NSF grant number: DMI-0200439
NSF Program Name: Manufacturing Enterprise Systems

Models and Algorithms for Integrated Multi-Stage Production/Distribution Problems

Philip Kaminsky
Hyun-soo Ahn
Osman Engin Alper
University of California, Berkeley

Abstract: We summarize a series of models and results related to the effective integration of production and transportation decision making in a multi-stage environment. For a deterministic two stage supply chain, we develop and analyze effective optimal algorithms. For continuous time stochastic version of this model, we partially characterize the optimal policy structure, and develop heuristics. For a discrete time stochastic version of this model, we partially characterize the optimal policy structure. Finally, we introduce a model integrating production and transportation outsourcing that we are currently analyzing.

1. Introduction: For many firms, costs associated with transportation and distribution and inventory holding costs are a large percentage of total product costs. Indeed, US industry spends more than $350 Billion on transportation and more than $250 Billion on inventory holding costs annually (Lambert and Stock, 2001). In many industries, parts and subcomponents are manufactured across a variety of sites. For example, in the personal computer industry, where lowering product cost is of paramount importance, the location of final assembly is often different from the location of component assembly. Similarly, in the pharmaceutical industry, manufacturing facilities are expensive to build, and pharmaceutical products have a limited profitable life span (since after patent protection expires, generic manufacturers can manufacture the same product). Thus, 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.

These examples highlight the importance of coordinating both production and inventory policies in multi-stage supply chains. However, although there have been tremendous academic and practical efforts focused on minimizing the inventory level for components and parts within a facility while maintaining efficient production, there has been significantly less work on coordinating production and distribution **simulta-**

neously, particularly when production faces capacity constraints. Efforts in this direction are complicated by the non-linear nature of transportation costs. Indeed, over the last twenty years, driven by the adoption of JIT, CONWIP, and flexible manufacturing systems, manufacturers have devoted significant effort to setup reduction, and thus manufacturing 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 doesn't cost much less to operate than a full one so it is crucial for successful decision making approaches for multi-stage manufacturing supply chains to explicitly account for these non-linear transportation costs.

The objective of our ongoing research project is therefore to **develop models** of multi-stage production/distribution systems, **analyze** these models, and use this analysis to **develop useful managerial insights**, and **effective algorithms** for planning, controlling, and operating these systems. In this paper, we survey a variety of recent results, starting with a simple deterministic model, and continuing on to a variety of stochastic models. We conclude by sketching an outline of our ongoing exploration of third party logistics contracts, and their impact on production planning.

2. Literature Review: There has been a variety of literature on multi-stage inventory systems. Clark and Scarf (1960) show that a base stock policy is optimal in a finite horizon periodic problem with no fixed cost and no capacity constraint. Federgruen and Zipkin (1984a,1984b) show that the optimality of the order-up-to policy is still valid in the case of discounted and average infinite horizon problems. Chen (1994, 1998) demonstrates the effectiveness of the (R, nQ) policy in a similar setting, while Chen and Song (2001) characterize the structure of the optimal policy in a multi-stage inventory model with Markov modulated demand. A few authors have considered the control of capacitated production systems. Federgruen and Zipkin (1986a, 1986b) consider an inventory policy in a capacitated production system under stationary demand with no fixed cost. Glasserman and Tayur (1994, 1995, 1996) study a multistage stationary production-inventory model with capacitated production and demonstrate the effectiveness of IPA (infinitesimal perturbation analysis) as a tool to obtain an optimal policy. Kapuscinski and Tayur (1998) extend the model to periodic demand. Parker and Kapuscinsky (2002) consider a periodic review problem of a two-stage capacitated echelon inventory system with zero or deterministic leadtime and show that the optimal policy is a modified base stock policy. However, in their model, the replenishment decision is free of economies of scale in shipping quantity so that it is reasonable to conjecture that a variant of an order-up-to policy should be optimal. In most of these papers, the optimal policy is either a base stock policy, or a policy with monotone structure. An exception to this is Duenyas et al. (2003), who consider a periodic review model with fixed cost and lost sales when the system has "must-meet" deterministic demand and random demand at each period. They show that the optimal shipping quantity and rationing policy are not necessarily monotone, because the inventory on hand may not satisfy existing demands, and can be conserved for deterministic demand in subsequent periods. Our analysis indicates that the optimal policy can be very complex, as well as counter-intuitive, when both production and distribution are considered in a system with non-linear distribution costs and finite production rate. We are not aware of any research which considers similar systems.

There has been another stream of literature on the control of production systems. Ha (1997) characterizes the structure of the optimal policy in make-to-stock system as a switching curve type policy. Carr and Duenyas (2000) consider a make-to-stock/make-to-order system. Ahn et al. (1999), Iravani et al. (2003), Duenyas and Patane-Anake (1998) characterize the optimal policy in a serial production line with single or multiple resources. Duenyas and Tsai (2002) consider a two-stage production system where stage one can sell its finished goods and provide the components to the downstream stage at the same time. However, in contrast to our proposed research, very little research considers multiple units shipped at a fixed cost.

There has also, of course, been a long history of research into deterministic *single stage* single item lot sizing models, starting with the seminal works of Wagner and Whiten(1958) for the uncapacitated model, and Florian and Klein (1971) for the capacitated model. 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 don't explicitly mention transportation costs, some of the formulations are general enough to encompass our model. In addition, Chan et al. (1997) and the references therein consider a concave cost network flow version of a multistage inventory/transportation problem (without capacitated production), and develop heuristics for the problem. However, in contrast to the heuristics these authors have focused on, we focus on polynomial approaches to obtain the optimal solution.

3. The Deterministic Model: The initial deterministic model we consider consists of 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 at stage one, and capacitated production at both stages. Items are manufactured at stage one, and then held in inventory at stage one prior to shipping. Items are transported to stage two, where they are again held in inventory. Additional capacitated production is completed at stage two (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 one and stage two, as well as transportation levels between stage one and stage two, must be determined.

Formally, we consider a two-stage, n period model of a supply chain, illustrated in Figure 1. As described above, an infinite supply of raw

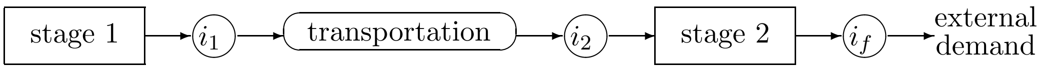

Figure 1: Model 2SPDP

material is available at stage one. Each period, $x_t^1, t = 1, 2, ..., n$ units are produced at stage 1, at a cost of c_i^t, where $x_t^1 \leq C_t^1$. Units can be held in inventory at the stage 1 buffer, where a holding cost $h_t^1, t = 1, 2, ..., n$ is charged per unit at period t. In addition, s_t units are also shipped to a buffer located before stage two (with no shipping lead time). 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 two, where a holding cost of h_t^2 is charged per unit at period t. Alternatively, $x_t^2 \leq C_t^2$ units can enter production, at a cost of c_t^2 per unit. After production, items can be held in finished goods inventory at the post stage two buffer where a holding cost h_t^f is charged per unit at period t or, or they can be shipped to meet demand d_t. We note that in this model, all demand must be met, and call this the two stage production distribution problem (2SPDP).

We summarize each period's order of events below:

1. Stage one production x_t^1 is determined, and production cost c_1^t 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 one buffer.

4. Stage two production x_t^2 is determined, and production cost c_2^t is charged per unit manufactured.

5. Holding cost h_t^2 is charged on the i_t^2 units remaining in the pre stage two buffer.

6. x_t^2 units are added to the finished goods buffer.

7. Demand is filled from the finished goods buffer.

8. Holding cost h_t^f is charged on the units remaining in the finished goods buffer.

We define the following parameters: let d_t be the external demand at period t, d_{tn} be the cumulative demand from t to period n. In addition, let h_t^1 be the holding cost per unit in the post stage one buffer at period t, h_t^2 be the holding cost per unit in the pre stage two buffer at period t, h_t^f be the holding cost per unit in the finished goods buffer at period t, C_t^1 be the production capacity at stage 1 at period t, C_t^2 be the production capacity at stage 2 at period t, f_t be the fixed cost for shipping at period t, v_t be the variable cost for shipping at period t, c_t^1 be the per unit production cost at stage 1 at period t, and c_t^2 be the per unit production cost at stage 2 at period t.

