# Inventory Control with Generalized Expediting

Eric Logan Huggins
huggins_e@fortlewis.edu
School of Business Administration
Fort Lewis College
Durango, CO 81301-3999

Tava Lennon Olsen
olsen@olin.wustl.edu
John M. Olin School of Business
Washington University in St. Louis
St. Louis, MO 63130-4899

September, 2001; Revised August 2003; December 2005

**Abstract**

We consider a single-item, periodic review inventory control problem where discrete stochastic demand must be satisfied. When shortages occur, the unmet demand must be filled by some form of expediting. We allow a very general form for the cost structure of expediting. This expediting might include in-house rush production, outsourcing, or even lost sales. However, we explicitly consider the case where expedited production is allowed to produce up to a positive inventory level. For the infinite horizon discounted problem, we characterize the structure of the optimal expediting policy and show that an $(s, S)$ policy is optimal for regular production. In certain cases we demonstrate that it may indeed be optimal to use expedited production to build up inventory. A heuristic for policy calculation is given; a numerical study tests the heuristic and add insight into the results.

# 1 Introduction

Traditionally, managers have controlled inventory by setting high enough inventory levels that the likelihood of stockouts is low, and allowing parts to be backordered when shortages do occur. However, the shift towards lean inventory has caused many managers to reduce inventories, which in turn may increase the likelihood of stockouts. Moreover, the cost of backorders (already high, though hard to estimate) is certainly not decreasing in today's competitive market. Therefore, we consider a problem where the cost of holding inventory may be relatively high and backorders are strictly forbidden. Of course, shortages will occur with stochastic demand; in our problem, these shortages must either be filled by some form of expediting or by lost sales.

We allow expediting to take two general forms. The first, which comprises the majority of the paper, allows the possibility of production above and beyond the backlog. The cost associated with this rush production is assumed to be concave non-decreasing and therefore it incorporates special cases such as fixed and linear components, models with economies of scale, etc. For many of our results the cost function need only be non-decreasing. The second form of expediting, which may be used in combination with the first, is one where only unmet shortages may be met and no added production is allowed. Naturally this case incorporates the case of lost sales. We consider this form of expediting alone, where it need only have non-decreasing costs, and together with rush production, where we study the special case of linear and fixed costs.

Thus, we study a single-item, periodic review inventory control problem where discrete stochastic demand cannot be backlogged. The traditional problem studied in Scarf (1960) and Veinott (1966) allows for backorders when shortages occur; under this condition, $(s, S)$ policies are optimal. Similarly, Zheng (1991) allows for backorders and shows that $(s, S)$ policies are optimal for the infinite horizon case. We use Zheng's results extensively for our main theorem with a modification to exclude backorders (i.e., we require that $s \geq -1$). As in these papers, we do not consider capacity constraints. One might imagine our results being applied to a multi-item setting where the cost of capacity is reflected in the production costs.

Our problem is concerned with meeting shortages in a timely fashion. The literature on this subject appears under various names, such as expediting, emergency orders, and dual supply modes. Daniel (1963) derives optimal policies where there are two supply modes with lead times of 0 or 1 periods. Fukuda (1964) extends these results to the case where the lead times are $k$ or $k+1$ periods. Whittmore and Saunders (1977) show that only one supply mode is optimal when the difference in lead times is more than one period. Interestingly, these three papers all use different terms for expediting: Daniel uses "Emergency", Fukuda uses "Negotiable Leadtime", and Whittmore and Saunders use "Two Supply Options". Other related papers on expediting include: Moinzadeh and Schmidt (1991), Chiang and Gutierrez (1998), Johansen and Thorstenson (1998), Tagaras and Vlachos (2001), Groenevelt and Rudi (2002), and Sethi et al. (2003). (See references there-in for further literature.)

One motivating environment for our work was the inventory control problem faced by a large automobile parts supplier in Michigan, which we will refer to as '*PartCo*'. PartCo produces mostly engine parts for one of the "big three" automobile manufactures. At PartCo, inventory levels are relatively low – about "half a day's worth" according to our contacts. However, many of the parts they supply to the automobile manufacturer are essential in keeping the assembly lines moving. The cost of shutting down an assembly line at the manufacturer is extremely high, "unacceptable" according to our contacts at PartCo. Therefore, backorders are avoided at all costs by either producing extra parts by running overtime or by shipping parts by air so that they arrive in a few hours or even minutes, rather than overnight. Both of these expensive practices are "common".

In the above example expediting corresponds to in-house production. However, another possible application includes outsourced or subcontracted production. Bradley (1997), (2004), and (2005) has looked at (and provided motivation for) subcontracting in a number of forms. There has also been some work that looks at the interaction between subcontracting and capacity investment (see, e.g., Van Mieghem, 1999) as well as some continuous time models of outsourcing (see, e.g., Arslan et al., 2001, and Zheng, 1994). Overtime, or a "vendoring option", in the context of inventory systems with production quotas was considered in Hopp et al. (1993), Duenyas et al. (1993), and Duenyas et al. (1997). Each paper provides structural results for a number of different models. The most related model to ours is Model 2 in Duenyas et al. (1997). In this paper an order-up-to policy is assumed for regular production and therefore the paper concentrates on evaluating the amount of overtime to use in a system where backlogs are allowed. To the best of our knowledge, our work is the first to provide structural results for both steps in a two stage decision where at both points one can produce beyond demand backlog.

Our second option for expediting (which may be used as a supplement to the first) does not allow for excess production. As such it may also be viewed as a shortage or lost sales cost. There have been several related approaches to dealing with shortages in the literature. Smith (1977) considers an $(S-1, S)$ system without backorders where a per unit penalty $L$ is assessed for units of unmet demand. In the model of Çetinkaya and Parlar (1998), backorders are allowed and incur fixed and per unit costs. They show that a myopic base stock policy is optimal over the infinite horizon under certain conditions. Lovejoy and Sethuraman (2000) consider a novel model where a production schedule must always be met by rushing production (at the risk of producing defective products) and/or using overtime; however, they do not explicitly consider stochastic demand. Mohebbi and Posner (1999) provides performance analysis for a model with both emergency orders and lost sales.

Closest to our work, Aneja and Noori (1987) consider a problem where unmet demand is met by "some external arrangement" with both per-unit and fixed costs. They assume that if a shortage occurs, the inventory level will be brought up to 0 and they show that $(s, S)$ policies are optimal over the finite horizon when the demand density is non-increasing. Ishigaki and Sawaki (1991) extend this work to give a condition based on the problem parameters for $(s, S)$ policies to be optimal for a finite horizon model with both fixed and per unit holding and lost sales costs. To the best of our knowledge, our work is the first to extend lost sales models with fixed and per unit costs to the infinite horizon and the first to consider general non-decreasing lost sales costs. However, given that the results fall out as a natural extension to Zheng's (1991) work, we do not want to overstate this part of our contribution and cannot guarantee that it has not appeared as a note in a paper or in an inventory textbook somewhere.

Finally, for our results to hold, we require that our demand distributions be logconcave. More specifically, the CDF (cumulative distribution function) $F(\cdot)$ of demand must be logconcave, and hence by definition in Rosling (2002), $f(x+1)/F(x)$ is non-increasing in $x$ where $f(\cdot)$ is the associated PMF (probability mass function). We use this assumption exclusively in proving that $(s, S)$ policies are optimal for regular production due to the fact that logconcave functions are closed with respect to composition, as shown in Ibragimov (1956).

Rosling gives an excellent overview of logconcavity and lists several useful properties in his 2002 paper. He defines logconcavity for both the CDF $F(\cdot)$ and frequency function (PDF or PMF) $f(\cdot)$ and relates the two concepts to each other and to the concept of monotone convolution ratios (MCRs). He also points out that this assumption is not terribly restrictive as most commonly assumed distributions are in fact logconcave; for example, the normal, uniform, exponential, beta and gamma are all continuous logconcave distributions and the binomial, Poisson, and discrete uniform are all discrete logconcave distributions. In his work, he analyzes systems with general backlogging costs and shows that the cost rate is quasi-convex, and hence standard inventory policies hold, under the assumption of logconcave demand. A function $f(x)$ is quasi-convex, according to Rosling (2002), "if and only if $-f(x)$ is unimodal." Lastly, An (1997) considers logconcave functions without assuming differentiability, as in the discrete case which we assume.

Because our problem has two decision variables (one for expedited production and the other for regular production), our proof is divided into two main steps. In the first step, we characterize the structure of the optimal expediting production policy. Then, in the second step, we show that an $(s, S)$ policy is optimal for regular production. To do so, we first rely on a relationship between two optimal cost functions (one for expediting, the other for regular production). Then, because we have quasi-convex cost functions, we utilize the concept of logconcavity to complete our proof. Finally, our proof yields results for the infinite horizon case, bypassing any finite horizon results. The two step nature of the problem combined with end of horizon effects make finite horizon results difficult to prove. We have found examples via numerical analysis where, near the end of the horizon, the parameters vary non-monotonically, preventing us from using monotonicity results for the finite horizon. Further, the infinite horizon model coupled with a quasi-convex cost function fits naturally into the framework of Zheng (1991).

