setup reduction, and thus setup costs play a smaller and smaller role in manufacturing decision making. Most transportation, however, exhibits natural economies of scale; in many cases, an empty truck does

not cost much less to operate than a full one. Thus, it is often crucial for successful decision making approaches in multi-stage manufacturing supply chains to explicitly account for these non-linear transportation costs.

In this paper, we begin to explore this problem by considering a two stage production transportation problem. We consider a two stage supply chain which faces a deterministic stream of external demands for a single product. We assume an infinite supply of raw materials, and capacitated production at both stages. Items are manufactured at stage 1, and then held in inventory after stage 1. Items are transported to stage 2, where they are again held in inventory. Additional capacitated production is completed at stage 2 (that is, value is added to each item, but no new items are created), items are held in finished goods inventory after this stage, and this inventory is used to meet final demand. Each period, production levels in stage 1 and stage 2, as well as transportation levels between stage 1 and stage 2, must be determined. We consider so called “non-speculative cost structures,” for both manufacturing and transportation costs. These assumptions ensure that it is never cheaper to manufacture or ship a given quantity by itself earlier than it is needed, although it may be cheaper to manufacture or ship that quantity with another quantity; i.e., economies of scale do exist. We use the characteristics of this model and these “non-speculative cost structures” to develop a new model, which has only one manufacturing stage, and two inventory storage buffers with transportation between them. Alternatively, this new model can be viewed as a two stage production model, with capacity constraint

but linear costs at the first stage, and non-linear costs but no capacity constraints at the second stage. We develop effective algorithms for this new model, and relate the solution of this new model to solution of our original model.

There has been a long history of research into deterministic *single stage* single item lot sizing models, starting with the seminal works of Wagner and Whitin (1958) for the uncapacitated model, and Florian and Klein (1971) for the capacitated model. We discuss the relationship between this work and capacitated single stage models such as Florian and Klein (1971) in more detail in Section 3.4. Aggarwal and Park (1990), Federgruen and Tzur (1991), and Wagelmans *et al.* (1992) developed faster exact algorithms for the uncapacitated case, while Love (1973) and Baker *et al.* (1978) developed more general algorithms for the capacitated case. Various authors have considered multi-stage production models under deterministic *constant* demand. Muckstadt and Roundy (1993) survey and summarize many of these results. The work presented in this paper allows for time-varying multi-period demand. A variety of research has been devoted to heuristic and mathematical-programming-based methods for deterministic multilevel, multi-product lot sizing problems. Baker (1993) summarizes various heuristic approaches which appeared in the literature up to that point, and subsequently, Harrison and Lewis (1996), Katok *et al.* (1998), and Armentano *et al.* (2001) have proposed additional heuristics for these complex problems. Although these papers do not explicitly mention transportation costs, some of the formulations are general enough to encompass our model. However, in contrast to the heuristics these authors develop, we present optimal polynomial approaches.

Finally, Zangwill (1966), and Crowton and Wagner (1973) develop optimal dynamic-programming-based algorithms for uncapacitated multi-stage lot sizing problems without considering transportation. Related approaches are also described in Zipkin (2000).

2. The model

Consider a two stage, n period model of a supply chain, illustrated in Fig. 1. As described above, an infinite supply of raw material is available at stage 1. Each period, x_t^1 , $t = 1, 2, \dots, n$ units are produced at stage 1, at a cost of c_t^1 , where $x_t^1 \leq C_t^1$, the maximum production capacity at stage 1 in period t . Units can be held in inventory at the stage 1 buffer, where a holding cost h_t^1 , $t = 1, 2, \dots, n$ is charged per unit at time t . In addition, s_t units are also shipped to a buffer located before stage 2 (with a shipping lead time of l time units, so that the s_t units will arrive at

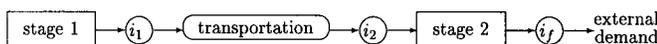


Fig. 1. Model 2SPDP.

time $t + l$). If a shipment occurs, a fixed cost of f_t is charged independent of the number of units shipped, and a variable cost of v_t per unit is charged. Units can be held in inventory in the buffer before stage 2, where a holding cost of h_t^2 is charged per unit at time t . Alternatively, $x_t^2 \leq C_t^2$ units can enter production, at a cost of c_t^2 per unit, where C_t^2 is the maximum stage 2 production capacity in period t . After production, items can be held in finished goods inventory at the post stage 2 buffer where a holding cost h_t^f is charged per unit at time t or they can be shipped to meet demand d_t . We note that in this model, all demand must be met, and that starting inventories are assumed to be zero, so that no demand can be met before period $l + 1$. We call this the Two Stage Production Distribution Problem (2SPDP). In addition, we make several assumptions about the relative value of the various costs, which we detail in subsequent sections.

We summarize each period's order of events below:

1. Stage 1 production x_t^1 is determined, and production cost c_t^1 is charged per unit manufactured.
2. The shipping quantity s_t is determined and units are shipped. Fixed and variable shipping cost ($f_t + s_t v_t$ if $s_t > 0$) is charged.
3. Holding cost h_t^1 is charged on the i_t^1 units remaining in the post stage 1 buffer.
4. Units arrive in the pre stage 2 buffer (l periods after they are shipped).
5. Stage 2 production x_t^2 is determined, and production cost c_t^2 is charged per unit manufactured.
6. Holding cost h_t^2 is charged on the i_t^2 units remaining in the pre stage 2 buffer.
7. x_t^2 units are added to the finished goods buffer.
8. Demand is filled from the finished goods buffer.
9. Holding cost h_t^f is charged on the units remaining in the finished goods buffer.

The mathematical model for the general two stage production transportation problem follows:

Parameters

- d_t = external demand at time t ;
- d_{tn} = cumulative demand from t to n ;
- h_t^1 = holding cost per unit in the post stage 1 buffer at time t ;
- h_t^2 = holding cost per unit in the pre stage 2 buffer at time t ;
- h_t^f = holding cost per unit in the finished goods buffer at time t ;
- C_t^1 = production capacity at stage 1 at time t ;
- C_t^2 = production capacity at stage 2 at time t ;
- f_t = fixed cost for shipping at time t ;
- v_t = variable cost for shipping at time t ;
- c_t^1 = per unit production cost at stage 1 at time t ;
- c_t^2 = per unit production cost at stage 2 at time t ;
- l = number of periods delivery lead time.