Also, let x_t^1 be the production quantity at stage one at period t, x_t^2 be the production quantity at stage two at period t, s_t be the shipping quantity at period t, y_t be the shipment indicator variable at period t, i_t^1 be the inventory level at the stage one buffer at the end of period t, i_t^2

be the inventory level at the pre stage two buffer at the end of period t and i_t^f be the inventory level at the post stage two buffer at the end of period t. We model our problem as follows:

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

$s.t.$

$$\begin{aligned}
x_t^1 &\le C_t^1 & t = 1 \dots n \, (1)\\
i_t^1 &= i_{t-1}^1 - s_t + x_t^1 & t = 1 \dots n \, (2)\\
s_t &\le y_t d_{tn} & t = 1 \dots n \, (3)\\
x_t^2 &\le C_t^2 & t = 1 \dots n \, (4)\\
i_t^2 &= i_{t-1}^2 - x_t^2 + s_t & t = 1 \dots n \, (5)\\
i_t^f &= i_{t-1}^f + x_t^2 - d_t & t = 1 \dots n \, (6)\\
i_t^f, i_t^1, i_t^2, x_t^1, x_t^2, s_t &\ge 0 & t = 1 \dots n \, (7)\\
y_t &\in \{0, 1\} & t = 1 \dots n \, (8)
\end{aligned}$$

Constraints (1) and (4) are capacity constraints for both stages. Constraints (2), (5), and (6) are the inventory balance equations for the three buffers. Constraint (3) ensures that the fixed cost is incurred when items are shipped.

In addition, we have also considered a general version of this model, in which rather than being characterized by a fixed and a linear cost, the shipping costs can be a general concave cost function: $f_t(s_t)$.

Kaminsky and Simchi-Levi(2001) present efficient algorithms for the model described above, both with the fixed and linear transportation cost, and the concave transportation cost under the following general assumptions:

- increasing holding costs:

$$h_t^1 < h_t^2 < h_t^f \quad \forall t \in 1...T.$$

- non-speculative assumptions:

$$c_t^1 + h_t^1 > c_{t+1}^1 \text{ and } c_t^2 + h_t^2 > c_{t+1}^2 \, \forall t \in 1...n-1.$$

We also assume that the following cost function holds for the case with the fixed and linear transportation cost,

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

and for the case with concave cost,:

$$f_i(x) + (h^2 - h^1)x > f_j(x) \qquad \forall i < j, x > 0.$$

In all cases, we prove the following property, which substantially reduces the complexity of the problem:

Property 1. The optimal values for $x_t^2, t = 1, 2, ..., T$ are determined by the following recursive equations:

$$x_t^2 = \min\{C_t^2, d_t + \sum_{i=t+1}^{T}(d_i - x_i^2)\} \qquad (9)$$

Indeed, this result enables us to develop a simpler 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 one. Each period, x_t^1 units are produced at stage one, where $x_t^1 \le C_t^1$, at a cost of c_t^1. Units can be held in inventory at the stage one 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 two), and a fixed cost of f_t is charged if $s_t > 0$. Units can be held in inventory in buffer two, where a holding cost of h_t^2 is charged per unit at time t or shipped to meet demand $d'_t = x_t^2, t = 1, 2, ..., n$, where x_t^2 is determined using Equation (9). Thus, we have eliminated a stage from the original model. We call this equivalent model 2SPDP′.

For the case with fixed and linear transportation cost, we further identify the following structural properties:

Property 2. In any optimal solution to 2SPDP′, shipping only occurs in periods t where the inventory in the stage two buffer, $i^2_{t-1} = 0$.

Property 3. In any optimal solution to 2SPDP′, the shipping quantity at time t, s_t is equal to some partial sum of future demands $\sum_{i=t}^{a} d_i, a \geq t$.

If we are given a production schedule, i.e, if the variables $x^1_t, t = 1, 2, ..., n$ are already determined, we can efficiently solve the problem of determining what quantity to ship in each period using a dynamic programming approach .

Given an efficient approach to this shipping problem, we define a block [s,t] to be a set of consecutive periods such that $i^1_{s-1} = 0$, $i^1_t = 0$, and $i^1_a > 0, s \leq a < t$.

This definition leads to the following two properties:

Property 4. In any optimal schedule, for any block, production will be at capacity in all periods except possibly for the first period.

Property 5. In any block, the total production in that block is equal to some partial sum of consecutive demands.

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

We have used all of the properties defined above to develop a dynamic programming approach to solving this problem. By carefully avoiding duplication of effort, we developed an algorithm which allows us to solve the original problem in $O(n^4)$ time, even when production capacity varies over time.

This problem is much more difficult in the general concave cost case. In particular, Properties 2 and 3 do not hold, so the approach we developed for the fixed+linear case will not work. However, we have developed a new (more complex) set of structural properties, and a new $O(n^8)$ algorithm for the constant capacity case (see Kaminsky and Simchi-Levi (2001)).

These results have recently been generalized by van Hoesel et al (2004).

4. Continous Time Stochastic Model: For our first stochastic model, in Ahn and Kaminsky (2004), we consider a two stage push-pull supply chain. The orders arrive at stage 2, the final stage, according to a Poisson process with rate λ. Two separate operations, which take place at different locations, are required to convert the raw materials to finished goods. We assume that an infinite supply of raw material is available at stage 1. A single server whose processing time follows i.i.d exponential distribution with rate μ_i, $i = 1, 2$ is available at each of the two stages. Items are produced at stage 1, the push stage, and can then be either held as inventory at stage 1, or shipped to stage two, in which case shipping cost are incurred. Items will be processed at stage two, the pull stage, *only to meet outstanding orders*. Non-negative holding costs are incurred on any intermediate inventory, and penalty costs are incurred for orders waiting to be filled. The holding cost is incurred at a rate h_i $(i = 1, 2)$ per unit time per item, while penalty is incurred at rate b per unit time per outstanding order. We assume that holding costs are non-decreasing (i.e., $h_1 \leq h_2 \leq b$). Shipping is assumed to be immediate, and a fixed cost, K, is incurred for each shipment to reflect shipping economies of scale. This model is illustrated in Figure 2.

We model this problem as a Markov Decision Process, and attempt to minimize the long run average cost per unit time. Without loss of generality, we uniformize λ, μ_1 and μ_2 such that $\lambda + \mu_1 + \mu_2 = 1$. The states of the MDP are $S(t) = (n_1(t), n_2(t), n_3(t))$ where $n_1(t)$ denotes the number of units held in inventory at stage 1 at time t, n_2 denotes the number of units held in inventory at stage 2 (before the completion of stage 2 operation) at time t and n_3 denotes the number of outstanding orders at time t. We assume that only Markovian stationary deterministic policies are under consideration (and call the class of these policies Π). In other words, the policies in Π specify the decision as a function of the current state only. Since processing times and interarrival times are exponential, it is easy to see that $\{S(t)\}$ is a continuous time Markov chain where the transition rate at any state is bounded by 1, and therefore $\{S(t)\}$ is uniformizable. By using the results of Lippman (1975), we can translate the original continuous time optimization problem into an equivalent (discrete time) Markov decision process with state $s = (n_1, n_2, n_3)$. Decision can be made upon the arrival of a new order or at any service completion. It is easy to see that stage 2 will always be busy as long as it is has product to process and outstanding orders to fill (i.e., $n_2 > 0$ and $n_3 > 0$). At any decision epoch, the production decision at stage 1 must be made (whether or not to initiate production of a unit at stage 1), and if $n_1 > 0$, the distribution decision must be made (i.e., the quantity greater than or equal to zero which should be shipped must be decided). Using the uniformization technique, we write the equivalent discrete time dynamic programming.

To develop insight into the structure of optimal solutions for this model, we performed computational testing using the value iteration algorithm with a variety of parameters. For the value iteration, we truncated the state space such that $n_i \leq 100$ by redirecting any transition to a state outside the truncated state space to the nearest state. Through an extensive computational study, we observed the following. In all cases, it is optimal to not ship if $n_2 > 0$. (We prove a related result below) For fixed values of n_1, the shipping decision increases monotonically in n_3. That is, there is some level of n_3, which is a function of n_2, below which it is optimal to not ship, and above which it is optimal to ship. We note that in all observed cases, this level was greater than 0. However, for fixed values of n_1, the shipping quantity fluctuates in increasing n_3. For fixed values of n_3, shipping quantity increases in n_1, and then cycles. The maximum and minimum values of the optimal shipping quantities increase with increasing n_3. This result is surprising, since it implies that in some cases it is optimal to ship a smaller quantity as the number of outstanding orders increases. Note that we expended considerable effort in eliminating truncation of countable state space and round-off error as potential reasons for this striking behavior. For example, we tried conducting a value iteration on a considerably larger state space, and extrapolating the relative value function for states outside the boundary, but the optimal policy was still non-monotonic.