We provide explicit heuristic approximations for calculating the parameters in models with fixed and linear expediting costs (in both the single mode and dual mode models). These approximations are heavily based on traditional lost sales models with an adjustment for fixed and linear costs and a further adjustment for expediting to a positive inventory level. As exact numerical methods are available, we have focused on explicit approximations so that they are readily spreadsheet implementable. We focus on continuous time approximations and our primary reference sources are Archibald (1981) and Porteus (1985). Broader discussion on related heuristics may be found in Lee and Nahmias (1993).

As described above, the contribution of our paper is fivefold. First, we allow for two different modes of expediting, each with general cost structures. Second, in the first mode of expediting we explicitly explore the possibility that production may be used to produce beyond zero. Third, we generalize previous work on lost sales models to the infinite horizon with general cost functions. Fourth, we provide a novel methodology for dealing with two-step decisions. Finally, we provide heuristics for explicit calculation of problem parameters.

This paper is organized as follows. In Section 2, we present our basic model involving rush production. Section 3 explains the two time period viewpoint that will be used throughout the paper and then uses it to characterize the structure of optimal overtime production policies and shows that the optimal regular production policy is an $(s, S)$ policy. It also extends the model to the two modes of expediting described above. In Section 4, we develop heuristic approximations and numerically analyze the policies found in Section 3. Finally, Section 5 concludes the paper.

## 2 Model, Notation, and Assumptions

This section outlines our base model with one type of expediting and presents notation and assumptions used throughout the paper. We consider a model where a manager must make two inventory decisions each day (or period). At the beginning of the day, the current inventory level is known and the manager must then decide what inventory position to produce up to with regular production. After regular production is determined, stochastic demand is realized and inventory updated accordingly. The manager must now decide how many expedited orders to produce. However, backlogged demand is not allowed for the following day so that if inventory is negative then the decision becomes whether to just produce up to zero units, or if not, up-to what positive level to produce to start the next day with positive inventory. Whatever inventory position is chosen becomes the starting inventory level for the next day, a holding cost is charged for any positive inventory, and the cycle continues.

Of course, every decision incurs costs. In reality, the costs of expedited production are relatively high compared the costs of regular production, which our assumptions take into account. We assume that all costs are stationary and will be discounted per period by a factor $\alpha$. Regular production incurs a fixed cost of $K_r \geq 0$ and a per unit cost of $c_r \geq 0$. Expedited production incurs a cost $e(x)$ for $x$ units and a general holding cost function $h(x)$ is assessed to all positive inventory, $x$, after expediting. Assumptions on $e(x)$ and $h(x)$ follow. However, we first define the following normalized functions

$$h_r(x) \equiv h(x) + (1-\alpha)c_r x$$

and

$$e_r(x) \equiv e(x) - c_r x.$$

Define $\delta(x) = 1$ if $x > 0$ and $\delta(x) = 0$ otherwise. We make the following assumptions.

(A1) The per period demand distribution is discrete and logconcave. Demand in each period is *i.i.d.*

(A2) If $D$ is the generic demand in a period then $0 < E[D] < \infty$.

(A3) $0 < \alpha < 1$.

(A4) $e_r(x) \geq \alpha K_r \delta(x)$ for $x \geq 0$.

(A5) $e_r(x)$ is non-negative, non-decreasing in $x$ with $e_r(0) = 0$ and $\lim_{x\to\infty} e_r(x) = \infty$.

(A6) $h_r(x)$ is non-negative, non-decreasing in $x$ with $h_r(0) = 0$ and $\lim_{x\to\infty} h_r(x) = \infty$.

Note that (A4) implies that the fixed costs associated with expediting are at least as great as those associated with performing regular production in the following period. Similarly, (A5) implies that the marginal cost of expediting is at least as great as the marginal cost of regular production. Clearly a fixed and per unit cost associated with expediting of $K_e$ and $c_e$, respectively, fits this model if $K_e \geq \alpha K_r$ and $c_e > c_r$. However, other more general models, such as ones with quantity discounts, will also fit the assumptions.

We define:

$D_t$ = the demand during period $t$

$x_t$ = the inventory level at the start of regular time during period $t$

$y_t$ = the inventory position chosen for regular production during period $t$

$\tilde{x}_t$ = the inventory level at the start of expediting during period $t$

$\tilde{y}_t$ = the inventory position chosen for expedited production during period $t$

Note that $\tilde{x}_t = y_t - D_t$ and $x_{t+1} = \tilde{y}_t$. From the problem description we know that the inventory position chosen cannot be less than the inventory level with which we start, and after expediting the inventory position cannot be negative since we do not allow backorders. Thus, $y_t \geq x_t^+$ and $\tilde{y}_t \geq \tilde{x}_t^+$, where we define $x^+ = x$ if $x \geq 0$ and $x^+ = 0$, otherwise. Similarly, we define $x^- = -x$ if $x \leq 0$ and $x^- = 0$, otherwise. Note that we write $y_t \geq x_t^+$ without loss of generality as by definition $x_t$ is not allowed to be negative. We are assuming that initial inventory is non-negative (i.e., $x_0 \geq 0$).

Due to the dual-period nature of this decision, we will consider two types of time periods: the regular-period and the expediting-period. We consider these two periods so that we may analyze various costs starting at different instances in the production cycle. The regular-period begins at the start of regular time with known inventory $x_t$ and ends after the expediting inventory position $\tilde{y}_t$ has been chosen and produced. The expediting-period begins at the start of expediting with known inventory $\tilde{x}_t$ and ends after the demand $D_{t+1}$ has been realized. Note that during a regular-period, all variables will have the same subscript $t$; during an expediting-period, expediting variables will have the subscript $t$ and regular production variables will have the subscript $t+1$. These periods are displayed on the timeline in Figure 1 below.

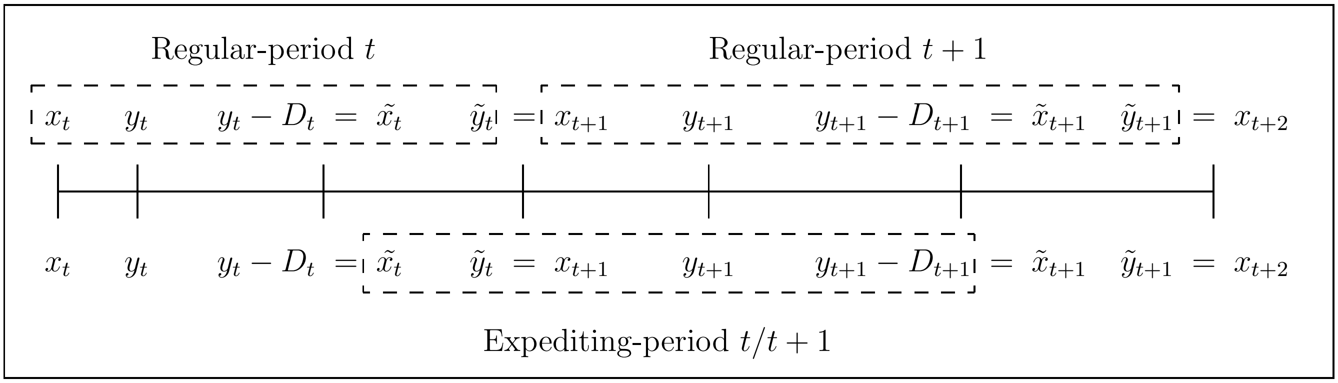


Figure 1: Regular- and Expediting-Periods

Of course, the separation between the time periods is artificial as no randomness is resolved between the end of a regular period and the start of the following regular period (only costs assessed). Therefore one could write down the model as a Markov decision problem with two actions to be decided at the start of each expediting period (see equation (4)). However, as these actions are highly interdependent we found the best way to derive structural results for them is to look at the two different time periods.

For both regular- and expediting-periods, this problem has functions representing the cost per stage, the total cost, and the optimal cost. Here we give an idea of what these functions represent; we will formally define these functions below. We will let $g$ be the cost per stage for a regular-period and $\tilde{g}$ be the cost per stage for an expediting-period. Similarly, we will let $f_\pi$ be the total infinite horizon cost starting in regular time and $\tilde{f}_\pi$ be the total infinite horizon cost starting with expediting under some admissible policy $\pi$. Finally, $f^*$ and $\tilde{f}^*$ will be the minimum total infinite horizon costs starting in regular time and the expediting period, respectively. Note that all of the cost functions depend on the initial inventory, either $x_0$ or $\tilde{x}_0$.

To define our cost per stage for a regular-period, consider that during period $t$, the costs incurred are $K_r\delta(y_t - x_t) + c_r(y_t - x_t) + e(\tilde{y}_t - \tilde{x}_t) + h(\tilde{y}_t^+)$. By observing that $\tilde{y}_t = x_{t+1}$ and $y_t - D_t = \tilde{x}_t$, and observing that $E[c_r D_t]$ is a finite constant (by Assumption A2) that will not affect the optimization the cost per period can be redefined as (cf., Veinott, 1966):

$$g(x_t, y_t, \tilde{x}_t, \tilde{y}_t) \equiv K_r\delta(y_t - x_t) + e_r(\tilde{y}_t - \tilde{x}_t) + h_r(\tilde{y}_t)$$

where $y_t \geq x_t^+$ and $\tilde{y}_t \geq \tilde{x}_t^+$. Note that $g$ is non-negative (by Assumptions A5 and A6). We define our expediting-period cost per stage as:

$$\tilde{g}(\tilde{x}_t, \tilde{y}_t, y_{t+1}) \equiv e_r(\tilde{y}_t - \tilde{x}_t) + h_r(\tilde{y}_t) + \alpha K_r\delta(y_{t+1} - \tilde{y}_t) \tag{1}$$

where $\tilde{y}_t \geq \tilde{x}_t^+$ and $y_{t+1} \geq \tilde{y}_t^+$. This can also be seen to be non-negative.