Decision variables

- x_t^1 = production quantity at stage 1 at time t ($x_0^1 \equiv 0$);
- x_t^2 = production quantity at stage 2 at time t ($x_u^2 \equiv 0, u = 0, 1, 2, \dots, l$);
- s_t = shipping quantity at time t ;
- y_t = shipment indicator variable at time t ;
- i_t^1 = inventory level at the stage 1 buffer at the end of period t ;
- i_t^2 = inventory level at the pre stage 2 buffer at the end of period t ($i_u^2 \equiv 0, u = 1, 2, \dots, l$);
- i_t^f = inventory level at the post stage 2 buffer at the end of period t .

$$(P) \min \sum_{t=1}^n (i_t^1 h_t^1 + i_t^2 h_t^2 + i_t^f h_t^f + y_t f_t + s_t v_t + c_t^1 x_t^1 + c_t^2 x_t^2),$$

subject to

$$x_t^1 \leq C_t^1 \quad t = 1, 2, \dots, n, \tag{1}$$

$$i_t^1 = i_{t-1}^1 - s_t + x_t^1 \quad t = 1, 2, \dots, n, \tag{2}$$

$$s_t \leq y_t d_{(t+l)n} \quad t = 1, 2, \dots, n, \tag{3}$$

$$x_t^2 \leq C_t^2 \quad t = 1, 2, \dots, n, \tag{4}$$

$$i_t^2 = i_{t-1}^2 - x_t^2 + s_{t-l} \quad t = l+1, l+2, \dots, n, \tag{5}$$

$$i_t^f = i_{t-1}^f + x_t^2 - d_t \quad t = 1, 2, \dots, n, \tag{6}$$

$$i_t^1, i_t^2, i_t^f, x_t^1, x_t^2, s_t \geq 0 \quad t = 1, 2, \dots, n, \tag{7}$$

$$y_t \in \{0, 1\} \quad t = 1, 2, \dots, n, \tag{8}$$

Constraints (1) and (4) are capacity constraints in the two stages. Constraints (2), (5) and (6) are the inventory balance equations for the three buffers. Constraint (3) ensures that the fixed cost is paid each time product is shipped.

In Section 3, we detail some further restrictions on the values of the parameters of this model, and then describe effective algorithms for these restricted cases. In Section 4, we present effective algorithms for the case with a more general transportation cost function. Finally, in Section 5, we list several remaining open questions related to this model.

3. Analysis and algorithms

3.1. Lead times

Since in this model we assume that the starting inventory is zero, and we require all demand to be met, we do not allow demand in any period before period $l + 1$. Also, observe that given an instance with positive lead time l , we can create an equivalent model (that is, a model with the same optimal solution) by creating a new demand vector d' such that $d'_t \equiv d_{t+l}, t = 1, 2, n - l$, and setting the new lead time $l' = 0$. Thus, without loss of generality, we assume zero lead time for the rest of this paper.

3.2. Assumptions

For the remainder of the paper, we make the following two assumptions. First, we assume that

$$h_t^1 < h_t^2 < h_t^f \quad \forall t \in 1 \dots n. \tag{9}$$

This is justified, as value is added at each manufacturing or transportation step. In addition, we make the following set of assumptions, often known as the *non-speculative* assumptions:

$$c_t^1 + h_t^1 > c_{t+1}^1 \quad \forall t \in 1 \dots n - 1, \tag{10}$$

$$c_t^2 + h_t^2 > c_{t+1}^2 \quad \forall t \in 1 \dots n - 1. \tag{11}$$

This assumption is true in the case of identical costs in each period, or when the time-value of money is considered. We make a similar assumption for transportation costs:

$$x(v_t + h_t^2 - h_t^1) + f_t > xv_{t+1} + f_{t+1} \quad \forall t \in 1 \dots n - 1, x > 0. \tag{12}$$

We note that this implies

$$v_{t+1} < v_t + (h_t^2 - h_t^1) \quad \forall t \in 1 \dots n - 1, \tag{13}$$

and that identical transportation costs, for example, meet this condition. In Section 4, we consider a more general transportation cost function.

3.3. Determining x_t^2

These assumptions lead to the following recursive formulation for determining $x_t^2, t = 1, 2, \dots, n$:

$$x_t^2 = \min \left\{ C_t^2, d_t + \sum_{i=t+1}^n (d_i - x_i^2) \right\}, \tag{14}$$

where $\sum_{i=n+1}^n (d_i - x_i^2) \equiv 0$. Equation (14) schedules production in stage 2 as late as possible, subject to capacity constraints.

Property 1. Equation (14) gives a set of optimal values for $x_t^2, t = 1, 2, \dots, n$.

Proof. The proof follows in a straightforward way from assumptions (9) and (10). ■

3.4. Model 2SPDP'

We note that given an instance of the model described above, Property 1 allows us to develop an alternate model with the same optimal production and shipping quantities. As in the model described above, an infinite supply of raw material is available at stage 1. Each period, $x_t^1, t = 1, 2, \dots, n$ units are produced at stage 1, where $x_t^1 \leq C_t^1$, at a cost of c_t^1 . Units can be held in inventory at the stage 1 buffer, where a holding cost h_t^1 is charged per unit at time t . s_t units can also be shipped to another buffer (we refer to this as buffer 2), and a fixed cost of f_t is charged if $s_t > 0$. Units can be held in inventory in buffer 2, where a holding cost of h_t^2 is charged per unit at time t or shipped to meet demand $d'_t = x_t^2, t = 1, 2, \dots, n$, where x_t^2 is determined using Equation (14). Thus, we have eliminated a stage from the original model. We call this equivalent model 2SPDP', and focus on 2SPDP' for the rest of the paper. This model is illustrated in Fig. 2.

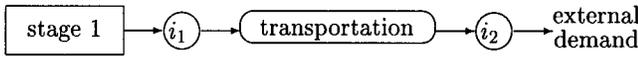


Fig. 2. Model 2SPDP'.

Intuitively, this model appears to have an interesting relationship to the traditional single level, single product capacitated lot sizing model addressed by Florian and Klein (1971), as well as subsequent authors. On one hand, it seems to be in some sense easier to solve than that model, since the fixed cost is decoupled from capacitated production decisions. On the other hand, given the fact that production at stage 1 is time varying, and the amount shipped is bounded by the amount produced, this model appears to be related to the single stage production model with time-varying capacity, which is known to be NP-hard, even in many special cases (Florian *et al.*, 1980; Bitran and Yanesse, 1982). Indeed, our methodology, particularly for the concave transportation cost case discussed in Section 4, is related to the approach employed by Florian and Klein. However, because the non-linear transportation costs are decoupled from the manufacturing costs, we use different structural properties of the optimal solution, and are in some cases able to solve problems with time-varying production capacities in polynomial-time.