Figure 3 illustrates that optimal shipping quantity is not always monotone in the amount of on-hand inventory at stage one or the number of outstanding orders. In this example, it is op-

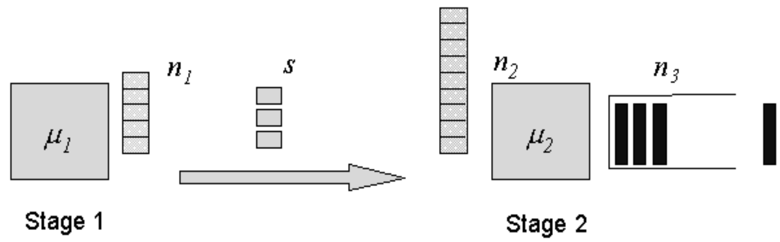

Figure 2: 2-stage Push-Pull System

timal to ship 17 units at state (19,0,3), but 14 units at state (27,0,3), even though this second state has more units to ship from stage one.

We characterize a sufficient condition for *not* shipping in this model:

Lemma 1. *For any state (n_1, n_2, n_3), it is optimal not to ship as long as $n_2 > 0$.*

This result can be easily shown using an interchange argument.

Unfortunately, the lack of monotonicity in the shipping quantities makes additional structural properties of this model difficult to determine. Thus, we are motivated to consider a variety of restrictions of this model. In the following subsection, we consider a version of this model for which the possible set of shipping quantities is limited. In Section 4.3, we characterize the impact of this restriction on the optimal objective value.

4.1 Restricting the Possible Shipping Quantities: As we observed above, the optimal policy of original formulation does not possess the monotonicity that we have hoped for. In addition, our original formulation, which considers shipping every quantity between zero and the current inventory level at each decision epoch, is computationally inefficient. We therefore consider a simplified version of the problem, in which the firm ships the minimum of the maximum transportation capacity and the current inventory level whenever the decision to ship is made. We note that this may be a reasonable restriction in practice, since firms may experience a physical limit on the quantity that can be shipped at a time (one truckload, for example). In addition, we assume that shipments only occur when inventory is zero at stage two, which we proved above is optimal in our original model, although not necessarily in this model. As we report in Section 4.3, computational testing shows that this restricted model performs about as well as the original model if the shipping quantity is selected appropriately.

At state $s = (n_1, n_2, n_3)$, the set of feasible policies, A_s is given by

$$A_s = \left\{ (i,q) | i = \begin{cases} 1, & \text{if produce;} \\ 0, & \text{otherwise.} \end{cases}, \right.$$
$$\left. q \in \left\{0, \mathbf{1}_{\{n_2=0\}} \min[n_1, Q]\right\} \right\}.$$

Using the uniformization technique, the optimal average cost, g and the relative value func-

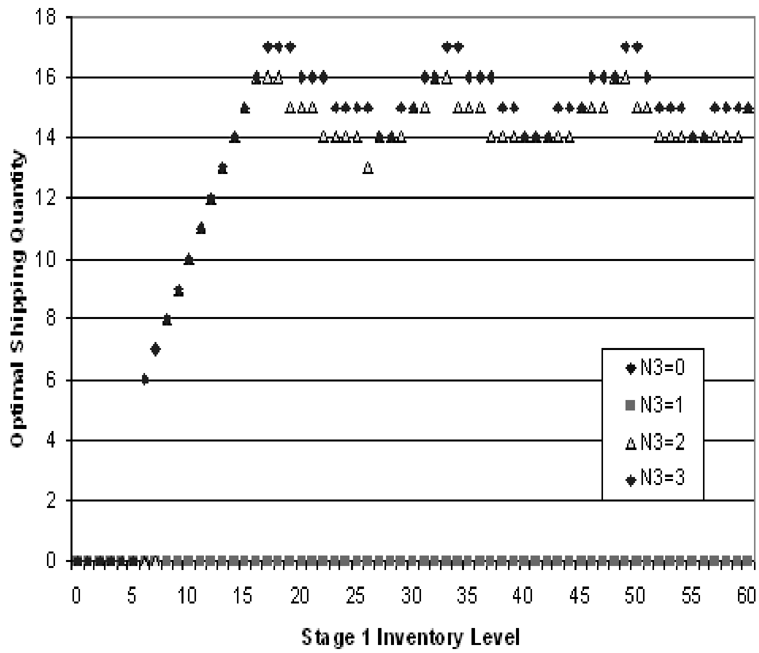

Figure 3: An example of optimal shipping quantity when $n_3 = 0, 1, 2$ and 3.

tion, $v(\cdot)$, must satisfy the following equations:

$$v(n_1,n_2,n_3)+g=\sum_{i=1}^{2}h_in_i+bn_3$$
$$+\min\left[\begin{array}{l} K+(h_2-h_1)Q \\ \ +\lambda v(n_1-Q,Q,n_3+1) \\ \ +\mu_2 v(n_1-Q,Q-1,n_3-1) \\ \ +\mu_1\min[v(n_1-Q+1,Q,n_3), \\ v(n_1-Q,Q,n_3)], \\ \lambda v(n_1,0,n_3+1)+\mu_2 v(n_1,0,n_3) \\ \ +\mu_1\min[v(n_1+1,0,n_3),v(n_1,0,n_3)] \end{array}\right]$$
$$\text{for } n_1\geq Q, n_2=0, n_3>0.$$

$$v(n_1,n_2,n_3)+g=\sum_{i=1}^{2}h_in_i+bn_3$$
$$+\min\left[\begin{array}{l} K+(h_2-h_1)n_1+\lambda v(0,n_1,n_3+1) \\ \ +\mu_2 v(0,n_1-1,n_3-1) \\ \ +\mu_1\min[v(1,n_1,n_3),v(0,n_1,n_3)], \\ \lambda v(n_1,0,n_3+1)+\mu_2 v(n_1,0,n_3) \\ \ +\mu_1\min[v(n_1+1,0,n_3),v(n_1,0,n_3)] \end{array}\right]$$
$$\text{for } n_1<Q, n_2=0, n_3>0.$$

$$v(n_1,n_2,n_3)+g=\sum_{i=1}^{2}h_in_i+bn_3$$
$$+\lambda v(n_1,n_2,n_3+1)+\mu_2 v(n_1,n_2-1,n_3-1)$$
$$+\mu_1\min[v(n_1+1,n_2,n_3),v(n_1,n_2,n_3)]$$
$$\text{for } n_2>0, n_3>0.$$

$$v(n_1,n_2,n_3)+g=\sum_{i=1}^{2}h_in_i+bn_3$$
$$+\lambda v(n_1,n_2,1)+\mu_2 v(n_1,n_2,0)$$
$$+\mu_1\min[v(n_1+1,n_2,0),v(n_1,n_2,0)]$$
$$\text{for } n_2\geq 0, n_3=0.$$

In the next subsection, we characterize the optimal policy of this model.

4.2 The Optimal Policy for the Capacitated Shipping Problem – Counterintuitive Observations: As we mentioned above, the restriction on shipping quantity simplifies computational efforts. Furthermore, as the shipping decision itself is monotone in the original problem and shipping quantity initially increases then fluctuates, we had anticipated that the monotonicity would continue to hold when we imposed a reasonable monotone shipping quantity which resembled the spirit of the optimal shipping quantity, such as $\min(n_1,Q)$. However, even this simplification does not result in a simple optimal policy. In fact, in many examples, the optimal policy is not only complex, but also counterintuitive, and very different from results found in the literature for more simple make-to-order models. This counterintuitive behavior seems to be due to the non-linear shipping costs. Later in the paper, we computationally compare this model to an analogous model with linear shipping costs. In this subsection, we discuss the structure (or lack of structure) of the optimal policy when shipping quantity is bounded by Q. Clearly, the overall performance of the system will be greatly affected by the choice of Q. In the next section, we discuss an heuristic approach to selecting a good Q value.