Let $\pi$ be an admissible policy if $y_t \geq x_t^+$ and $\tilde{y}_t \geq \tilde{x}_t^+$ for all $t$ and $y_t$ and $\tilde{y}_t$ are chosen in a non-anticipatory fashion. In other words, $y_t$ may only depend on $x_t$ and $(x_i, y_i, D_i, \tilde{x}_i, \tilde{y}_i)$ where $i < t$; $\tilde{y}_t$ may only depend on $\tilde{x}_t$ and $(\tilde{x}_{i-1}, \tilde{y}_{i-1}, x_i, y_i, D_i)$ where $i \leq t$. Let $\Pi$ be the set of all such policies. For the regular-period, for $\pi \in \Pi$, let

$$f_\pi(x_0) \equiv \lim_{N\to\infty} E_{\mathcal{D}_0}\left[\sum_{t=0}^{N-1} \alpha^t g(x_t, y_t, \tilde{x}_t, \tilde{y}_t)\right]$$

where $\mathcal{D}_0 = \{D_0, D_1, D_2, \ldots\}$. Note that this function does not include $-c_r x_0$ or $c_r E[D]/(1-\alpha)$, but as these are fixed finite costs they do not affect the eventual minimization. For the expediting-period, for $\pi \in \Pi$, let

$$\tilde{f}_\pi(\tilde{x}_0) \equiv \lim_{N\to\infty} E_{\mathcal{D}_1}\left[\sum_{t=0}^{N-1} \alpha^t \tilde{g}(\tilde{x}_t, \tilde{y}_t, y_{t+1})\right]$$

where $\mathcal{D}_1 = \{D_1, D_2, D_3, \ldots\}$. Note that the limit is known to exist in both cases since $g(x_t, y_t, \tilde{x}_t, \tilde{y}_t) \geq 0$. and $\tilde{g}(\tilde{x}_t, \tilde{y}_t, y_{t+1}) \geq 0$.

Finally, we define the optimal cost functions starting in the regular time and expedited time periods, respectively, as

$$f^*(x) \equiv \min_{\pi \in \Pi} f_\pi(x) \tag{2}$$

and

$$\tilde{f}^*(\tilde{x}) = \min_{\pi \in \Pi} \tilde{f}_\pi(\tilde{x}) \tag{3}$$

where $x \in \mathcal{Z}^+$, $\tilde{x} \in \mathcal{Z}$, and $\mathcal{Z}$ and $\mathcal{Z}^+$ are the set of all integers and non-negative integers, respectively. Since $\tilde{g}(\tilde{x}, \tilde{y}, y) \geq 0$, Proposition 1.1 of Bertsekas (1995, p. 137) holds and the optimal cost function $\tilde{f}^*$ satisfies

$$\tilde{f}^*(\tilde{x}) = \min_{\tilde{y} \geq \tilde{x}^+,\ y \geq \tilde{y}^+} E_D \left[ \tilde{g}(\tilde{x}, \tilde{y}, y) + \alpha \tilde{f}^*(y - D) \right]. \tag{4}$$

As would be expected, there is a strong relationship between $f^*(\cdot)$ and $\tilde{f}^*(\cdot)$; this relationship is given in the following lemma, which also provides other properties on the cost functions that will be used in the following section.

**Lemma 1**

$$f^*(x) = \min_{y \geq x^+} \left\{ K_r \delta(y - x) + E_D[\tilde{f}^*(y - D)] \right\} < \infty \tag{5}$$

*and*

$$\tilde{f}^*(\tilde{x}) = \min_{\tilde{y} \geq \tilde{x}^+} \{ e_r(\tilde{y} - \tilde{x}) + h_r(\tilde{y}) + \alpha f^*(\tilde{y}) \} < \infty, \tag{6}$$

*where the corresponding policy that solves equation (4) or jointly solves (5) and (6) is the optimal stationary policy for the system. Further, define*

$$f^*(-) = f^*(x), \ \textit{for } x < 0,$$

*(since from equation (5) $f^*(x)$ is constant over $x < 0$); then for any $\tilde{x} < 0$*

$$\tilde{f}^*(\tilde{x}) \geq \alpha f^*(-). \tag{7}$$

*Finally,*

$$K_r + f^*(0) \geq f^*(-). \tag{8}$$

**Proof:** See Appendix.

# 3 Optimal Production with Expediting

This section derives the optimal policies for expediting and regular-time production. We include an extension where two modes of expediting are possible.

## 3.1 Optimal Expediting Policy

This subsection characterizes the optimal expedited production policy. The following theorem shows that if the inventory at the start of the expediting period is non-negative then expedited production will not be used; this is a direct consequence of Assumptions (A4) and (A5) and could almost certainly also be proven using a sample path argument. Theorem 1 also provides further structure for the optimal expediting policy when the normalized expediting cost function is either concave or consists of a fixed and per unit cost.

**Theorem 1** *Let $\tilde{y}^*(\tilde{x})$ be the smallest minimizer in (6).*

1. *If $\tilde{x} \geq 0$ then $\tilde{y}^*(\tilde{x}) = \tilde{x}$.*

2. *If $e_r(\cdot)$ is concave and $\tilde{x} < 0$ then $\tilde{y}^*(\tilde{x})$ is non-increasing in $\tilde{x}$.*

3. *If $e_r(x) = K_e + c_e x$ for some fixed $K_e$ and $c_e$, then $\tilde{y}^*(\tilde{x})$ is constant across $\tilde{x} < 0$.*

**Proof:** From (4), for $\tilde{x} \geq 0$,

$$\begin{aligned}
\tilde{f}^*(\tilde{x}) &= \min_{\tilde{y} \geq \tilde{x}, y \geq \tilde{y}^+} \{e_r(\tilde{y} - \tilde{x}) + h_r(\tilde{y}) + \alpha K_r \delta(y - \tilde{y}) + \alpha E[\tilde{f}^*(y - D)]\} && (9)\\
&\geq \min_{y \geq \tilde{x}} \left\{ \min_{\tilde{y} \geq \tilde{x}} \{e_r(\tilde{y} - \tilde{x}) + h_r(\tilde{y}) + \alpha K_r \delta(y - \tilde{y})\} + \alpha E[\tilde{f}^*(y - D)] \right\} \\
&= \min_{y \geq \tilde{x}} \{h_r(\tilde{x}) + \alpha K_r \delta(y - \tilde{x}) + \alpha E[\tilde{f}^*(y - D)]\} && (10)
\end{aligned}$$

Note that $e_r(\tilde{y} - \tilde{x}) + h_r(\tilde{y}) + \alpha K_r \delta(y - \tilde{y}) \geq h_r(\tilde{x}) + \alpha K_r \delta(y - \tilde{x})$ for $\tilde{y} \geq \tilde{x}$ by Assumptions (A4) - (A6). Therefore as (10) is (9) evaluated at $\tilde{y} = \tilde{x}$ we have that the optimal expediting policy for $\tilde{x} \geq 0$ is not to produce and therefore $\tilde{y}^*(\tilde{x}) = \tilde{x}$.

Assume $e_r(\cdot)$ is concave and suppose that $\tilde{x}_1 < \tilde{x}_2 < 0$ and $\tilde{y}^*(\tilde{x}_1) < \tilde{y}^*(\tilde{x}_2)$; we wish to find a contradiction. By definition of $\tilde{y}^*(\tilde{x}_2)$ (since $\tilde{y}^*(\tilde{x}_1) < \tilde{y}^*(\tilde{x}_2)$),

$$e_r(\tilde{y}^*(\tilde{x}_1) - \tilde{x}_2) + h_r(\tilde{y}^*(\tilde{x}_1)) + \alpha f^*(\tilde{y}^*(\tilde{x}_1)) > e_r(\tilde{y}^*(\tilde{x}_2) - \tilde{x}_2) + h_r(\tilde{y}^*(\tilde{x}_2)) + \alpha f^*(\tilde{y}^*(\tilde{x}_2)).$$

But, by the concave non-decreasing nature of $e_r(\cdot)$,

$$e_r(\tilde{y}^*(\tilde{x}_1)-\tilde{x}_1)-e_r(\tilde{y}^*(\tilde{x}_1)-\tilde{x}_2)-e_r(\tilde{y}^*(\tilde{x}_2)-\tilde{x}_1)+e_r(\tilde{y}^*(\tilde{x}_2)-\tilde{x}_2)\geq 0$$

so

$$e_r(\tilde{y}^*(\tilde{x}_1)-\tilde{x}_1)+h_r(\tilde{y}^*(\tilde{x}_1))+\alpha f^*(\tilde{y}^*(\tilde{x}_1))>e_r(\tilde{y}^*(\tilde{x}_2)-\tilde{x}_1)+h_r(\tilde{y}^*(\tilde{x}_2))+\alpha f^*(\tilde{y}^*(\tilde{x}_2)).$$

which contradicts the optimality of $e_r(\tilde{y}^*(\tilde{x}_1))$.