The following properties, analogous to the well-known *zero inventory ordering property* for uncapacitated lot sizing, hold for 2SPDP' when conditions (1) and (2) hold:

Property 2. *In any optimal solution to 2SPDP', shipping only occurs in periods t where the inventory in the stage 2 buffer, $i_{t-1}^2 = 0$, where $i_0^2 \equiv 0$.*

Proof. Assume an optimal solution in which in some period $j, j = 1, 2, \dots, n, s_j > 0$ and $i_{j-1}^2 = \gamma > 0$. Let $k, 1 \leq k < j$ be the period when the items in inventory at the stage 2 buffer originally shipped. Now create a new solution identical to the first, but with $s'_k = s_k - \gamma$, and $s'_j = s_j + \gamma$. Assumption (13) implies that this is a less expensive solution. ■

The following property follows immediately from Property 2.

Property 3. *In any optimal solution to 2SPDP', the shipping quantity s_t at time $t, t = 1, 2, \dots, n$ is equal to some partial sum of future demands $\sum_{i=t}^a d_i, t \leq a \leq n$.*

3.5. Production schedule given

In this section, we address the problem of solving the transportation part of the 2SPDP' model given a production schedule, i.e., if the variables $x_t^1, t = 1, 2, \dots, n$ are already determined. We call this the Shipping problem.

Based on the Properties 2 and 3, we observe that a property analogous to Wagner and Whitin's (1958) *Inventory Decomposition Property* is true:

Property 4. *For the Shipping problem, if $i_t^1 = 0$ for some t , it is optimal to consider periods 1 through $t - 1$ by themselves.*

This suggests a dynamic programming approach for the Shipping problem. Note that in this case we can ignore manufacturing costs, and the holding costs in stage 1, since they are sunk costs. To account for incremental holding costs, we define $h_t = h_t^2 - h_t^1, t = 1, 2, \dots, n$. This dynamic programming approach is similar to the standard Wagner and Whitin (1958) lot sizing algorithm, except that infeasible solutions are ignored. Let $v(t)$ be the minimum cost to go starting in period t , given that $i_{t-1} = 0$ and let v_{ab} be the cost of shipping in period a to exactly meet the total demand in the interval from period a to period $b - 1, 1 \leq a < b \leq n + 1$. Then,

$$v(t) = \left\{ \min_{t < i \leq n+1} (v_{ti} + v(i)) : \sum_{j=1}^t x_j^1 \geq \sum_{j=1}^{i-1} d_j \right\},$$

where $v(n + 1) = 0$. Note that we ensure feasibility by checking that we have made enough to ship. Let $D_q = \sum_{i=1}^q d_i, q = 1, 2, \dots, n$, and observe that cost of shipping in period a to exactly meet the total demand in the interval from period a to period $b - 1, 1 \leq a < b \leq n + 1$, is:

$$v_{ab} = \left(s_a + (D_{b-1} - D_{a-1})v_t + \sum_{i=a}^{b-2} h_i \left(\sum_{j=i+1}^{b-1} d_j \right) \right). \quad (15)$$

This dynamic program can be represented by the network illustrated in Fig. 3. An arc between nodes i and j represents a shipment at time i to serve demands in times $i, i + 1, i + 2, \dots, j - 1$. This is a directed, acyclic network with $n + 1$ nodes and $O(n^2)$ arcs. The shortest path through this network represents an optimal shipping plan, and can be found in $O(n^2)$ time.

3.6. An effective algorithm for 2SPDP'

We now present an algorithm for 2SPDP'.

First, we define a *block* as follows:

Definition 1. A block $[s, t]$ is a set of one or more consecutive periods from period s to period $t, 1 \leq s \leq t \leq n$, such that $i_{s-1}^1 = 0, i_t^1 = 0$, and either $s = t$ or $i_a^1 > 0, s \leq a < t$. Recall that $i_0^1 \equiv 0$.

Clearly, any feasible schedule can be divided into blocks, and the first block starts (and possibly ends) at time 1.

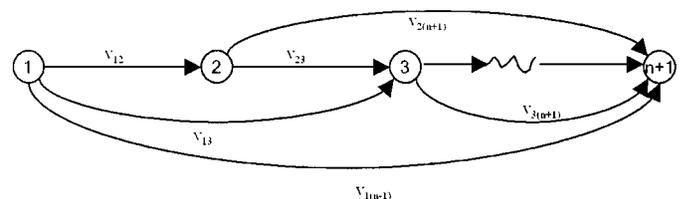


Fig. 3. DP Network with the production quantities being given.

Furthermore, in any optimal schedule, the following properties can be proved for each block.

Property 5. *In any optimal schedule, for any block, production will be at capacity in all periods except possibly for the first period.*

We note that this first period can be the only period in the block, and that production can be zero in this period.

Proof. If the block is one period long, the property is trivially true. Assume by contradiction that in an optimal solution, in block $[s, t]$, production is less than capacity in some period $q, s < q \leq t$. Since this is a block, we know that $i_{q-1}^1 > 0$. But this implies that $\min(i_{q-1}^1, C_q^1 - x_q^1)$ production in the block from the periods $s, s + 1, \dots, q - 1$ could be moved to period q . By Assumption (10), this will decrease costs. Thus, this block cannot be optimal. ■

Property 6. *In any block, the total production in that block is equal to some partial sum of consecutive demands.*

Proof. This follows from the definition of a block and Property 3. ■

This property, combined with the definition of a block, implies that any block $[s, t]$ serves all demands in some interval $a, a + 1, \dots, b$.

These properties, along with Property 2, suggest the following dynamic-programming-based approach to solving this problem. Let $v(t, a)$ be the minimum cost to go starting in period t , given that $i_{t-1}^1 = 0$, i.e., a block ended in period $t - 1$, and that this block covered demands through period $a - 1$. Let $v_{st}^{ab}, s \leq t \leq n + 1, a \leq b \leq n + 1$, be the minimum possible cost of a block which lasts from period s to period $t - 1$, and services demand from period a to period $b - 1$. Note that if it is not possible for a block that lasts from period s to period $t - 1$ to service demand from period a to period $b - 1$ due to capacity constraints, we set $v_{st}^{ab} = \infty$. We observe that there will be $O(n^3)$ blocks with non-infinite cost, since for each possible a, b , and t , there is only one possible s (due to Property 5, which states that production will be at capacity in all blocks except for the first block in a period, and Property 6, which implies that this production will exactly meet demand in this interval).