First, by studying a variety of computational examples, we observe that neither production or shipping decisions are monotone in most of the state variables. In particular,

1. Even if it is optimal to ship at state $(n_1,0,n_3)$, it is not necessarily optimal to ship at state $(n_1,0,n_3+1)$.

2. Even if it is optimal to ship at state $(n_1,0,n_3)$, it is not necessarily optimal to ship at state $(n_1+1,0,n_3)$, even in the case

when $n_1 > Q$ so the shipping quantity remains the same.

3. The production policy at stage 1 is not necessarily monotone in the inventory level at stage 1, that is, even if it is optimal not to produce at state (n_1, n_2, n_3), it may be optimal to produce at state $(n_1 + 1, n_2, n_3)$.

4. The production policy at stage 1 is not necessarily monotone in the number of outstanding orders. Even if it is optimal to produce at state (n_1, n_2, n_3), it is not necessarily optimal to produce at state $(n_1, n_2, n_3 + 1)$.

While in some examples the optimal policy is monotone (mostly when fixed shipping cost is close to zero or small), there are many other examples showing that such monotonicity does not hold in general. As can be seen in Figure 4, the optimal policy can not be characterized by a simple switching curve. First, consider the lack of monotonicity in n_3. In the example, when $n_1 = 10$, it is optimal not to ship when $n_3 \in \{0, 1, 5, 6, 7\}$, but optimal when n_3 takes on any other value. Although this seems counterintuitive, consider the following possible explanation. In the optimal policy, units shipped from stage 1 cover current and (possibly) future outstanding orders. Any shipment decision balances increased holding cost in stage two with backorder cost. In addition, any shipment less than Q also accounts for increased shipping costs. Also, once a shipment is made, the next shipment cannot be made until inventory at stage 2 is once again zero, and the time until inventory at stage two is once again zero depends on the backorder level during the first shipment, and the amount shipped. For very low backorder levels ($n_3 = 1$ in this example), increased holding and transportation costs offset increased backorder costs, so it is optimal not to ship. For certain low backorder levels ($n_3 = 2$ in this example), the benefit from shipping, decreased backorder costs, may outweigh the benefits of waiting and producing more before shipping. This effect is enhanced by the fact that backorder level is likely to be relatively low at the time of the next shipment, since the amount being shipped is relatively small, and much of it is already allocated to existing backorders. However, for other small backorder levels ($n_3 = 5, 6, 7$ in this example), the benefit of delaying shipment (consequently shipping more in a later time) dominates the benefit of shipping immediately. Finally, as the backorder level gets larger, it again makes sense to ship immediately, as the backorder level at the next shipment will be large regardless of whether or not some quantity is immediately shipped.

The non-monotonicity of shipping in n_1 can be explained similarly. When there are few units in stage 1, shipping immediately is suboptimal since doing so fails to take advantage of the economies of scale in shipping. However, as the number of units in stage 1 increases, the shipping cost per unit decreases. However, at some point, increasing stage 1 holding cost may begin to dominate, so that it therefore becomes optimal to wait until additional orders arrive before shipping.

Figure 5 shows an example where the optimal production policy is not monotone in backorder levels. For instance, it is optimal to produce for $n_3 \leq 3$ or $n_3 \geq 8$, but optimal not to produce for $4 \leq n_3 \leq 7$ when $n_1 = 7$. For low backorder levels, it is optimal to produce, but not to ship since backorder cost is not high enough to justify immediate shipping (with resultant increased per unit shipping charges). However, as

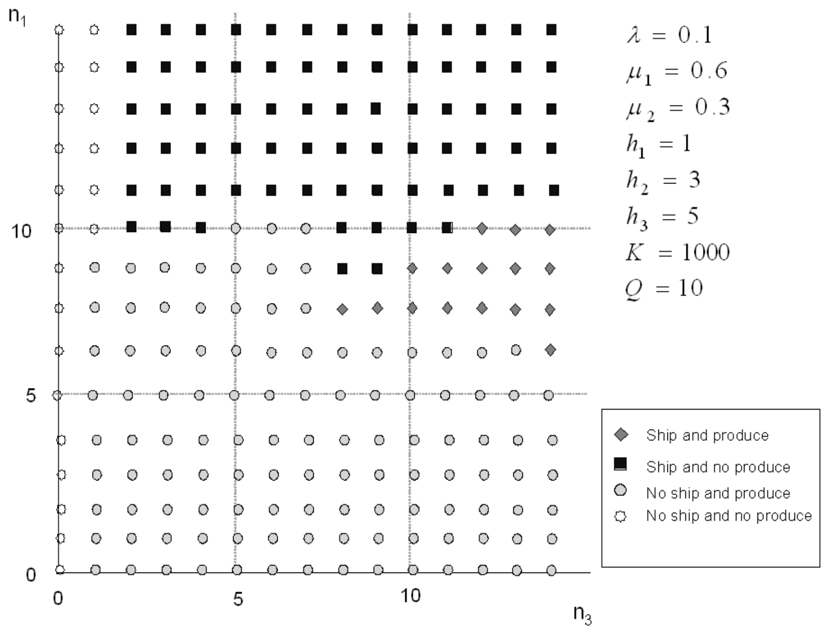

Figure 4: An optimal policy at $n_2 = 0$ in a capacitated shipping problem.

the backorder level gets larger, it makes sense to ship immediately for $n_3 \geq 4$ as in Figure 5. While it is optimal not to produce after a shipment for medium backorder levels ($n_3 = 4, 5, 6, 7$ in this example), it is optimal to produce after a shipment for high backorder levels. This is true because with high backorder, the time until the next time inventory is again zero at stage two decreases. Thus, production must continue to ensure that there is enough material to include with the next shipment. On the other hand, with lower backorder, it will be longer until the next shipment is required. Indeed, the time at which inventory at stage two is zero after a shipment stochastically decreases in the number of outstanding orders.

This phenomenon is affected by the relative production and arrival rates. When the production rate at stage 1 is higher than the rate at stage 2, it is more likely to be optimal to wait before producing at stage one, as the time until another shipment is required is greater. As backorder costs increase, the value of waiting until the next shipment decreases, so production is more likely, although only if relative processing rates suggest that this production is likely to be shipped relatively soon.

This example contradicts the intuition established by the optimal solution to many make-to-order models. The conventional intuition suggests that the optimal production rate should increase as the number of outstanding orders increases so that the system can clear orders as soon as possible. Therefore, as the backorder cost becomes high, it becomes urgent to reduce outstanding orders. However, at the push-pull boundary, the finite capacity at stage 2 dampens such urgency. When there is enough time to produce units at stage 1 while stage 2 is busy, the production can be delayed. On the other hand, the benefit of increasing shipping quantities increases when backorder cost is small and shipping cost is high. Therefore, it becomes optimal not to ship, but to produce when there are few backorders.

The non-monotonicity of optimal production policy in n_1 is even more intriguing. Observe in Figure 6 that it is optimal to produce at state $(9, 0, 2)$ although it is optimal not to produce at state $(8, 0, 2)$. Most previous analysis of related models implies that it is optimal to produce until the inventory level reaches a certain point (although this point may be state dependent). This example contradicts that intuition, since it is optimal to produce at a particular state, $(9, 0, 2)$, but not at a state with lower inventory, $(8, 0, 2)$. In general, this counterintuitive behavior seems to occur when n_2 is low. In fact, when $n_2 = 0$ or 2, the production policy is characterized by a monotone switching curve in the opposite direction to what intuition may suggest

Although this seems perplexing, consider the following possible explanation. When the fixed shipping cost is extremely high as in this example, it is optimal to ship Q units when a shipment occurs. Therefore, stage 1 will always produce Q units before a shipment occurs. Now, if holding costs are similar at stage 1 and stage 2, it makes little difference if inventory is held at stage 1 or at stage 2. Thus, when the inventory level at stage 1 is close to Q, it becomes optimal to produce more units (up to Q), then ship to stage 2, rather than storing inventory at stage 1. However, when the inventory level at stage 1 is low, it is optimal to wait for more outstanding orders before producing additional units to ship. Thus, in this case, the optimal policy is monotone in the counter-intuitive direction (i.e., the more inventory there is, the more valuable it is to produce more.) However, as the level of inven-