Finally, suppose $e_r(x)=K_e+c_ex$. For $\tilde{x}<0$ expedited production is required and therefore

$$e_r(\tilde{y}-\tilde{x})+h_r(\tilde{y})+\alpha f^*(\tilde{y})=K_e+c_e(\tilde{y}-\tilde{x})+h_r(\tilde{y})+\alpha f^*(\tilde{y}).$$

Thus,

$$\tilde{f}^*(\tilde{x})=K_e-c_e\tilde{x}+\min_{\tilde{y}\geq 0}\left\{c_e\tilde{y}+h_r(\tilde{y})+\alpha f^*(\tilde{y})\right\},$$

where the term inside the minimization is independent of (and hence constant in) $\tilde{x}$. $\square$

Theorem 1 shows that if inventory is negative and the expediting cost function is concave then the produce-up-to amount is non-decreasing in backlog. In other words, a generalized $(s,S)$ policy (as defined in Porteus, 1971) is optimal for expedited production when there are concave expediting costs. According to Porteus's definition the ordering policy $\tilde{y}$ is "generalized $(s,S)$" if there exists $(\tilde{s},\tilde{S})$ such that $\tilde{y}(\tilde{x})=\tilde{x}$ for $\tilde{x}\geq\tilde{s}$ and $\tilde{y}(\tilde{z})\geq\tilde{y}(\tilde{x})\geq\tilde{S}\geq\tilde{s}$ for $\tilde{z}<\tilde{x}<\tilde{s}$. In our case, $\tilde{s}=0$ and $\tilde{S}=\tilde{y}(-1)\geq 0$ (recall that inventory is assumed to be discrete). For the special case of fixed plus linear costs $\tilde{S}$ is the fixed base-stock level that is optimal when expediting (i.e., when $\tilde{x}<\tilde{s}=0$).

In Assumption (A4) we assumed that $e_r(x)\geq\alpha K_r\delta(x)$ for $x\geq 0$. This was used in Theorem 1 to show that expediting is not used if inventory is non-negative entering the expedited production period. Without this restriction it may well be optimal for expedited production to be used even if inventory is positive. In this case, we would no longer expect an $(s,S)$ policy to be optimal for regular production (as will be shown to be the case in the following subsection). Instead there may well be multiple thresholds depending on whether expediting would be likely to be used or not. Note also that Assumption (A4) precludes a convex $e_r(\cdot)$ function.

The following lemma sets the stage for incorporating the optimal expediting policy into the optimal regular production policy.

**Lemma 2** *Define for* $z \geq 0$

$$p(z) \equiv \min_{y \geq 0}\{e_r(z+y) + h_r(y) + \alpha f^*(y)\} - \alpha(f^*(0) - \delta(z)(f^*(0) - f^*(-))), \qquad (11)$$

*which is well-defined by Lemma 1, then* $p(z)$ *is non-decreasing in* $z$ *with* $p(0) = 0$ *and*

$$\tilde{f}^*(\tilde{x}) = h_r(\tilde{x}^+) + p(\tilde{x}^-) + \alpha f^*(\tilde{x}) \qquad (12)$$

*where* $h_r(\tilde{x}^+) + p(\tilde{x}^-)$ *is quasi-convex in* $\tilde{x}$ *with* $\lim_{|\tilde{x}| \to \infty}(h_r(\tilde{x}^+) + p(\tilde{x}^-)) = \infty$.

**Proof:** By definition $\tilde{y}^*(-z)$ is the minimizer of the right hand-side of (11) for $z \geq 0$. Thus,

$$p(0) = e_r(0) + h_r(0) + \alpha f^*(0) - \alpha f^*(0) = 0$$

since, from Theorem 1, $\tilde{y}^*(0) = 0$ and, by Assumptions (A5) and (A6), $e_r(0) + h_r(0) = 0$.

We now show that $p(\cdot)$ is non-decreasing in $z$. For $z_1 > 0$, if $\tilde{y}^*(-z_1) > 0$,

$$\begin{aligned} p(z_1) &= \tilde{f}(\tilde{y}^*(-z_1)) - \alpha f^*(-) \\ &\geq 0 = p(0) \end{aligned}$$

from Lemma 1. If, for $z_1 > 0$, $\tilde{y}^*(-z_1) = 0$ then

$$\begin{aligned} p(z_1) &= e_r(z_1) + \alpha(f^*(0) - f^*(-)) \\ &\geq 0 = p(0) \end{aligned}$$

from Lemma 1 since $e_r(z_1) \geq \alpha K_r$ by Assumption (A4). Thus, in both cases, $p(z_1) \geq p(0)$ for $z_1 > 0$. Now for $0 < z_1 \leq z_2$,

$$\begin{aligned} p(z_1) &= e_r(z_1 + \tilde{y}^*(-z_1)) + h_r(\tilde{y}^*(-z_1)) + \alpha(f^*(\tilde{y}^*(-z_1)) - f^*(-))\} \\ &\leq e_r(z_1 + \tilde{y}^*(-z_2)) + h_r(\tilde{y}^*(-z_2)) + \alpha(f^*(\tilde{y}^*(-z_2)) - f^*(-))\} \\ &\leq e_r(z_2 + \tilde{y}^*(-z_2)) + h_r(\tilde{y}^*(-z_2)) + \alpha(f^*(\tilde{y}^*(-z_2)) - f^*(-))\} = p(z_2) \end{aligned}$$

where the first inequality follows from the definition of minimum and the second from the fact that $e_r(x)$ is non-decreasing in $x$. Thus $p(\cdot)$ is non-decreasing in $z$.

From equation (6)

$$\tilde{f}^*(\tilde{x}) = e_r(\tilde{y}^*(\tilde{x}) - \tilde{x}) + h_r(\tilde{y}^*(\tilde{x})) + \alpha f^*(\tilde{y}^*(\tilde{x})).$$

If $\tilde{x} \geq 0$ then, from Theorem 1, this implies

$$\tilde{f}^*(\tilde{x}) = e_r(0) + h_r(\tilde{x}) + \alpha f^*(\tilde{x}) = h_r(\tilde{x}^+) + p(\tilde{x}^-) + \alpha f^*(\tilde{x}),$$

where the final equality follows since $p(0) = 0$. If $\tilde{x} < 0$ then

$$h_r(\tilde{x}^+) + p(\tilde{x}^-) + \alpha f^*(\tilde{x}) = p(\tilde{x}^-) + \alpha f^*(-) = \tilde{f}^*(\tilde{x}),$$

where the final equality follows from (6) and the definition of $p(\cdot)$.

Since $p(z)$ is non-decreasing in $z$, $p(x^-)$ is non-increasing in $x$. Further, since $p(0) = h_r(0) = 0$ and, for $x \geq 0$, $h_r(x)$ is non-decreasing in $x$ we have that $h_r(\tilde{x}^+) + p(\tilde{x}^-)$ is quasi-convex. That $\lim_{\tilde{x}\to\infty}(h_r(\tilde{x}^+) + p(\tilde{x}^-)) = \infty$ follows immediately from Assumption (A6). Further, by an argument similar to (8), $f^*(y) - f^*(-) \geq -2K_r$, for any $y \geq 0$. Therefore, by Assumption (A5), $\lim_{\tilde{x}\to-\infty}(h_r(\tilde{x}^+) + p(\tilde{x}^-)) = \infty$. This completes the proof. □

The function $p(z)$ in Lemma 2 represents the penalty associated with a backlog of size $z$ following regular production and demand. If expedited production above zero was not allowed then $p(z)$ would simply equal $e_r(z) - \alpha\delta(z)(f^*(0) - f^*(-))$.

## 3.2 Optimal Regular Time Production Policy

In this subsection we characterize the structure of the optimal regular time production policy. Combining (12) and (5) we have that

$$f^*(x) = \min_{y \geq x^+} \{K_r\delta(y - x) + G(y) + \alpha E_D[f^*(y - D)]\} \tag{13}$$

where we define

$$G(y) \equiv E_D[h_r((y - D)^+) + p((y - D)^-)].$$

Using the fact that demand is logconcave and $h_r(\tilde{x}^+) + p(\tilde{x}^-)$ is quasi-convex we have that $G(y)$ is also quasi-convex, because quasi-convexity is preserved under convolution with log-concave functions, as discussed in Section 1. We can therefore characterize the optimal regular production policy.

**Theorem 2** *The optimal regular production policy is an $(s, S)$ policy with $-1 \leq s < S$.*

**Proof:** Equation (13) is the optimization equation from Zheng (1991) except for the $y \geq x^+$ rather than $y \geq x$. We have that $G(\cdot)$ is quasi-convex. Further, $\lim_{|\tilde{x}| \to \infty} G(y) = \infty$ since $h(\cdot)$ and $p(\cdot)$ are monotone and therefore the interchange of limit an expectation follows from the monotone convergence theorem. Thus, $G(\cdot)$ satisfies the required assumptions in Zheng. For $x < 0$, redefine $G(x) = \infty$ and allow backordering. Note that $G(\cdot)$ remains quasi-convex with $\lim_{|x| \to \infty} G(x) = \infty$. This is an equivalent optimization and the optimality of an $(s, S)$ policy follows directly from Zheng (1991). This infinite cost immediately implies it is better to order than not order when $x < 0$, and hence $s \geq -1$. $\square$

Note that Theorem 2 applies both to models with the possibility of expediting above zero and to models with lost sales. In the latter case the lost sales cost function need only satisfy Assumption (A5), which guarantees quasi-convexity of $G(\cdot)$; there are no structural results to prove for lost sales and hence Assumption (A4) is unnecessary.