In this case:

$$v(t, a) = \min_{t \leq s \leq n+1, a \leq b \leq n+1} (v_{ts}^{ab}, v(s, b)).$$

where $v(n + 1, n + 1) = 0$.

This dynamic program can be represented by the network illustrated in Fig. 4. Note that for simplicity, only the nodes are illustrated. In addition, there is an arc between every node (s, a) and (t, b) for which $s < t, a < b$, and $v_{st}^{ab} < \infty$. An arc between node (s, a) and node (t, b) represents a block starting at time s and ending at time $t - 1$ which covers demand from period a to period $b - 1$, and has cost v_{st}^{ab} . This is a directed, acyclic network with $O(n^2)$ nodes, and $O(n^3)$ arcs.

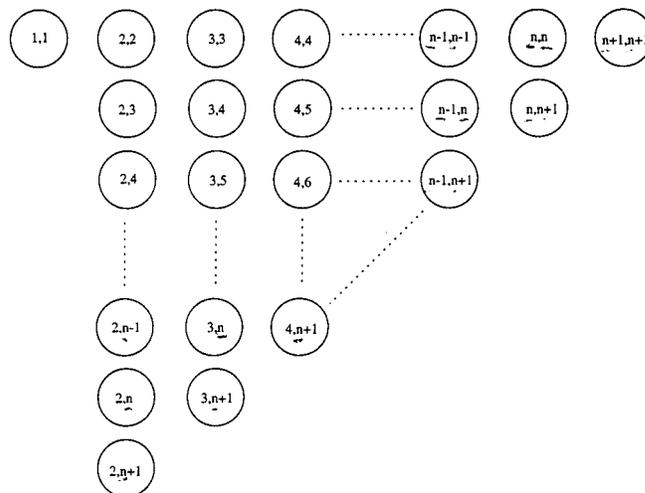


Fig. 4. 2SPDP' network. Arcs exist from node (s, a) to node (t, b) if $s < t, a < b$, and $v_{st}^{ab} < \infty$.

It remains to determine v_{st}^{ab} , the cost on each arc in the network described above. First, given a block $[s, t - 1]$ that serves demand from a to $b - 1$, production is known. In particular, production is at capacity in all periods except possibly for the first period (although this period may be the only period in the block). The production in the first period in the block must equal

$$\sum_{i=a}^{b-1} d_i - \sum_{i=s+1}^{t-1} C_i^1.$$

Since production is known for a particular block, we can use the dynamic programming approach outlined in Section 3.5 to find v_{st}^{ab} in $O(n^2)$ time. In particular, we can find the shortest path on a graph analogous to the one pictured in Fig. 3, with only demands from a to $b - 1$. We know that an arc from $t - 1$ to b must be in the shortest path, and we look for the shortest path between a , the period in which the first shipment must occur, to $t - 1$. Thus, the running time of this dynamic programming algorithm is dominated by the time to calculate arc costs, which is $O(n^5)$. This complexity can be somewhat reduced by the following observation.

For a given t and $b, 1 \leq t \leq b \leq n$, let $a' = \min\{a \leq b \mid (\exists s \leq a : \sum_{i=s}^{t-1} C_i \geq \sum_{i=a}^{b-1} d_i)\}$, and let s' be this s . In other words, $(a', b - 1)$ is the longest sequence of demand ending at period $b - 1$ which can be optimally covered by a single block ending at time $t - 1$, and this block extends from period s' to period $t - 1$. If we create a network as described above to find v_{st}^{ab} and solve it backwards, we end up with the shortest path from b to all feasible nodes $b' < b$.

Property 7. *Consider the shortest path representation described above. If we use it to find the optimal shipping schedule to determine the value v_{st}^{ab} as described above, then for any feasible value of $a, 1 \leq a \leq b$, the optimal shipping schedule*

is the (backwards) shortest path from $t - 1$ to a on the same graph.

We note that there must be a shipment at node $t - 1$ which covers demand through b , and there must be a shipment at time a .

Proof. The proof proceeds by comparing the network for the $v_{s't}^{a'b}$ problem described above, with the network for some v_{st}^{ab} problem. Recall that $(a', b - 1)$ is the longest sequence of demand ending at period $b - 1$ which can be optimally covered by a single block ending at time $t - 1$, and thus any other sequence of demand must be shorter, and is therefore covered by a block starting at a time $t, t \geq s'$. In other words, $a \geq a', s \geq s'$. Clearly, identical arcs in both networks will have identical costs. It remains to show that the identical arcs will appear in both networks on the nodes $a, a + 1, \dots, t - 1$. Since $a \geq a'$ and $s \geq s'$, all of the arcs in the v_{st}^{ab} problem are feasible, and thus will appear, in the $v_{s't}^{a'b}$ problem. Also, observe that since both blocks cover demand ending at the same period with items made in a block ending in the same period, each unit of production can be allocated to its demand such that the allocation is the same in the $v_{s't}^{a'b}$ and v_{st}^{ab} problems. This, combined with the fact that a shipment at time a covers demand starting at time a , implies that arcs representing shipments at time a will appear in both graphs. Thus, the (reverse) shortest path from $t - 1$ to a is the same in both graphs, and clearly the arc from $t - 1$ to b is present in both solutions. ■

For example, consider $v_{s't}^{a'b}$ where $s' = 4, t = 8, a' = 5, b = 9$. In this case, $(5, 8)$ is the longest sequence of demand ending at period 8 which can be optimally covered by a block ending at time 7. Thus, we know that for this block, production is at capacity in periods 5, 6, and 7, and possibly fractional in period 4. Given this production schedule, we can create a network of the type illustrated in Fig. 5, where each arc represents a feasible shipping policy as we discussed in Section 3.5. By finding the shortest path from node 4 to node 9 in this graph, we determine $v_{(4)(8)}^{(5)(9)}$. At the same time, we can find the shortest path from node 9 to all other nodes in the graph which can be reached (backwards) from node 9. Now, suppose we are interested in the cost of another block ending at time 7 which covers demand through time 8. This block must cover demand starting at a time greater than five, since by definition of a' and s' , $(5, 8)$ is the longest sequence of demand which could be covered by a block ending at time 7. Suppose the block from five to seven can feasibly cover demand from six to eight. Clearly,

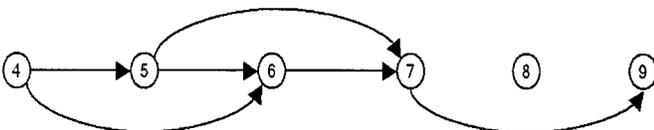


Fig. 5. DP example; reducing complexity.

this implies that a path from node 5 to node 9 must be present in Fig. 5. Furthermore, the cost $v_{(5)(8)}^{(6)(9)}$ is the cost of the shortest path from node 5 to node 9, which we have already found.