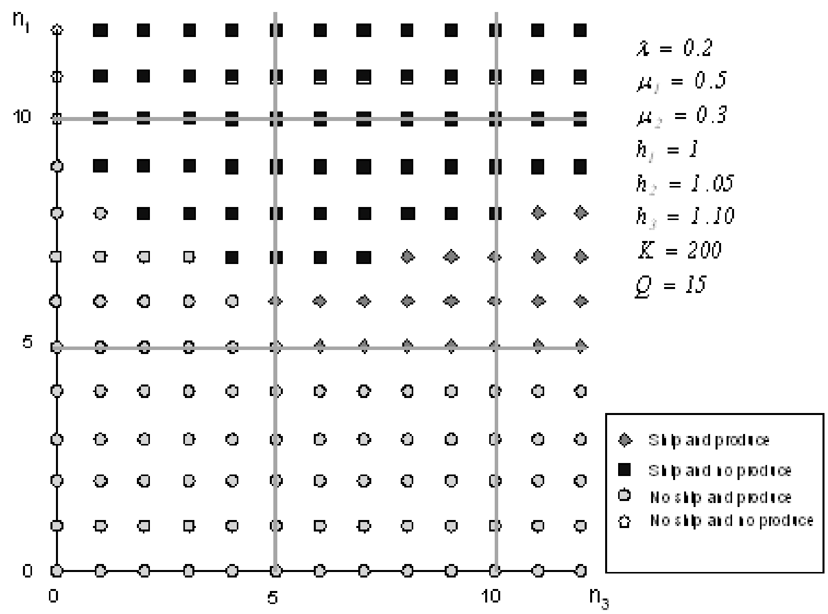

Figure 5: The non-monotonicity of production policy in n_3 when $n_2 = 0$.

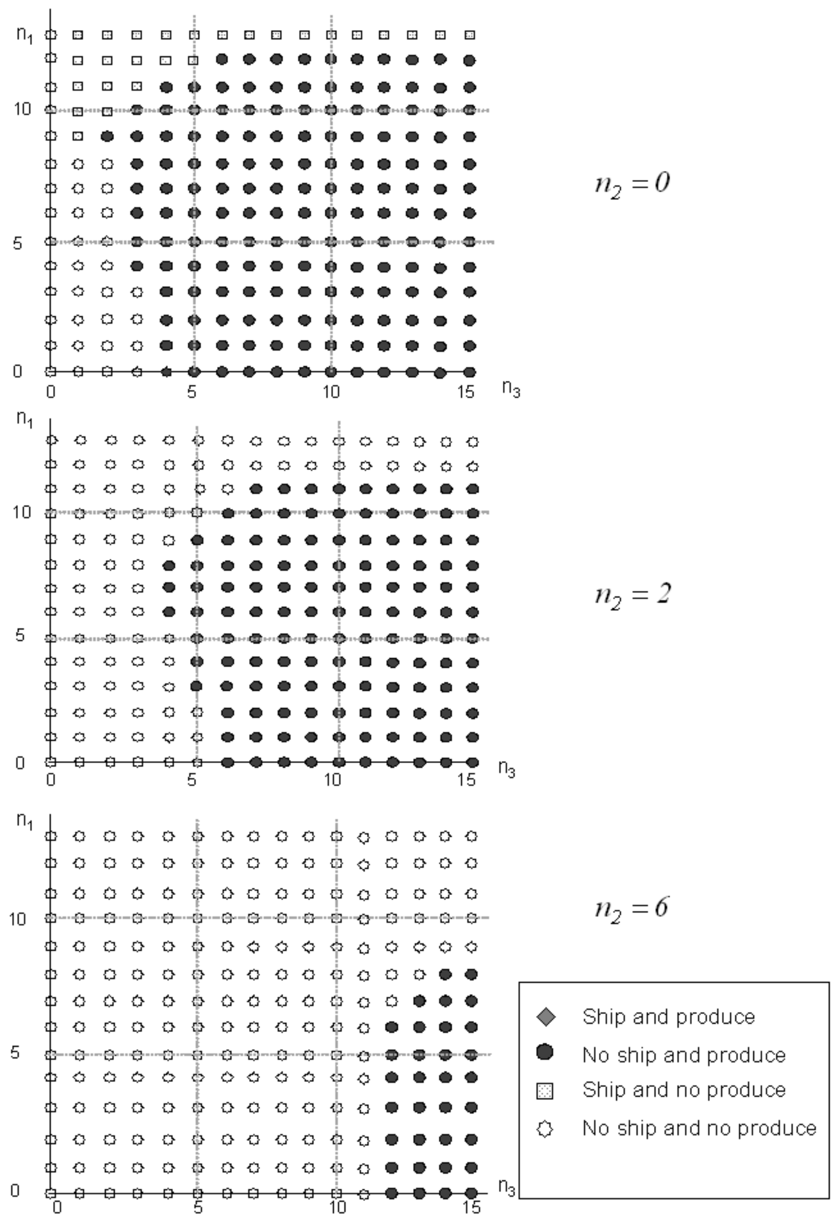

Figure 6: The non-monotonicity of production policy in n_1 ($\lambda = .05$,$\mu_1 = .50$, $\mu_2 = .45$, $h_1 = 1.0$, $h_2 = 1.05$,$b = 1.10$, $K = 1000.0$,and $Q = 12$)

tory at stage 2 increases, this effect diminishes, as the time until the next shipment increases. In Figure 6, for example, this "opposite direction monotonicity" disappears when $n_2 \geq 6$.

Although the optimal policy to this problem is clearly complex, we are able to partially analytically characterize its structure. We show that it is always optimal to ship if there are at least Q units of inventory and at least Q units of backorder.

Lemma 2. *For any state $(n_1, 0, n_3)$ such that $n_1 \geq Q$ and $n_3 \geq Q$, it is optimal to ship independent of production policy.*

Proof. We prove the claim by a sample path argument that any policy with delayed shipping can be improved by a policy with immediate shipping. Without loss of generality, we assume that $t = 0$. Let w be a sample path in a probability space $\mathbf{P}$ large enough to contain all future arrivals and service times. Suppose that in the optimal policy Π^*, shipping is delayed by $t_s > 0$. Let T denote the first time that $n_2(t)$ becomes zero after shipping Q units at time t_s on a given sample path, w. Also, let $p^{\Pi^*}(t)$ be the cumulative units produced in stage 1 by time t,$D^{\Pi^*}(t)$ be the cumulative units produced in stage 2 at time t under policy Π^*, and $A(t)$ be the cumulative number of arrivals to be system by time t. It can be easily shown that the stage 2 remains idle until time t_s (that is when the first shipment occurs) and works to produce Q consecutive units without idling. The process will reaches state $(n_1 + P^{\Pi}(T) - Q, 0, n_3 + A(T) - Q)$ under policy Π^* at time T and the total cost accumulated over this interval $[0, T]$, denoted as $C^{\Pi^*}((0, T]; w)$, can be expressed as the following:

$$C^{\Pi^*}((0,T];w) = \int_0^{t_s} \Big[h_1(n_1 + p^{\Pi^*}(t)) + b(n_3 + A(t))\Big]\, dt + \int_{t_s}^{T} \Big[h_1(n_1 + p^{\Pi^*}(t) - Q) + h_2(Q - D^{\Pi^*}(t)) + b(n_3 + A(t) - D^{\Pi^*}(t))\Big]\, dt$$

Consider another policy Π on the same sample path, that mimics the production of policy Π, but ships Q units at time zero instead of t_s. Under this policy, Q units are produced without idling and idling occurs in the interval $(T - t_s, T]$ at stage 2 (i.e., $D^{\Pi}(t) = D^{\Pi^*}(t + t_s)$ for $t \in [0, T - t_s]$ and $D^{\Pi}(t) = D^{\Pi^*}(T) = Q$ for $t \in [T - t_s, T]$.) After T, policy Π mimics the production and shipping decisions of policy Π^*. For every sample path, it can be easily shown that the sample paths under both policies coincide after T. The total cost accrued over this interval $[0, T]$, denoted as $C^{\Pi}((0, T]; w)$, is

$$\begin{aligned} C^{\Pi}((0,T];w) &= \int_0^{T} \Big[h_1(n_1 + p^{\Pi}(t) - Q) + b(n_3 + A(t) - D^{\Pi}(t))\Big] dt + \int_0^{T-t_s} \big[h_2(Q - D^{\Pi}(t))\big]\, dt \\ &= \int_0^{T} \Big[h_1(n_1 + p^{\Pi^*}(t) - Q)\Big]\, dt + \int_0^{T-t_s} \Big[b(n_3 + A(t) - D^{\Pi^*}(t + t_s)) + h_2(Q - D^{\Pi^*}(t + t_s))\Big] dt \\ &\quad + \int_{T-t_s}^{T} \big[b(n_3 + A(t) - Q)\big]\, dt. \end{aligned}$$

Comparing the cost associated with policy Π^* and the cost associated with policy Π, we have

$$C^{\Pi^*}((0,T];w) - C^{\Pi}((0,T];w) = \int_0^{t_s} Q(h_1 + b)dt \geq 0.$$

The result holds for every sample path. Therefore, for any policy under which the shipping is delayed, it is possible to construct an immediate shipping policy which is better for every sample path. This contradicts the optimality of policy Π^*. $\square$

Of course, the performance of this restricted model relative to the original model depends on the selection of a good Q value. In Section 4.3, we demonstrate that if we select the best possible Q value, the gap between the average cost of this policy and the average optimal cost is very small.