Corollary 1 provides further structure on the optimal policy. In particular, it shows that if expedited production is used to build up inventory in some period then regular-time production will not be used in the following period. It gives a myopic, but relatively restrictive, condition on when expedited production above zero is not used. Finally, it shows that the up-to amount when expediting will never exceed the optimal regular-time produce-up-to amount if the latter amount is positive.

**Corollary 1** *Let $(s^*, S^*)$ be the optimal regular production controls and let $\tilde{y}^*(\tilde{x})$ be the optimal produce-up-to amount when expediting (i.e., when $\tilde{x} < 0$).*

1. *If, for some $\tilde{x} < 0$, $\tilde{y}^*(\tilde{x}) > 0$ then $s^* < \tilde{y}^*(\tilde{x})$.*

2. *If, for some* $\tilde{x} < 0$, $e_r(y - \tilde{x}) - e_r(-\tilde{x}) + h_r(y) \geq \alpha K_r$ *for all* $y > 0$ *then* $\tilde{y}^*(\tilde{x}) = 0$.

3. *If* $S^* > 0$ *then* $\tilde{y}^*(\tilde{x}) \leq S^*$ *for any* $\tilde{x} \leq S^*$.

**Proof:** See Appendix □

Note that Condition 2 of Corollary 1 implies that, if $K_r = 0$ (i.e., there are no fixed costs associated with regular time production) then expedited production beyond zero will never be used. Both this and Condition 2 could likely also be proven using a sample path argument as the intuition behind the condition (and associated proof) is simply that the per period fixed cost of regular production in the next period is always less than the extra expediting cost of producing beyond zero plus the holding cost of the extra units of inventory.

## 3.3 Multiple Modes of Expediting

In many cases one may have two (or more choices) for the expediting options, each having differing cost structures. In this section we assume that there are two modes of expedited production. Rush production, which may occur up to any amount, and penalty production which may only be used to produce up to zero inventory. While general concave cost functions are easily modeled, here we restrict attention to fixed and linear costs for both rush and penalty production.

We assume that rush production of $x$ units costs $K_o\delta(x) + c_o x$ for some $K_o \geq \alpha K_r$ and $c_o > c_r$ and penalty production of $x$ units costs $K_p\delta(x) + c_p x$ for some $K_p \geq \alpha K_r$ and $c_p > c_r$. First observe that for $x \geq z \geq 0$,

$$K_o + c_o z + K_p + c_p(x - z) \geq K_o + K_p + \min(c_o, c_p)x;$$

thus it is never optimal for both modes of expediting to occur simultaneously. A formal proof is trivial and would mirror Theorem 1 or apply a sample path argument.

Define

$$C^* \equiv \min_{y \geq 0}\{(c_o - c_r)y + h_r(y) + \alpha f^*(y)\} - \alpha f^*(0)$$

and

$$\tilde{S} \equiv \arg\min_{y \geq 0} \{(c_o - c_r)y + h_r(y) + \alpha f^*(y)\}.$$

Then, when $\tilde{S} > 0$, $C^*$ represents the added cost (or, more correctly, $-C^* \geq 0$ represents the added benefit) in using expediting up-to $\tilde{S}$ instead of up-to 0. Using this and the above structure of expediting we have that, in Lemma 2, for $z \geq 0$,

$$p(z) = \min\{K_o + (c_o - c_r)z + C^*, K_p + (c_p - c_r)z\} + \delta(z)(f^*(0) - f^*(-)).$$

Hence, the values of $K_o$, $c_o$, $K_p$, $c_p$, and $C^*$ determine which mode of expedited production is best, when. In particular, if $c_o \neq c_p$, define

$$\tilde{s} \equiv \left\lfloor \frac{K_o + C^* - K_p}{c_o - c_p} \right\rfloor,$$

where, for any $x$, $\lfloor x \rfloor$ is the largest integer less than or equal to $x$. Note $\tilde{s}$ remains undefined if $c_o = c_p$ as it is unnecessary. The optimal expediting policy is represented in the following table.

| | $K_o + C^* > K_p$ | $K_o + C^* = K_p$ | $K_o + C^* < K_p$ |
|---|---|---|---|
| $c_o < c_p$ | PP for $\tilde{s} < \tilde{x} < 0$<br>RP for $\tilde{x} \leq \tilde{s}$ | RP for all $\tilde{x} < 0$ | RP for all $\tilde{x} < 0$ |
| $c_o = c_p$ | PP for all $\tilde{x} < 0$ | Any | RP for all $\tilde{x} < 0$ |
| $c_o > c_p$ | PP for all $\tilde{x} < 0$ | PP for all $\tilde{x} < 0$ | RP for $\tilde{s} < \tilde{x} < 0$<br>PP for $\tilde{x} \leq \tilde{s}$ |

Table 1: Regions in Which to Use Rush Production (RP) Versus Penalty Production (PP)

**Theorem 3** *An optimal stationary expediting production policy exists and has structure as in Table 1 where under rush production the optimal produce-up-to amount equals $\tilde{S}$ and under penalty production the up-to amount is zero by definition. Regular production follows an $(s, S)$ policy.*

**Proof:** Follows directly from the above, Theorem 2, and simple algebra. □

Note that when $C^* = 0$ (for example, when $K_r = 0$), the optimal expediting policy depends only on the relative expediting costs and is independent of the demand distribution.

# 4 Heuristic and Numerical Analysis

In this section, we develop heuristic approximations and perform numerical analyzes assuming linear plus fixed costs for expediting. In Section 4.1, we consider a single form of expediting (as in Sections 3.1 and 3.2) and develop heuristics for $s^*$, $S^*$, and $\tilde{S}^*$. In Section 4.2, we numerically analyze the model and compare it to the case where no excess expedited production is allowed; we also test the approximations developed in Section 4.1. In Section 4.3, we develop heuristic approximations for the two forms of expediting of Section 3.3 and discuss some numerical results.

In our problem, we assume that deliveries either occur overnight or effectively instantaneously with expediting. This will be true for a manufacturing plant delivering to neighboring manufacturing plants. However, one of the fundamental aspects of our model is that demand must be met. If there is a leadtime $L$ for production or shipping then clearly demand must be provided $L$ periods in advance. If expedited production just allows for a one period gain, then the model remains the same, just transposed by $L$ periods. We will consider a continuous approximation to our model with a delivery leadtime of $L$, where $L = 1$ approximates the model of the previous section.

## 4.1 Heuristic with One Mode of Expediting

Assume that $e_r(x) = K_e\delta(x) + c_e x$ for some fixed $K_e$ and $c_e$, $x \geq 0$. Then from Theorems 1 and 2, optimal values $s^*$, $S^*$, and $\tilde{S}^*$ exist for the model with expedited production as the only form of expediting. Calculating these optimal values exactly requires a lengthy iterative procedure, so here we develop heuristic estimates. We focus our attention on explicit estimates as the most likely to be implemented. Further, for an iterative heuristic one would need to show significant improvement either in run-time or complexity, over exact iterative solutions. Because we found our explicit heuristic to perform reasonably well (see the following subsection) we choose to avoid any such issues.

To develop these estimates, we follow closely the procedures discussed in Archibald (1981), Porteus (1985), and Lee and Nahmias (1993). The last is a survey article and gives appropriate references for the results we use. We first approximate the values of $s^*$ and $S^*$ by using a continuous-time model without excess expedited production. We estimate $S^*$ using $s^* + Q_r$ for some order quantity $Q_r$, after adjusting for "overshoot". We then use these values to estimate $\tilde{S}^*$, the up-to amount for overtime.

As discussed above, we wish to develop explicit estimates. We therefore use the simple EOQ value for the heuristic value of $Q_r$, $Q_r^H$, appropriately adjusted for discounting (see, e.g., equation (2.3) of Lee and Nahmias, 1992). That is we set

$$Q_r^H \equiv \sqrt{\frac{2\mu K_r}{h_r + c_r(1-\alpha)}}. \tag{14}$$

As mentioned above, we approximate our model using a continuous-time model without excess expedited production. This approximation leads to a long-run average cost function $C_r$ that resembles a lost sales model (see, e.g., p. 38 Lee and Nahmias, 1992) but is different due to the fixed cost of expedited production. We have

$$C_r = \frac{\mu}{Q_r}\Big[K_r + (c_e - c_r)E[D_L - s]^+ + K_e P[D_L > s]\Big] + h_r\left[\frac{Q_r}{2} + E[s - D_L]^+\right]$$

where $\mu$ is the average demand per unit time, $Q_r$ is the order quantity, and $D_L$ is the stochastic demand during the lead time.