Thus, the shortest path problem only needs to be solved for each value of t and $b \geq t$, or $O(n^2)$ times. This means that the time to calculate optimal shipping patterns for all blocks is bounded by $O(n^4)$, and given an optimal shipping pattern, the cost of that pattern can be calculated in $O(n)$ time. The overall running time of the dynamic programming algorithm for 2SPDP is $O(n^4)$. Since 2SPDP can be converted into an equivalent 2SPDP' in $O(n^2)$ time, we can solve the original problem in $O(n^4)$ time.

4. Concave transportation cost function

A more general concave cost function for transportation, $f_i(s_i)$, rather than fixed plus linear cost transportation cost function employed in the previous section, more accurately models real-world transportation costs which feature economies of scale, or quantity discounts. Unfortunately, in this case, Properties 2 and 3 do not hold, so the algorithms described above will not work. Thus, we are motivated to identify new properties, and develop additional, more complex algorithms.

In this section, we allow the transportation cost functions $f_i()$, $t = 1, 2, \dots, n$, to be any set of concave functions where transportation cost increases with the number of units shipped, and where the following restriction holds for any positive shipping quantity x :

$$f_i(x) + (h^2 - h^1)x > f_j(x) \quad \forall i, j : 1 \leq i < j \leq n, x > 0. \tag{16}$$

This restriction is clearly met if cost functions are identical, or if shipping costs reflect the time value of money.

In this case, the following properties, analogous to Properties 2 and 3 hold.

Property 8. *There exists an optimal solution to 2SPDP' in which shipping only occurs in periods t where the inventory in the stage 2 buffer, $i_{t-1}^2 < d_t$ where $i_0^2 \equiv 0$.*

Proof. Assume by contradiction that there exists a period t in which $s_t > 0$ and $i_{t-1}^2 > d_t$. Clearly, the quantity s_t could be shipped in the next period, which by assumption (16) would be less expensive. ■

Property 9. *There exists an optimal solution to 2SPDP' in which either the shipping quantity at time $t, 1 \leq t \leq n$, plus all previous shipping quantities, $\sum_{i=1}^t s_i$, is equal to some partial sum of demands $\sum_{i=1}^a d_i, t \leq a \leq n$, or the shipping quantity is all available inventory, $s_t = i_{t-1}^1 + x_t^1$, or both.*

Proof. Assume by contradiction that in an optimal solution, at time $t, 1 \leq t < n, s_t$ is a quantity sufficient to meet

all demand up to period u , $u \geq t$ and some quantity α which is equal to a fraction of the demand in period v , the first period with non-zero demand after period u . Furthermore, assume that $i_t^1 > 0$. This implies that the rest of the demand at time v is shipped at some time w , $t < w \leq v$, and that at least some of those units could have been shipped at time t . However, since the shipping function is concave, we know that either it is cheaper to ship the available units at time t and hold them at stage 2, or to ship the α units at time w (or we are indifferent between the current solution and one of these two solutions). Thus, either this is not an optimal solution, or an equivalent optimal solution can be found for which this property holds. ■

4.1. The Shipping problem

Note that for this problem, the *Shipping problem*, or the problem of determining shipping quantities once the production schedule is set, is not trivial, since the *Inventory Decomposition Property* does not hold. However, we can use an approach analogous to the approach described in Section 3.5 with additional consideration of possible starting inventory levels.

Observe that from Property 9, we know that in each period in which shipping occurs, either a quantity so that the total shipped amount is equal to a partial sum of demands, or a quantity sufficient to make the inventory at the stage 1 buffer zero, is shipped.

Consider a shipment made in period j , $1 \leq j \leq n$ which covers demand through period $k \geq j$, and possibly some subsequent demand. If this shipment exactly covers demand, there may be many possible values of k for which this is feasible. If, however, the shipment does not exactly cover demand, by Properties 8 and 9, all of the inventory manufactured up to that time must ship, which implies that

$$k = \min \left\{ i > j \mid D_i - \left(\sum_{l=1}^j x_l^1 \right) > 0 \right\} - 1.$$

Also, recall that in some optimal solution to 2SPDP', if $\alpha_t = i_{t-1}^2 > 0$, a shipment is made in period t if and only if $\alpha_t < d_t$. Furthermore, this is only possible if $\alpha_t = 0$, or if the previous shipment quantity equaled the total quantity produced up to that shipment. Thus, if we define the function

$$g_t(i) = \sum_{j=1}^i x_j^1 - D_{t-1},$$

then the set of possible α_t values in each period t , $1 \leq t \leq n$ in which shipments are made is limited to the positive values of $g_t(i)$, $i = 1, 2, \dots, t - 1$ (plus zero).

This suggests that we can use a shortest path approach similar to the one outlined in Section 3.5.

As before, we represent this problem by a directed network, although this one is more complicated than the network used previously. For each time period t , $1 \leq t \leq n$,

there is a node for each possible value of α_t if a shipment is made, which we call α_t^j , where the index j represents the particular feasible value of α_t , and $\alpha_t^0 = 0$. An arc between nodes (α_a^j) and (α_b^k) , $1 \leq a < b \leq n + 1$, represents a shipment at time a to serve demands through time $b - 1$, plus the additional quantity α_b^k (which is possibly zero). Note that once a shipment is made to serve demands through time $b - 1$, plus the additional quantity α_b^k , no additional shipment is made until time b , and the quantities i_{b-1}^1 and i_{b-1}^2 are fully determined. Thus, the cost on each arc is simple to calculate. This is a directed, acyclic network, with at most $O(n^2)$ nodes and at most $O(n)$ arcs per node (representing feasible shipping plans). The shortest path through this network represents an optimal shipping plan, and can be found in $O(n^3)$ time.