4.3 Computational Analysis: Our computational experiments are designed to characterize the loss associated with restricting the shipping quantity to the minimum of some Q value and the available inventory and characterize the impact of non-linear shipping costs. We tested a variety of combinations of system parameters:

- $h_1, h_2, b \in \{(1, 2, 5), (1, 4, 15)\}$
- $K \in \{250, 1000, 5000\}$
- $\lambda, \mu_1, \mu_2 \in \{(.2, .4, .4), (.15, .6, .25), (.15, .25, .6)\}$

We vary costs so that they increase a relatively small amount and a relatively large amount between stages, we consider a variety of fixed costs, and we consider processing rates which are much faster at stage one, and much faster at stage two. All of the examples are solved through a value iteration on a sufficiently large truncated state space to alleviate any boundary effect.

For each of the combinations of system parameters listed above, we tested a variety of Q values $Q = 1, 2, ..., 70$. Table 1 displays these results. Note that we essentially lose nothing by making this assumption. Indeed, in all cases the objective value for the best possible Q is almost identical to the optimal objective value for the original model. We have to consider four decimal digits to see a difference in most cases. In addition, this result is relatively insensitive to Q. Figure 7 graphs objective value versus Q value for one sample problem (The problem instance is: $h_1 = 1$, $h_2 = 2$,$b = 5$, $K = 250$, $\lambda = .2$, $\mu_1 = .4$,$\mu_2 = .4$). Observe that the objective value is very close to optimal for a large range of Q values. The performance of all the sample problems we considered was similar.

To determine the impact of non-linear shipping costs on total system cost, in the last column of Table 1, we present system cost if shipping cost is linear, and equal to K/Q where Q is the optimal Q value for the restricted shipping problem. In all cases, this system is significantly less expensive than the system with fixed shipping costs.

5. A Discrete Time Stochastic Model: In many cases, a discrete time model is more applicable. In addition, we were motivated to explore the discrete time model to help us understand which of the observations made in the continuous time case are general, and which seem to be specifically related to the continous time nature of that model. In this case, we consider a manufacturing system consisting of two stages facing stochastic demand through n periods. An infinite supply of raw material is available at stage one. Production at stage one is uncapacitated while the one at stage two is capacitated. Items are manufactured at stage one, and then held in inventory at stage one prior to shipping. Transported items are held in inventory at stage two until additional production is completed and then the external demand is met.

Our problem is to determine the production

h_1	h_2	b	K	λ	μ_1	μ_2	Optimal	Restrict. Ship	Linear
1	2	5	250	0.2	0.4	0.4	22.2961	22.2966	10.5471
1	4	15	250	0.2	0.4	0.4	38.6485	38.6502	29.4737
1	2	5	1000	0.2	0.4	0.4	37.7050	37.7056	14.5638
1	4	15	1000	0.2	0.4	0.4	59.7015	59.7025	34.9888
1	2	5	4000	0.2	0.4	0.4	68.0791	68.0804	26.5500
1	4	15	4000	0.2	0.4	0.4	101.1060	101.1076	50.8105
1	2	5	250	0.15	0.6	0.25	21.7331	21.7344	14.4781
1	4	15	250	0.15	0.6	0.25	42.1668	42.1690	33.8936
1	2	5	1000	0.15	0.6	0.25	34.5979	34.599685	18.6450
1	4	15	1000	0.15	0.6	0.25	60.1309	60.133752	40.1438
1	2	5	4000	0.15	0.6	0.25	59.9318	59.934515	26.9787
1	4	15	4000	0.15	0.6	0.25	95.6139	95.6182	52.6442
1	2	5	250	0.15	0.25	0.6	16.9865	16.9876	9.2840
1	4	15	250	0.15	0.25	0.6	26.3363	26.3505	21.8587
1	2	5	1000	0.15	0.25	0.6	29.8138	29.8154	11.1591
1	4	15	1000	0.15	0.25	0.6	43.6287	43.6309	24.3597
1	2	5	4000	0.15	0.25	0.6	55.4480	55.4504	26.1596
1	4	15	4000	0.15	0.25	0.6	78.3368	78.3406	41.6330

Table 1: Performance When Shipping Quantity is Restricted, and with Linear Shipping Costs

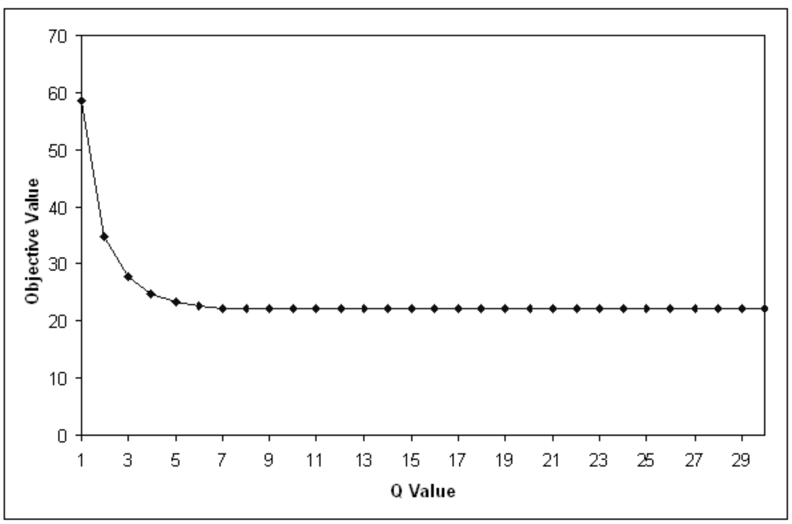

Figure 7: Q vs. Objective for restricted shipping.

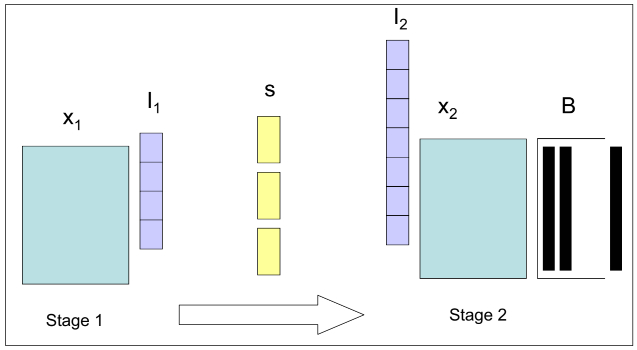

Figure 8: Two-Stage System

level at first stage, as well as the transportation level between stage one and stage two at each period. Notice that the first stage is operated in a push-based manner, utilizing the economies of scale in transportation. On the other hand the second stage is clearly a pull-based stage in which the production level is directly determined by the external demand. Although production is unrestricted at stage one, shipment from stage one to stage two is restricted by the amount of inventory at stage one. Similarly the production at stage two is limited by the amount of inventory plus shipment at stage two as well as the production capacity at stage two.

5.1 The Sequence of Events: In each (discrete) time period, the following events take place in the following order:

1. Make shipment and production decisions.
2. Receive shipment at stage two inventory.
3. Realize demand, update inventory position at stage two.
4. Receive production at stage one inventory.
5. Calculate the costs.

5.2 Notation and Formulation: We employ the following notation in our analysis.

I_t^1: Inventory level at stage one at the beginning of period t.
I_t^2: Inventory level at stage two at the beginning of period t.
B_t: Backlog level at stage two at the beginning of period t.
x_t^1: Production level at stage one at period t.
x_t^2: Production level at stage two at period t.
s_t: Shipment level from stage one to stage two at period t.
h_1: Unit holding cost of inventory at stage one.
h_2: Unit holding cost of inventory at stage two.
p: Unit penalty cost of backlogged demand.
c_1: Unit production cost at stage one.
c_2: Unit production cost at stage two.
C: Production capacity of stage two.
K: Fixed cost for shipping.
N: Number of periods.
d_t: External demand (random disturbance) realized at period t.
α: Discount factor.