We plan to set $Q_r = Q_r^H$ and minimize this equation with respect to $s$ (rather than using lengthier iterative procedures). Define $\phi(x) = P(D_L = \lceil x \rceil)$, where, for any $x$, $\lceil x \rceil$ is the smallest integer greater than or equal to $x$, so that

$$P[D_L > s] = \int_s^\infty \phi(x)dx.$$

Then, setting $\partial C_r/\partial s = 0$ yields

$$(\mu(c_e - c_r) + Q_r h_r)P(D_L > s) + K_e\mu\phi(s) = Q_r h_r$$

The inclusion of the $K_e\mu\phi(s)$ term makes this equation a function of both the probability mass function and the cumulative probability, which may be solved implicitly but not explicitly; hence, we employ another estimate.

Our goal is to solve $\phi(s) \approx \frac{Q_r h_r}{K_e \mu}$, but not exactly since by ignoring the cumulative term, our estimate for $\phi(s)$ would be too high. Moreover, we want the estimate to increase as the numerator increases, to decrease as the denominator increases, and to remain small enough such that the fraction for the cumulative distribution is less than 1. To advance these goals, we reconsider the problem in terms of a simple newsvendor model where the numerator is the "holding" cost and the denominator is the "penalty" cost and estimate

$$\phi^H \equiv \phi\left(P^{-1}\left(\frac{K_e\mu}{h_r Q_r^H + K_e\mu}\right)\right) \tag{15}$$

which yields the approximation

$$s^H \equiv P^{-1}\left(\frac{\mu(c_e - c_r) + K_e\mu\phi^H}{\mu(c_e - c_r) + h_r Q_r^H}\right). \tag{16}$$

Clearly, this is a rough estimate, but it does factor in both the linear and fixed costs of expediting and it is trivial to calculate (assuming all distributions and costs are known).

We are now ready to estimate $S^*$. As discussed on p. 139 of Porteus (1985), the average order size in the discrete model will be larger than $S - s$. We use the suggested adjustment for this overshoot as follows,

$$S^H \equiv Q_r^H + s^H - \frac{E[D]}{2} - \frac{Var[D]}{2E[D]} \tag{17}$$

where $D$ is the generic demand in a period.

Finally, relaxing our assumption of no excess expedited production above, we must estimate when to use excess expedited production and how much. We use a similar procedure as above and estimate $\tilde{S}^H \equiv Q_e^H + s^H - E[D]/2 - Var[D]/2E[D]$, where we are yet to estimate the excess expedited production amount $Q_e^H$.

The difference between a regular production cycle and expedited production cycle is approximately

$$C_r - C_e \approx \frac{\mu K_r}{Q_r^H} + \frac{h_r Q_r^H}{2} - \frac{h_r Q_e}{2}.$$

Similarly, the expected benefit of using excess expedited production (over expediting to zero) is approximately

$$-C_O \approx (C_r - C_e)\frac{Q_e}{\mu} - (c_e - c_r)(Q_e + s^H)$$

which is maximized by

$$Q_e^H \equiv \frac{\mu}{h_r}\left(\frac{K_r}{Q_r^H} + \frac{h_r Q_r^H}{2\mu} - c_e + c_r\right). \tag{18}$$

Finally, we define

$$\tilde{S}^H \equiv \min\{S^H, \max\{Q_e^H - \frac{E[D]}{2} - \frac{Var[D]}{2E[D]}, 0\}\} \tag{19}$$

since $\tilde{S}^* \leq S^*$ and if $Q_e^H - E[D]/2 - Var[D]/2E[D] \leq 0$, the optimal policy is simply to order up to 0 with expedited production.

Equations (14) - (19) define explicit heuristic estimates $s^H$, $S^H$, and $\tilde{S}^H$ for $s^*$, $S^*$, and $\tilde{S}^*$, respectively. In the next section we discuss their accuracy.

## 4.2 Numerical Analysis with One Mode of Expediting

In this section, we numerically analyze the expedited production model with linear plus fixed costs, and compare it to the case where excess expedited production is not allowed and to the approximations developed in the previous section. Briefly, our results show that although excess expedited production may be optimal, the benefits are minimal. However, our heuristic policy performs quite well.

We compute the optimal parameters for regular and expedited production ($s$, $S$, and $\tilde{S}$) using C++ code to iteratively perform all calculations. In our experiment, we set $c_r = 10$ and varied all other parameters. The regular setup cost, $K_r$, took on the values of 100 and 200. We ignored the case where $K_r = 0$ since excess expedited production is never optimal in that case. We let the holding cost $h_r = 0.01, 0.05$, and 0.1. For expedited production, we let the per unit cost $c_e = 10.1$, 10.5, 11, and 15 where the smaller costs correspond to the expediting being performed by a second shift or some other form of inexpensive expediting, rather than the largest cost which represents traditional "time-and-a-half." The expediting setup cost $K_e = 150$ and 250. Finally, we let $\alpha = 0.99, 0.999$, and 0.9999, roughly corresponding to quarterly, weekly, and daily discounting. These variations lead to $2^2 3^2 4 = 144$ combinations, although only 108 (three-fourths) are feasible since some of the combinations violate Assumption (A4) with $K_e < \alpha K_r$.

Our results require discrete, logconcave probability distributions, and we initially ran the experiment for the Poisson distribution and discretized versions of the Normal, Uniform, and Exponential distributions, truncated when necessary. However, all distributions provided similar results for the expedited production parameters so we chose to focus on the discretized Normal distribution with mean 25 and standard deviation 5.

For each combination we made several calculations: We calculated the optimal parameters $s^*$, $S^*$ and $\tilde{S}^*$ as well as the inventory/expediting costs (the total discounted cost less the expected cost of regular production). Our first result is that with our data, excess expedited production is frequently optimal, 41% of the time. However, the benefit of excess expedited production is minimal, saving on average only **0.25%** on inventory/expediting costs (with a maximum benefit of 0.80%) when compared to the model where expedited production just fills the shortage. Unsurprisingly, excess expedited production mostly occurs with low expediting linear costs (10.1, 10.5) and high regular production fixed costs (200); it never occurs when $c_e = 15$. This calls into question whether excess expedited production is practical or not. This question would have to be answered by the inventory managers themselves, but in some cases (where perhaps a "free" production day could be used to produce a different product, for plant maintenance, etc.), it may make sense. On the other hand, since the savings are generally low, managers may choose to forego the option of excess overtime production.

Secondly, we compared our heuristic to optimality, and it performed quite well. We calculated $s^H$, $S^H$ and $\tilde{S}^H$ for the same data and compared inventory/expediting costs. On average, the heuristic was only **0.46%** higher than the optimal inventory cost, with a minimum of 0.00% and a maximum of 2.77%. The heuristic appears to be unbiased in that it there is no clear correlation between any of the cost parameters and its performance.

Our estimates for $s^*$ appear to be good and our estimates for $S^*$ appear to be very good, although the estimates for $\tilde{S}^*$ are a little too high. For each heuristic, we determined the mean absolute percentage error (MAPE) and the maximum absolute percentage error (Max) when compared to optimality. Also, we calculated the frequency of each heuristic underestimating (Under), exactly estimating (Exact), and overestimating (Over), with the results listed in the table below.

| | $s^H$ | $S^H$ | $\tilde{S}^H$ |
|---|---|---|---|
| MAPE | 8% | 1% | 6% |
| Max | 31% | 5% | 73% |
| Under | 45% | 52% | 16% |
| Exact | 10% | 9% | 55% |
| Over | 44% | 39% | 30% |

Table 2: Heuristic Accuracy

Note that the data for $\tilde{S}^H$ was for all outcomes, including 64/108 (59%)where $\tilde{S}^* = 0$. Considering only the cases where both $\tilde{S}^*$ and $\tilde{S}^H$ are positive, the last column should read 11%, 73%, 36%, 5%, and 60%. Although our values of $\tilde{S}^H$ are somewhat inaccurate, they do not have a huge effect on inventory/expediting costs due to the infrequency of using expedited production. Note that simply assuming that no excess expedited production is allowed would create a much simpler heuristic where only $s^H$ and $S^H$ need to be determined.

### 4.3 Heuristic and Numerical Analysis with Two Modes of Expediting

In this section, we discuss a heuristic for the model with two forms of expediting as outlined in Section 3.3. We follow a similar process as in Section 4.1 to develop the heuristic, but new difficulties arise. Then, we perform another numerical analysis assuming that both methods of expediting are available.

In order to estimate $Q_r^H$, we again consider the long-run average cost $C_r$. However, this cost now depends on $\tilde{s}^*$, which in turn depends on $-C_O$ and the rest of our parameters. So, at this point we assume that no excess expediting is allowed (at minimal cost, as seen in the previous section). Thus, $-C_O = 0$, $\tilde{S}^H \equiv 0$ and $\tilde{s}^H$ may easily be calculated as in Section 3.3 as

$$\tilde{s}^H \equiv \left\lfloor \frac{K_o - K_p}{c_o - c_p} \right\rfloor .$$

Recall from Section 3.3 that four different possibilities arise. One is to only use rush production, in which case, simply follow the heuristic developed in Section 4.1. Another is to only use penalty production; in this case, use the same technique as before, but substitute the penalty production costs for the expedited production costs in the heuristic calculations. Third, it may turn out to use penalty production for small shortages and rush production for large shortages, or fourth vice-versa, where $\tilde{s}$ is the threshold. We focus on the third case and find that

$$\begin{aligned} C_r = \frac{\mu}{Q_r} &\left[(c_p - c_r)E[(D_L - s)I(s \geq D_L \geq s + \tilde{s}^H)] + K_p P[s \geq D_L \geq s + \tilde{s}^H]\right] \\ &+ \frac{\mu}{Q_r}\left[(c_o - c_r)E[(D_L - s)I(D_L > s + \tilde{s}^H)] + K_o P[D_L > s + \tilde{s}^H]\right] \\ &+ \frac{\mu K_r}{Q_r} + h_r\left[\frac{Q_r}{2} + E[s - D_L]^+\right], \end{aligned}$$

where $I(\cdot)$ is the indicator function which is 1 if the event occurs and zero if it doesn't.