4.2. Properties of the optimal solution

Next, we present several properties of an optimal solution to 2SPDP'. In the next section, we use these properties to develop an algorithm for the constant capacity version of this problem.

First, recall that we defined a block as follows:

Definition 1. A block $[s, t]$ is a set of one or more consecutive periods from period s to period t , $1 \leq s \leq t \leq n$, such that $i_{s-1}^1 = 0$, $i_t^1 = 0$, and either $s = t$ or $i_a^1 > 0$, $s \leq a < t$. Recall that $i_0^1 \equiv 0$.

In addition, we define an interval as follows:

Definition 2. An interval $[q, r]$ is a set of consecutive periods from q to r , $1 \leq q \leq r \leq n$, such that the following set of conditions holds:

- $i_{q-1}^1 = 0$;
- $i_r^1 = 0$;
- if $q \neq 1$, there exists a time l , $q - 1 \leq l \leq n$ such that $\sum_{i=1}^{q-1} s_i = D_l$;
- there exists a time m , $r \leq m \leq n$ such that $\sum_{i=1}^r s_i = D_m$;
- if $q \neq r$, there **does not** exist a period p in the interval $q, q + 1, \dots, r - 1$ and a time w such that both $\sum_{i=1}^p s_i = D_w$, and $i_p^1 = 0$.

In other words, blocks define periods in which starting and ending inventory is zero, and no intermediate inventories are zero. Intervals add the additional restriction that the final shipment is such that the total amount shipped up to that time including the shipment in the final period of the interval, is a partial sum of demands. Clearly, any feasible schedule can be divided into one or more intervals, each of which can be subdivided into one or more blocks.

The definition of an interval implies that:

Property 10. In any interval, the total production in an interval is equal to some partial sum of demands.

This property implies that any interval $[q, r]$ serves all demands in some interval $a, a + 1, \dots, b$, and no additional demands.

In addition, the following properties can be demonstrated:

Property 11. *There is at least one optimal solution in which production is at capacity in all periods of a block except perhaps for the first period.*

Proof. Assume by contradiction that in a block, partial production occurs in some period a later than the first period of the block. Clearly, production of $\min(C_a - x_a, i_{a-1}^1)$ can be shifted from period $a - 1$ to period a without increasing cost, by assumption (11). This will either decrease costs, create a new block structure, or eliminate partial production in period a . If the later occurs, and $a - 1$ is not the first period of the block, continue to apply this same approach to shifting production. ■

We note that zero production blocks one period long can exist. However, Property 11 implies that there will be no zero production periods within a block with positive production, since such zero production blocks would have to be in the first period, and this would imply that inventory was zero at the end of this period, so that a new block would start in the next period. Thus, we conclude that:

Property 12. *There is at least one optimal solution for which all periods of zero production are in their own single period blocks. In other words, $x_i = 0$ if and only if $i_{i-1}^1 = i_i^1 = 0$.*

Finally, we characterize production in an interval by demonstrating that:

Property 13. *There is at least one optimal solution in which production is at capacity or at zero except possibly for one period during the interval.*

Proof. Assume that we are given an optimal solution for which Properties 11 and 12 apply, but for which Property 13 does not apply. Without loss of generality, assume that there is a single interval G for which Property 13 does not apply, let a represent the second latest period in the interval for which production is below capacity but above zero, and let b represent the latest such period. By Property 11, both a and b must be the first period in their respective blocks. We will show by construction that we can either create an equivalent solution in which the two below capacity production periods are reduced to a single below capacity production period within the interval, or one of the following occurs: the interval is divided into two new, shorter intervals, or a block in the interval is divided into two new, shorter blocks. Since we can repeatedly apply this construction approach, and intervals and blocks can be no shorter than one period, this is sufficient to prove the result.

Let i be the minimum of

- the minimum stage 1 inventory at the end of each time period in the interval $[a, b]$ excluding those periods which are the final period in their respective blocks;
- the fractional inventory at stage 2 (that is, the inventory which meets some but not all of the demand in a future period) at the end of those periods which are the final period in their respective blocks.

If $i \leq C_b - x_b$, then it is clearly feasible to shift production of i units to period b without altering the shipping schedule. Since we already have an optimal solution, this shift cannot result in lower cost. If this shift results in the same total cost, make the shift. This either results in more, shorter blocks within the interval, or more, shorter intervals. Furthermore, if all blocks in the interval are of length one, this must result in more, shorter intervals. We observe that it is *not possible for this shift to result in higher costs*, since this would imply that it must be cheaper to manufacture some of the units currently produced in period b in period a , and since production is less than capacity at a , this is feasible.

If $i > C_b - x_b$, then it is clearly feasible to shift production of $C_b - x_b$ units to period b . By the same reasoning as above, we see that this must result in a schedule of equal cost, while increasing production to capacity in period b . ■

4.3. An effective algorithm for the constant capacity case

Unfortunately, the presence of zero production blocks with the intervals makes it difficult to develop an effective algorithm for the general case. However, the definition of an interval, along with these properties, suggests the following dynamic-programming-based approach to solving this problem for the constant capacity case. For simplicity, we use the same notation as in Section 3.6, although here we are referring to an interval rather than a block.

Let $v(t, a)$ be the minimum cost to go starting in period t , given that an interval ended in period $t - 1$, and that this interval covered demands through period $a - 1$. Let v_{st}^{ab} be the minimum possible cost of an interval which lasts from period s to period $t - 1$, and services demand from period a to period $b - 1$. Note that if it is not possible for an interval that lasts from period s to period $t - 1$ to service demand from period a to period $b - 1$ due to capacity constraints, we set $v_{st}^{ab} = \infty$. In this case:

$$v(t, a) = \min_{t \leq s \leq n+1, a \leq b \leq n+1} (v_{ts}^{ab}, v(s, b)),$$

where $v(n + 1, n + 1) = 0$.

This dynamic program can be represented by the same network as in Section 3.6 (illustrated in Fig. 4). Recall that although only the nodes are illustrated, there is an arc between every node (s, a) and (t, b) if $s < t$ and $a < b$. An arc between node (s, a) and node (t, b) represents an

interval starting at time s and ending at time $t - 1$ which covers demand from period a to period $b - 1$, and has cost v_{st}^{ab} . This is a directed, acyclic network G1 with $O(n^2)$ nodes, and $O(n^2)$ arcs per node, for a total of $O(n^4)$ arcs, since in this case, for a given a, b and t , there are a variety of possible s values due to the possibility of zero production.