$\vec{I_t}$: State vector.
$\vec{u_t}$: Control vector.
D_t: Random disturbance space.
S_t: State space.
C_t: Control space.
U_t: State constrained control space.
π: Policy (control law).
μ_t: Control function.

We employ this notation to define the following dynamic program:

$$d_t \in D_t, \text{ where } D_t \subseteq Z^+.$$
$$\vec{I_t} = (I_t^1, I_t^2, B_t) \in S_t, \text{ where } S_t \subseteq Z^{3+}.$$
$$\vec{u_t} = (x_t^1, s_t) \in C_t, \text{ where } C_t \subseteq Z^{2+}.$$
$$\vec{u_t} \in U_t(\vec{I_t}), \text{ where}$$
$$U_t(\vec{I_t}) = \{(x_t^1, s_t) \in C_t | s_t \leq I_t^1\}.$$
$$\pi = \{\mu_0, ..., \mu_{N-1}\}, \vec{u_t} = \mu_t(\vec{I_t}), \text{ where}$$
$$\mu_t(\vec{I_t}) \in U_t(\vec{I_t}), \text{ for all } \vec{I_t} \in S_t.$$
$$\vec{I_{t+1}} = f_t(\vec{I_t}, \mu_t(\vec{I_t}), d_t), \text{ where}$$
$$f_t((I_t^1, I_t^2, B_t), (x_t^1, s_t), d_t) = [I_t^1 + x_t^1 - s_t, I_t^2 + s_t - min(B_t + d_t, I_t^2 + s_t, C), B_t + d_t - min(B_t + d_t, I_t^2 + s_t, C)].$$
$$\min J_\pi(\vec{I_0}) = E\left\{\sum_{t=0}^{N-1} \alpha^t g_t(\vec{I_t}, \mu_t(\vec{I_t}), d_t)\right\}, \text{ where}$$
$$g_t((I_t^1, I_t^2, B_t), (x_t^1, s_t), d_t) = (I_t^1 + x_t^1 - s_t)h_1 + [I_t^2 + s_t - min(B_t + d_t, I_t^2 + s_t, C)]h_2 + [B_t + d_t - min(B_t + d_t, I_t^2 + s_t, C)]p + K\mathbf{1}_{(s_t>0)} + x_t^1 c_1 + min(B_t + d_t, I_t^2 + s_t, C)c_2.$$
$$J_N(\vec{I_N}) = 0, \text{ and}$$
$$J_t(\vec{I_t}) = min_{\vec{u_t}}\Big\{x_t^1 c_1 + K\mathbf{1}_{(s_t>0)} + (I_t^1 + x_t^1 - s_t)h_1 + h_2 E[I_t^2 + s_t - min(B_t + d_t, I_t^2 + s_t, C)] + pE[B_t + d_t - min(B_t + d_t, I_t^2 + s_t, C)] + c_2 E[min(B_t + d_t, I_t^2 + s_t, C)] + \alpha E\Big[J_{t+1}\Big(f_t(\vec{I_t}, \mu_t(\vec{I_t}), d_t)\Big)\Big]\Big\}.$$

5.3 Results: To gain insight into the structure of optimal policy for this model, we performed computational testing using the value iteration algorithm with a variety of parameter values. For the value iteration, we truncated the state space with a conservative upper bound, redirecting any transition to a state outside the truncated space to the nearest state in. Through an extensive computational study, we observed that neither optimal production nor shipping decisions are monotone in most of the state variables. The following examples are taken from the optimal solution to a given set of particular parameters.

1. The production level at stage one is not monotone in outstanding orders. For example, while $x = 9$ at (9,8,8); $x = 0$ at (9,8,9).
2. The production level at stage one is not monotone in inventory level at stage one. For example, while $x = 0$ at (10,8,5); $x = 5$ at (11,8,5).
3. The production level at stage one is not monotone in inventory level at stage two. For example, while $x = 0$ at (9,8,9); $x = 9$ at (9,9,9).
4. The shipment level is not monotone in outstanding orders. For example, while $s = 9$ at (9,8,12); $s = 1$ at (9,8,13).
5. The shipment level is not monotone in inventory level at stage two. For example, while $s = 7$ at (13,2,8); $s = 13$ at (13,3,8).

As can be observed from the examples, the optimal policy is not only complex but also counterintuitive. This counterintuitive behavior seems to stem from the nonlinear shipping costs. The production capacity at stage two further complicates the structure of the optimal policy.

We were able to prove the following two propositions, which are fairly intuitive.

Proposition 1. *$I_t^2 \geq C$ implies $s_t = 0$.*

Proof. For sake of contradiction, suppose that in the optimal policy, π, in period t the shipment level is positive although $I_t^2 \geq C$. Consider the policy, $\overline{\pi}$, which is the same as π except $\overline{s}_t = 0$, and $\overline{s}_{t+1} = s_{t+1} + s_t$. But $J_\pi(\vec{I_0}) - J_{\overline{\pi}}(\vec{I_0}) = s_t(h_2 - h_1) > 0$, which contradicts the optimality of π. □

Proposition 2. *$I_t^2 \geq 2C$ implies $x_t = 0$.*

Proof. For sake of contradiction, suppose that in the optimal policy, π, in period t the production level is positive although $I_t^2 \geq 2C$. Consider the policy, $\overline{\pi}$, which is the same as π except $\overline{x}_t = 0$, and $\overline{x}_{t+1} = x_{t+1} + x_t$. Notice that the optimal shipment level at the next period is zero, $s_{t+1} = 0$, by the previous result; thus such a policy, $\overline{\pi}$, exists. But $J_\pi(\vec{I_0}) - J_{\overline{\pi}}(\vec{I_0}) = x_t h_1 > 0$, which contradicts the optimality of π. □

5.4 Modifications of the Initial Model: The lack of monotonicity in both the shipping and production levels makes the structural properties of this model very difficult to determine. Thus, we are motivated to consider various modifications of this model in the hope of better characterizing optimal policy structure. We considered the following:

1. In this model, we restrict shipment to either zero or the minimum of the inventory at hand and Q, where Q is an additional parameter specifying the shipment upper bound.
2. In this model, we restrict production to an (s,S) type policy. Hence the production decision is eliminated from the problem, reducing the dimension of the control space to one.
3. In this model, we restrict shipment to an (s,S) type policy; i.e. whenever the inventory at stage two drops below s $min(S - I_t^2, I_t^1)$ is shipped, otherwise there is no shipment. Hence the shipment decision is eliminated from the problem, reducing the dimension of the control space to one.
4. In this model, we restrict shipment to either Q (given $I_t^1 \geq Q$) or zero, where Q is an additional parameter specifying the shipment level. Hence the shipment decision becomes a binary one.

For all the modified models, we performed the same computational testing using the value iteration algorithm with a variety of parameter values. Although the computational effort is substantially reduced by all of the modified models, the structure of the optimal policy is hardly simplified by the first three of them. The optimal policy structure of the fourth modified model is remarkably well behaved compared to the others, and we were able to partially characterize the optimal policy structure more extensively than in the other cases. Unfortunately, the complete characterization of the optimal policy structure eluded us in all of the models we considered. It appears that both discrete and continuous time stochastic versions of this model behave quite counter-intuitively, and optimal polices are in both cases very complex.

6. Third Party Logistics Contracting: Finally, building on prior research, over the next year we will investigate third party logistics contracting in a multi-stage production/distribution environment. As we have argued above, in a supply chain, production and distribution operations are generally the most important operational functions. It is crucial to integrate these two functions by planning and scheduling them jointly in a coordinated way to achieve optimal operational performance of the supply chain. On the other hand, outsourcing non-core competencies has now become a widely accepted practice across many industries. Running an inhouse trucking fleet is a difficult task and not a core competency. Hence, acquiring transportation contracts and coordinating production and distribution operations with them have become critical operational and tactical points of interest in a supply chain.

Transportation contracts in the modern era often specify in advance the frequency and volume to be reserved by the carrier for a particular customer's future deliveries. The most common types of transportation contracts in research literature and practice are the following:

1. Long-term (or forward buy or fixed commitment) contracts, which specify a fixed amount of service to be delivered at some point in the future.