It is likely that under most logconcave demand functions, smaller shortages occur more frequently due the unimodal nature of the distribution. Hence to estimate $Q_r^H$, $s^H$ and $S^H$, we suggest using only the costs for whichever method is used for smaller shortages, in this case penalty production, and follow the previous technique.

Numerically, we analyzed the case with both methods of expediting by repeating our experiment from Section 5.2 including three linear penalty production costs ($c_p = 10.2, 12, 16$) and two penalty production fixed costs ($K_p = 50, 250$). These additions lead to 648 feasible combinations. Our first result is that all four possibilities arise: only expedited production occurred 20% of the time; only penalty production, 27%; expedited production for small shortages, penalty production for large shortages, 6%; and vice-versa, 46%. Clearly, these results depend upon the chosen cost parameters, but most importantly, all situations are possible.

Finally, with both methods of expediting, our heuristic did not perform as well. Overall, the average inventory/expediting cost of the heuristic was **5.52%** higher than optimal, with a minimum of 0.16% and a maximum of 17.02%. Interestingly, the error depends on $\alpha$, with the heuristic being off by 11.99%, 2.88%, 1.70% for $\alpha = 0.99, 0.999, 0.9999$, respectively. These results suggest that a more accurate heuristic would depend more on the discount factor. Lastly, our estimate for $S^*$ continued to be very good with an MAPE of only 1.58% whereas our estimate for $s^*$, unsurprisingly, worsened with an MAPE of 15.55%.

# 5 Conclusion and Extensions

In this paper we have modeled an inventory control problem where stochastic demand must always be met and shortages may be filled by expediting. Our goal was to determine optimal policies for expediting and regular production and to gain insight about this problem.

We first considered a model with one mode of expediting. Under a variety of assumptions on holding and expediting costs, we characterized the structure of the optimal expediting policy and showed that the optimal regular production policy is $(s, S)$. We then considered a model with two forms of expediting. Again, we explored the structure of the optimal expediting policy and showed that the optimal regular production policy is $(s, S)$. We presented heuristics for explicit parameter calculation and tested them using a numerical study. They were shown to perform quite well. Finally, we used numerical analysis to gain insights into the different optimal expediting policies and the frequency of excess expedited production and associated savings.

Capacity constraints may apply to either regular production, expedited production, or possibly both. However, if both types of production are constrained, we cannot guarantee that demand will always be met. As capacitated inventory problems with fixed order costs are extremely challenging in their own right (see, e.g., Gallego and Scheller-Wolf, 2000), we leave such extensions as the subject of future research.

As discussed in Section 4, if expedited production just allows for a one period gain then the model remains equivalent. However, the more interesting case is where expedited production allows for a choice of the number of periods shipping will take. In this case one would expect an $L$ dimensional state-space where one keeps track of the inventory due in $k$ periods for $k = 1, 2, \ldots, L-1$. The decision on which periods to ship by premium freight would depend on the specific cost structure. One would need conditions to ensure that the expediting production cost function, $p(\cdot)$, remains quasi-convex and then the same proof would be able to be used to show $(s, S)$ production for regular time. This is left as the subject for future research.

A number of other extensions to this work are natural to consider including non *i.i.d.* demand, demand forecasting, and multi-echelon supply-chains. The latter has been explored to some extent for the case of no fixed production costs in Huggins and Olsen (2003, 2005).

## Acknowledgments

We thank Brad Johnson and Michael Knox, our contacts at 'PartCo,' for sharing their insights about the real inventory challenges faced by their employer. We also thank Sven Axsater for introducing us to Kaj Rosling's work prior to its publication as well as the anonymous referees and associate editors for their extremely helpful suggestions for improving this paper. We particularly note that the simple form for the proof of Theorem 2 (replacing that of Huggins, 2002) is due to a referee and much of the form of the heuristic was suggested by the AE. Finally, this research was funded in part by NSF grant DMI-9713727.

# Appendix

*Proof of Lemma 1*

We first show that:

$$f_\pi(x_0) = K_r\delta(y_0 - x_0) + E_{D_0}\left[\tilde{f}_\pi(y_0 - D_0)\right]. \tag{20}$$

where $y_0 \geq x_0^+$ and

$$\tilde{f}_\pi(\tilde{x}_0) = e_r(\tilde{y}_0 - \tilde{x}_0) + h_r(\tilde{y}_0) + \alpha f_\pi(\tilde{y}_0). \tag{21}$$

for $\tilde{y}_0 \geq \tilde{x}_0^+$.

To wit,

$$\begin{aligned}
f_\pi(x_0) &= \lim_{N\to\infty} E_{\mathcal{D}'}\left[\sum_{t=0}^{N-1} \alpha^t g(x_t, y_t, \tilde{x}_t, \tilde{y}_t)\right] \\
&= K_r\delta(y_0 - x_0) + \lim_{N\to\infty} E_{\mathcal{D}'}\left[\sum_{t=0}^{N-1} \alpha^t \tilde{g}(\tilde{x}_t, \tilde{y}_t, y_{t+1})\right] \\
&= K_r\delta(y_0 - x_0) + \lim_{N\to\infty} E_{D_0}\left[E_{\mathcal{D}^\infty}\left[\sum_{t=0}^{N-1} \alpha^t \tilde{g}(\tilde{x}_t, \tilde{y}_t, y_{t+1})\middle| D_0\right]\right] \\
&= K_r\delta(y_0 - x_0) + E_{D_0}\left[\lim_{N\to\infty} E_{\mathcal{D}^\infty}\left[\sum_{t=0}^{N-1} \alpha^t \tilde{g}(\tilde{x}_t, \tilde{y}_t, y_{t+1})\middle| D_0\right]\right] \\
&= K_r\delta(y_0 - x_0) + E_{D_0}\left[\tilde{f}_\pi(\tilde{x}_0)\right] \\
&= K_r\delta(y_0 - x_0) + E_{D_0}\left[\tilde{f}_\pi(y_0 - D_0)\right]
\end{aligned}$$

where $y_0 \geq x_0^+$. The first equality is true by definition, the fourth equality is true by the Monotone Convergence Theorem since $\tilde{g}(\cdot) \geq 0$, and the fifth equality is true by the definition of $\tilde{f}_\pi$ and because the system is Markovian. The proof of (21) follows immediately as no expectations are needed. Minimizing these equations over $\pi$ yields the relationships in (5) and (6). Below we show that such a minimum indeed exists.

Observe that $f_\pi(x)$ and $\tilde{f}_\pi(\tilde{x})$ are non-negative for all policies $\pi \in \Pi$ and for all $\tilde{x}, x \in \mathcal{I}$ because they are the sum of non-negative functions $g$ and $\tilde{g}$, respectively. Thus, to show that $f^*(x)$ and $\tilde{f}^*(\tilde{x})$ are finite, it suffices to show that there exists a policy $\gamma$ such that $\tilde{f}_\gamma(\tilde{x}) < \infty$ and $f_\gamma(x) < \infty$ for all $x \in \mathcal{Z}^+$, $\tilde{x} \in \mathcal{Z}$. Let this $\gamma$ be such that whenever the inventory level is negative produce up to 0; otherwise, do nothing. Note that this policy applies to both expedited and regular production and that this policy is stationary.