Unfortunately, determining v_{st}^{ab} , the cost on each arc in the network described above, is more complicated for this concave cost case than for the fixed plus linear cost case described previously. We need an efficient method of both determining the manufacturing schedule within an interval, and for determining the shipping schedule for all the items manufactured during the interval.

Recall that an interval is made up of one or more blocks. From Property 10, the total production quantity in an interval is known. Furthermore, by Property 13, the production schedule within an interval consists of some single period zero production blocks, and for the remaining blocks, production is at capacity in all periods except perhaps one period during one of the blocks. As in Florian and Klein (1971), we set ϵ equal to the partial production (if D is the total demand served by the interval, $\epsilon = (D \bmod C)$). Thus, we know the cumulative amount produced between the start of the interval, period s , and any period in the block, can take on at most the values $\{0, \epsilon, C, C + \epsilon, \dots, (\lfloor D/C \rfloor), (\lfloor D/C \rfloor) + \epsilon\}$. Let the cumulative production in the interval at period $j - 1$ be X_j .

This suggests the following procedure to determine the optimal policy for an interval from s to $t - 1$, $1 \leq s \leq t \leq n + 1$. Construct a network G2 with a single starting node for time s , and nodes for each possible value of $X_j, j = s + 1, s + 2, \dots, t$ in the graph. An arc between two nodes in the graph X_i and $X_j, i < j$, represents a block which starts at the time of the first node, i , and ends at the period immediately preceding the time of the second node, period $j - 1$. At time $j - 1$, X_j units have been produced. Thus, an arc uniquely defines a production schedule within a block, and indicates when the final shipment in the block occurs. If a particular block is not feasible, either because a partial production period has appeared in an earlier block, or because production is not sufficient to meet demand up to that point, the arc does not appear in the network. The cost on each arc includes the production cost for all the material made during that block, as well as the holding and shipping costs associated with that material. Thus, a path through the network represents a feasible production plan for the interval. This network will have $O(n^2)$ nodes, and $O(n^3)$ arcs since zero production only occurs in single period, zero production blocks, so the shortest path can thus be found in $O(n^3)$ time.

Note that although Florian and Klein (1971) employ a similar approach to determine production levels in a capacitated single stage production system, their model only requires production decisions, whereas ours requires production and shipping decisions. Thus, in the graph they use

to make production decisions, arcs exist for each time period. We need to introduce the concept of shipping blocks, and use arcs which represent production in a block, rather than a single time period.

The production costs associated with each arc are easy to determine. Shipping and holding costs require the determination of a shipping schedule. However, by Property 9 and the definition of a block, all of the shipments within a block (excluding the final shipment) occur such that the total amount shipped at each period is equal to a partial sum of demands. Thus, except for the first shipment during a block, shipping only occurs when stage 2 inventory is zero, and the first shipment occurs in the first period t for which $i_{t-1}^2 < d_t$, so the shipping schedule for a block can easily be found by using a dynamic programming approach exactly analogous to the one used in Section 3.5. The optimal cost of a particular interval can therefore be found in $O(n^5)$ time using G2, and since G1 has $O(n^4)$ arcs representing intervals, the optimal solution to the initial problem can be found in $O(n^9)$ time.

The complexity of this algorithm can be reduced by making the following observation. Consider each of the arcs terminating at a particular node X_j in the network described above. These arcs represent a variety of blocks which have two things in common: (i) they terminate with a final shipment at time $j - 1$; and (ii) the final shipment covers demand starting at time $j - 1$, and completing when the quantity represented by X_j is used to meet demand. Thus, we can use the approach described at the end of Section 3.6, and the reasoning behind Property 7, to solve the shortest path in reverse, and find at the same time the shipping schedules for all of the blocks described by arcs terminating at X_j . This implies that the costs on all the arcs in G2 can be found in $O(n^4)$ time, so the optimal solution to the initial problem can be found in $O(n^8)$ time.

5. Conclusions and future research

We have introduced a two stage model which combines production lot sizing and shipping decisions, and described effective algorithms for the fixed charge case, and the concave shipping cost case with constant manufacturing capacities.

This initial exploration of 2SPDP leaves many remaining questions. For example:

- Is there a more efficient algorithm for this problem?
- Can the concave cost problem be solved under more general capacity assumptions?
- Can an efficient algorithm be found for more general transportation cost structures than those analyzed in this paper, such as those which do not meet restriction (16)?
- Our algorithms require a fixed lead time. Can effective algorithms be developed for time-varying lead times. In this case, the model with positive lead times is not immediately

equivalent to a model with zero lead time, so a different approach will have to be used.

In addition, many supply chains have additional stages. Our approach was in many ways dependent on the fact that we could reduce the 2SPDP model to the 2SPDP' model, which has only a single production stage. Although this approach can be generalized to multi-stage models as long as there is only a single transportation stage (that is, a single stage with economies of scale), it cannot be generalized in an obvious way to the case with transportation between more than two of the stages. We would like to find efficient algorithms for this n stage nSPDP problem, although this is likely to require a different solution approach. Also, we are currently investigating a variety of stochastic versions of this model.

Finally, this paper would be incomplete without some remarks concerning the weaknesses of our model and analysis. As mentioned above, our results are quite specific to our cost structures. In particular, we require a two stage supply chain in which economies of scale are present in transportation but not in manufacturing, and in which transportation to the final customer is not a consideration. We argued in the Introduction that in many cases manufacturing setup costs have become negligible; in some cases, however, manufacturing setup costs are still significant. In addition, our model is a deterministic model, although in most situations, demand is stochastic, and managers need to react to changing demand. Nevertheless, our model and solution algorithm can serve as a useful tool in situations for which our assumptions are practical. Also, in this paper, we demonstrate that polynomial-time optimal algorithms can be found for a multi-stage problem, in contrast to most previous multi-stage research, which has focused on heuristics. We hope to build on this knowledge in future research, as we attempt to find efficient algorithms for more complex multi-stage models.

Acknowledgement

Research supported by National Science Foundation Contracts DMI-9732795 and DMI-0092854.