2. Option contracts, in which the buyer prepays the reservation price (or premium) for the service capacity and then pays the execution (or exercise) price for each unit of capacity used.

3. Flexibility contracts, in which a fixed amount of supply is determined when the contract is signed, but the amount to be delivered and paid for can differ by no more than a given percentage established upon signing the contract. These contracts are used in practice to share risk between the parties, and are equivalent to a relevant combination of a long-term contract and an option contract.

We are interested in developing and analyzing mathematical models of supply chain networks integrating production and distribution operations through the use of transportation contracts. In this general setting we plan to consider the following research problems:

1. We will consider a manufacturer in possession of a transportation contract of one of the types described above over a given finite horizon. This manufacturer faces stochastic demand, and has access to a spot market for expedited shipping. At each period, given the demand at that period the manufacturer have to decide between immediate shipment and consolidating orders to utilize the transportation contract. We aim to determine the optimal operation of the production and distribution functions of the manufacturer in this setting, while gaining important practical insights for supply chain operation.

2. Given the solution of the first problem, we are interested in jointly optimizing the acquisition and the operation of the transportation contracts. For this purpose, we assume a large pool of available transportation contracts with various schedules and contract parameters.

3. Given the solution of the second problem, now intend to consider the problem from the

transportation provider's point of view and tackle the questions of designing and pricing transportation contracts in a market with a pool of buyers under competition.

References:

Aggarwal, A. and J. K. Park. 1990. Improved Algorithms for Economic Lot-Size Problems. Working Paper, Laboratory for Computer Science, MIT, Cambridge, MA.

Ahn. H., I. Duenyas and R. Zhang. 1999. Optimal Stochastic Scheduling of a 2-Stage Tandem Queue with Parallel Servers. *Advances in Applied Probability*, **38**, pp. 1095-1117.

Ahn, H-S, and P. Kaminsky. 2004. Production and Distribution Policy in a Two Stage Stochastic Push-Pull Supply Chain. *Submitted for publication*.

Armentano, V.A., Esther Berretta, R., and P.M. Franco. 2001. Lot Sizing in Capacitated Multi-Stage Serial Systems. To appear in *Production and Operations Management.*

Baker, K.R. 1993. Requirements Planning. In *Handbooks in Operations Research and Management Science*, the volume on Logistics of Production and Inventory, Graves, S.C., A.H.G. Rinnooy Kan, and P.H.Zipkin (eds.), North- Holland, Amsterdam, pp. 571-627.

Baker, K.R., P. Dixon, M.J. Magazine, and E.A. Silver. 1978. An Algorithm for the Dynamic Lot-Size Problem with Time-Varying Production Capacity Constraints. *Management Science* **24**, pp. 1710-1720.

Carr. S. and I. Duenyas. 2000. Optimal Admission Control and Sequencing in aMake-To-Stock/Make-To-Order Production System. *Operations Research*, **48**, pp. 709-720.

Chan, L. M. A., A. Muriel and D. Simchi-Levi. 1997. Supply Chain Management: Integrating Inventory and Transportation. Working Paper, Northwestern University.

Chen, F. and Y.-S. Zheng. 1994. Evaluating Echelon Stock (R, nQ) Policies in Serial Production/Inventory Systems with Stochastic Demand. *Management Science*, **40**, pp. 1262-1275.

Chen, F. 1998. Echelon Reorder Points, Installation Reorder Points, and the Value of Centralized Demand Information. Management Science 44 (12, part 2), S221-S234.

Chen, F. and J.-S. Song. 2001. Optimal Policies for Multi-Echelon Inventory Problems with Markov Modulated Demand. *Operations Research*, **49**, pp. 226-234.

Clark, A. J. and H. Scarf. 1960. Optimal Policies for a Multi-echelon Inventory Problem. *Management Science*, **6**, pp. 475-490.

Duenyas, I. and P. Patane-Anake. 1998. Base-Stock Control for Single Product Tandem Make-to-Stock Systems. *IIE Transactions*, **30**, pp. 31-39.

Duenyas, I., K. Frank and R. Zhang. 2003. Optimal Inventory Policies in Systems with Priority Demand Classes. To appear in Operations Research.

Duenyas, I. and C. Tsai. 2002. Centralized and Decentralized Control of a Two-Stage Tandem Manufacturing System with Demand for Intermediate and End Products, , Working Paper.

Federgruen, A. and M. Tzur. 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**, pp. 909-925.

Federgruen, A. and P. Zipkin. 1984a. Computational issues in an infinite-horizon, multiechelon inventory model. *Operations Re-*

search, **32**, pp. 818-836.

Federgruen, A. and P. Zipkin. 1984b. An efficient algorithm for computing optimal (s,S) policies. *Operations Research*, **32**, pp. 1268-1285.

Federgruen, A. and P. Zipkin. 1986a. An Inventory Model with Limited Production Capacity and Uncertain Demands I: The Average-cost Criterion. *Mathematics of Operations Research*. **11**. 193-207.

Federgruen, A. and P. Zipkin. 1986b. An Inventory Model with Limited Production Capacity and Uncertain Demands II: The Discounted-cost Criterion. *Mathematics of Operations Research*. **11**. 208-215.

Florian, M. and M.Klein. 1971. Deterministic Production Planning with Concave Costs and Capacity Constraints. *Management Science*, **18**, pp. 12-20.

Glasserman, P. and S. Tayur. 1994. The Stability of a Capacitated, Multi-Echlon Production-inventory System. *Operations Research*. **42**. pp. 913-925.

Glasserman, P. and S. Tayur. 1995. Sensitivity Analysis for Base-Stock Levels in Multi-Echelon Production-Inventory System. *Management Science*. **42**. pp. 263-281.

Glasserman, P. and S. Tayur. 1996. A Simple Approximation for Multi-Stage Capacitated Production-Inventory System. *Naval Res. Logist.* **43**. pp. 41-58.

Ha, A. 1997. Optimal Dynamic Scheduling Policy for a Make-to-Stock Production System. *Operations Research*, **43**, pp. 1093-1103.

Harrison, T.P., and H.S.Lewis. 1996. Lot Sizing in Serial Assembly Systems with Multiple Constrained Resources. *Management Science*, **42**, pp. 19-36.

Iravani, S. M. R., Posner, M. J. M. and J. A. Buzacott. 2003. Scheduling of a Moving Server in a Serial Production Process. Working Paper. Northwestern University.

Kaminsky, P. and D. Simchi-Levi. 2001. Production and Distribution Lot Sizing in a Two-Stage Supply Chain. To appear in IIE Transactions.

Katok, E., H.S. Lewis, and T.P. Harrison. 1998. Lot Sizing in General Assembly Systems with Setup Costs, Setup Times, and Multiple Constrained Resources. *Management Science*, **44**, pp. 859- 877.

Kaupscinski, R. and S. Tayur. 1998. A Capacitated Production-Inventory Model with Periodic Demand. *Operations Research*, **46**, pp. 899-911

Love, S.F. 1973. Bounded Production and Inventory Models with Piecewise Concave Costs. *Management Science*. **11**, pp. 313-318.

Lambert, D. M. and J.R. Stock. 2001. *Strategic Logistics Management, Fourth Edition.* McGraw-Hill, NY.

Lippman, S.A. 1975. Applying a New Device in the Optimization of Exponential Queuing Systems. *Operations Research* Res. 23, pp. 686-710.

Muckstadt, J.M. and R.O. Roundy. 1993. Analysis of Multistage Production Systems. In *Handbooks in Operations Research and Management Science*, the volume on Logistics of Production and Inventory, Graves, S.C., A.H.G. Rinnooy Kan, and P.H.Zipkin (eds.), North- Holland, Amsterdam, pp. 59-131.

Parker, R. and R. Kapuscinski. 2002. Optimal Inventory Policies for a Capacitated Two-Echelon System, Working paper.

Wagelmans, A. S. Van Hoesel and A. Kolen. 1992. Economic Lot Sizing: An $O(n \log n$

Algorithm that Runs in Linear Time in the Wagner Whitin Case. *Operations Research* **40**, Suppl. No. 1, pp. S145-S156.

Wagner, H.M. and T.M.Whitin. 1958. Dynamic Version of the Economic Lot Size Model. *Management Science* **5**, pp. 89-96.