References

- Aggarwal, A. and Park, J.K. (1990) Improved algorithms for economic lot-size problems. Working paper, Laboratory for Computer Science, MIT, Cambridge, MA.
- Armentano, V.A., Esther Berretta, R. and Franco, P.M. (2001) Lot sizing in capacitated multi-stage serial systems. *Production and Operations Management*, (in the press).
- Baker, K.R. (1993) Requirements planning, in *Handbooks in Operations Research and Management Science: Logistics of Production and Inventory*, Graves, S.C., Rinnooy Kan, A.H.G. and Zipkin, P.H. (eds.), North-Holland, Amsterdam, pp. 571–627.
- Baker, K.R., Dixon, P., Magazine, M.J. and Silver, E.A. (1978) An algorithm for the dynamic lot-size problem with time-varying production capacity constraints. *Management Science*, **24**, 1710–1720.
- Bitran, G.R. and Yanasse, H.H. (1982) Computational complexity of the capacitated lot size problem. *Operations Research*, **28**, 1174–1186.
- Crowton, W.B. and Wagner, M.H. (1973) Dynamic lot size models for multi-stage assembly systems. *Management Science*, **20**, 14–21.
- Federgruen, A. and Tzur, M. (1991) A simple forward algorithm to solve general dynamic lot sizing models with n periods in $O(n \log n)$ or $O(n)$ time. *Management Science*, **37**, 909–925.
- Florian, M. and Klein, M. (1971) Deterministic production planning with concave costs and capacity constraints. *Management Science*, **18**, 12–20.
- Florian, M., Lenstra, J.K. and Rinnooy Kan, A.H.G. (1980) Deterministic production planning: algorithms and complexity. *Management Science*, **36**, 240–243.
- Harrison, T.P. and Lewis, H.S. (1996) Lot sizing in serial assembly systems with multiple constrained resources. *Management Science*, **42**, 19–36.
- Katok, E., Lewis, H.S. and Harrison, T.P. (1998) Lot sizing in general assembly systems with setup costs, setup times, and multiple constrained resources. *Management Science*, **44**, 859–877.
- Love, S.F. (1973) Bounded production and inventory models with piecewise concave costs. *Management Science*, **11**, 313–318.
- Muckstadt, J.M. and Roundy, R.O. (1993) Analysis of multistage production systems, in *Handbooks in Operations Research and Management Science: Logistics of Production and Inventory*, Graves, S.C., Rinnooy Kan, A.H.G. and Zipkin, P.H. (eds.), North-Holland, Amsterdam, pp. 159–131.
- Wagelmans, A., Van Hoesel, S. and Kolen, A. (1992) Economic lot sizing: an $O(n \log n)$ algorithm that runs in linear time in the Wagner-Whitin case. *Operations Research*, **40**, S145–S156.
- Wagner, H.M. and Whitin, T.M. (1958) Dynamic version of the economic lot size model. *Management Science*, **5**, 89–96.
- Zangwill, W.I. (1966) A deterministic multi-product, multifacility production and inventory model. *Operations Research*, **14**, 486–507.
- Zipkin, P.H. (2000) *Foundations of Inventory Management*, McGraw-Hill, New York.

Biographies

Phil Kaminsky is an Assistant Professor in the Industrial Engineering and Operations Research Department at the University of California, Berkeley. Prior to receiving his Ph.D. in Industrial Engineering and Management Science from Northwestern University, he worked in production engineering and control at Merck and Co. His current research focuses on the analysis and development of robust and efficient techniques for the design and operation of logistics systems and supply chains. He is a co-author (with David Simchi-Levi and Edith Simchi-Levi) of *Designing and Managing the Supply Chain: Concepts, Strategies and Case Studies* (McGraw-Hill, 1999), winner of the Book-of-the-Year Award and Outstanding IIE Publication Award given in 2000 by the Institute of Industrial Engineers. He has consulted in the areas of production and logistics system control.

David Simchi-Levi is a Professor of Engineering Systems at Massachusetts Institute of Technology. His research currently focuses on developing and implementing robust and efficient techniques for logistics systems. He has published widely in professional journals on both practical and theoretical aspects of logistics and supply chain management. He is the founder and chairman of Logic Tools (<http://www.logic-tools.com/>), a software company focusing on developing decision support

systems for solving logistics and supply chain management problems. Professor Simchi-Levi is the co-author (with Julien Bramel) of *The Logic of Logistics: Theory, Algorithms and Applications for Logistics Management*, published by Springer in 1997. His new book, *Designing and Managing the Supply Chain: Concepts, Strategies and Case Studies*, written with P. Kaminsky and E. Simchi-Levi was published by Irwin/McGraw-Hill in August 1999. The book received the Book-of-the-Year award and the Outstanding IIE Publication award given in 2000 by the Institute of

Industrial Engineers. The book also received the Outstanding First Edition of the Year award given in 2000 by McGraw-Hill. The book was selected by Business 2.0, December 2001 issue, as the best source for slashing time and cost and increasing productivity in the supply chain. The book has been translated into Chinese, Japanese and Korean.

Contributed by the Supply Chains/Production-Inventory Systems Departments

Position Announcement

Editor—*The Engineering Economist*

The Engineering Economist is a cross-disciplinary journal devoted to issues in engineering economy and co-sponsored by the Institute of Industrial Engineers and the American Society for Engineering Education. The Editorial Board of *The Engineering Economist* invites applicants for the position of Editor beginning in January 2004. This renewable three year voluntary position enjoys great prestige and a commensurate work load. The recent agreement with Taylor & Francis has outsourced the copy editing, layout, and production processes, and provided some travel support for the editor. However, some level of institutional support for the editor's activities remains essential. A previous association with *The Engineering Economist* is preferred. Applications are desired immediately. Those interested should submit an abbreviated curriculum vita outlining their qualifications, a vision statement that discusses where their leadership would take the journal, an associated set of objectives and strategies, and a discussion

on how they would define a process of continuous improvement for *The Engineering Economist*. Interested persons are urged to contact Dr. Kim Needy in advance of submitting their application. Completed applications should be sent to:

Kim LaScola Needy, Ph.D., P.E., CFPIM
University of Pittsburgh
Department of Industrial Engineering
1041 Benedum Hall
Pittsburgh, PA 15261
412-624-9838 (ofc)
412-624-9831 (fax)
kneedy@enr.pitt.edu

Review of applications will begin September 15 with an expected decision by October 20. The search for the new editor will stay open until the position is filled.