Consider the cost per stage for expediting under this policy:

$$\tilde{g}_\gamma(\tilde{x}, \tilde{y}, y) = K_o\delta(\tilde{y} - \tilde{x}) + h_r(\tilde{y}) = \begin{cases} h_r(\tilde{x}) & \text{if } \tilde{x} \geq 0 \\ K_o & \text{if } \tilde{x} < 0 \end{cases}$$

since $\tilde{y} = \tilde{x}^+$. Thus, since $\gamma$ is stationary and $\tilde{g}_\gamma(\tilde{x}, \tilde{y}, y) \geq 0$, by Corollary 1.1.1 of Bertsekas (1995, p. 139) we have that

$$\tilde{f}_\gamma(\tilde{x}) = \begin{cases} h_r(\tilde{x}) + \alpha E[\tilde{f}_\gamma(\tilde{x} - D)] & \text{if } \tilde{x} \geq 0 \\ K_o + \alpha E[\tilde{f}_\gamma(0 - D)] & \text{if } \tilde{x} < 0. \end{cases}$$

Now note that

$$E[\tilde{f}_\gamma(0 - D)] = K_o(1 - p_0) + \alpha E[\tilde{f}_\gamma(0 - D)]$$

$$\Rightarrow E[\tilde{f}_\gamma(0 - D)] = \frac{K_o(1 - p_0)}{1 - \alpha} < \infty$$

Thus,

$$\tilde{f}_\gamma(\tilde{x}) = \begin{cases} h\tilde{x} + \alpha E[\tilde{f}_\gamma(\tilde{x} - D)] & \text{if } \tilde{x} \geq 0 \\ K_o + \alpha \frac{K_o(1-p_0)}{1-\alpha} & \text{if } \tilde{x} < 0 \end{cases}$$

and $\tilde{f}_\gamma(\tilde{x}) < \infty$ for $-\infty < \tilde{x} < 0$. Now, when $\tilde{x} \geq 0$, the inventory level will remain non-negative for some time $T$ and will then eventually become negative at time $T+1$. Note, $T < \infty$ almost surely (a.s.) since $E[D] > 0$ by Assumption (A2). While the inventory level is non-negative, there will be a holding cost of at most $h_r(\tilde{x})$ for $T$ discounted periods. When the inventory goes negative, to some value $N$ say, there will be an $\alpha^{T+1}$ discounted cost of $K_o + (\alpha K_o(1 - p_0))/(1 - \alpha)$ where $-\infty < N < 0$ *a.s.* and $-E[N] \leq E[D]$. So, for $\tilde{x} \geq 0$,

$$\tilde{f}_\gamma(\tilde{x}) \leq E\left[\sum_{t=0}^{T} \alpha^t h_r(\tilde{x}) + \alpha^{T+1}\left(K_o + \frac{\alpha K_o(1 - p_0)}{1 - \alpha}\right)\right]$$

$$\leq \frac{h_r(\tilde{x})}{1 - \alpha} + K_o + c_o E[D] + \alpha \frac{K_o(1 - p_0)}{1 - \alpha} < \infty.$$

Thus $\tilde{f}_\gamma(\tilde{x}) < \infty$ for $0 \leq \tilde{x} < \infty$. So, $\tilde{f}^*(\tilde{x}) \leq \tilde{f}_\gamma(\tilde{x}) < \infty$ for all $\tilde{x} \in \mathcal{Z}$. Now, $f^*(x)$ is also finite since

$$f_\gamma(x) = E[\tilde{f}_\gamma(x - D)] < \infty.$$

Thus, $f^*(x) \leq f_\gamma(x) < \infty$ for all $x \in \mathcal{Z}^+$. Therefore the optimal cost functions are both finite and satisfy the relationships in (5) and (6).

The fact that any solution to (4) or any joint solution to (5) and (6) must be the optimal solution to (3) and to (2) and to (3), respectively, follows directly from Proposition 1.1 of Bertsekas (1995, p. 137) Further, that the corresponding policy that solves equation (4) or jointly solves (5) and (6) is the optimal stationary policy for the system follows directly from Proposition 1.3 of Bertsekas (1995, p. 143).

We wish to show (7). For $\tilde{x} < 0$,

$$\begin{aligned}
\tilde{f}^*(\tilde{x}) &= \min_{\tilde{y} \geq \tilde{x}^+,\ y \geq \tilde{y}^+} E_D\left[\tilde{g}(\tilde{x}, \tilde{y}, y) + \alpha \tilde{f}^*(y - D)\right] \\
&\geq \min_{\tilde{y} \geq \tilde{x}^+,\ y \geq \tilde{y}^+} \tilde{g}(\tilde{x}, \tilde{y}, y) + \alpha \min_{y \geq 0} E_D\left[\tilde{f}^*(y - D)\right] \\
&\geq \alpha K_r + \alpha \min_{y \geq 0} E_D\left[\tilde{f}^*(y - D)\right] \\
&= \alpha f^*(-)
\end{aligned}$$

The first inequality follows from the definition of a minimum and because $\tilde{x} < 0$; the second follows because $\tilde{x} < 0$ and Assumptions (A5) and (A6) imply that $\tilde{g}(\tilde{x}, \tilde{y}, y) \geq e_r(\tilde{x})$ and then from Assumption (A4). Finally, (8) follows since

$$\begin{aligned}
f^*(-) &= K_r + \min_{y \geq 0} E_D\left[\tilde{f}^*(y - D)\right] \\
&\leq K_r + \min_{y \geq 0}\{K_r \delta(y) + E_D\left[\tilde{f}^*(y - D)\right]\} \\
&= K_r + f^*(0).
\end{aligned}$$

This completes the proof. □

*Proof of Corollary 1*

Assume, for some $\tilde{x} < 0$, $\tilde{y}^*(\tilde{x}) > 0$. We wish to show $s^* < \tilde{y}^*(\tilde{x})$. From equations (4) and (1)

$$\tilde{f}^*(\tilde{x}) = \min_{\tilde{y} \geq 0, y \geq \tilde{y}^+} \left\{e_r(\tilde{y} - \tilde{x}) + h_r(\tilde{y}) + \alpha K_r \delta(y - \tilde{y}) + \alpha E_D\left[\tilde{f}^*(y - D)\right]\right\}.$$

Now if $\tilde{y} \neq y$ then, since $e_r(\tilde{y} - \tilde{x}) + h_r(\tilde{y})$ is non-decreasing in $\tilde{y}$, then it is optimal to set $\tilde{y}$ equal to zero. Thus, for any given $y$, there are only two possible optimal values for $\tilde{y}$, namely 0 or $y$. In other words, either expediting is used to return the system to zero or, if expediting results in positive inventory, then regular time production is not used. This directly implies that $s^* < \tilde{y}^*(\tilde{x})$ if $\tilde{y}^*(\tilde{x}) > 0$. Further, if for any $\tilde{x} < 0$, $e_r(y - \tilde{x}) + h_r(y) \geq e_r(-\tilde{x}) + \alpha K_r$ for all $y > 0$, then $\tilde{y}^*(\tilde{x}) = 0$.

We wish to prove Condition 3. For $0 \leq \tilde{x} \leq S^*$, $\tilde{y}^*(\tilde{x}) = \tilde{x} \leq S^*$ (by Theorem 1). Now suppose $\tilde{x} < 0$. If expedited production results in positive leftover inventory then

$$\tilde{f}^*(\tilde{x}) = \min_{y \geq 0} \left\{ e_r(y - \tilde{x}) + h_r(y) + \alpha E_D \left[ \tilde{f}^*(y - D) \right] \right\}$$

and if expedited production produces up to zero then

$$\tilde{f}^*(\tilde{x}) = e_r(-\tilde{x}) + \min_{y \geq 0} \left\{ \alpha K_r \delta(y) + \alpha E_D \left[ \tilde{f}^*(y - D) \right] \right\}$$

Thus, by definition,

$$S^* = \arg\min_{y \geq 0} \left\{ \alpha K_r \delta(y) + \alpha E_D \left[ \tilde{f}^*(y - D) \right] \right\}$$

and

$$\tilde{y}^*(\tilde{x}) = \arg\min_{y \geq 0} \left\{ e_r(y - \tilde{x}) + h_r(y) + \alpha E_D \left[ \tilde{f}^*(y - D) \right] \right\}$$

By assumption $S^* > 0$ so that

$$S^* = \arg\min_{y > 0} \left\{ \alpha E_D \left[ \tilde{f}^*(y - D) \right] \right\}$$

But $e_r(y - \tilde{x}) + h_r(y)$ is non-decreasing in $y$ so $\tilde{y}^*(\tilde{x}) \leq S^*$. □

# References


[1] An, M. Y. 1997. Logconcavity: A Complete Characterization without Differentiability. Department of Economics, Duke University, Durham.

[2] Aneja, Y., and H. Noori. 1987. The Optimality of (s,S) Policies for a Stochastic Inventory Problem with Proportional and Lump-sum Penalty Cost. *Management Science*, **33**, 750-755.

[3] Archibald, B. 1981. Continuous Review (s,S) Policies with Lost Sales. *Management Science*, **27**, 1171-1178.

[4] Arslan, H., H. Ayhan and T. L. Olsen. 2001. Analytic Models for When and How to Expedite in Make-to-Order Systems. *IIE Transactions*, **33**, 1019-1029.

[5] Bertsekas, D. 1995. *Dynamic Programming and Optimal Control (Volume 2).* Athena Scientific, Belmont, Massachusetts.

[6] Bradley, J. R. 1997. *Managing Assets and Subcontracting Policies*, Ph.D. Dissertation, Stanford University, CA.

[7] Bradley, J. R. 2004. A Brownian Approximation of a Production-Inventory System with a Manufacturer That Subcontracts. *Operations Research*, **52**, 765 - 784.

[8] Bradley, J. R. 2005. Optimal Stationary Control of a Dual Service Rate M/M/1 Production-Inventory Model. *European Journal of Operational Research*, **161**, 812-837.

[9] Çetinkaya, S. and M. Parlar. 1998. Optimal Myopic Policy for a Stochastic Inventory Problem with Fixed and Proportional Backorder Costs. *European Journal of Operational Research*, **110**, 20-41.

[10] Chiang, C. and G. Gutierrez. Optimal Control Policies for a Periodic Review Inventory System with Emergency Orders. *Naval Research Logistics*, **45**, 187-204.

[11] Daniel, K. 1963. A Delivery-Lag Inventory Model with Emergency. In *Multistage Inventory Models and Techniques*. H. Scarf, D. Gilford, and M. Shelly (eds.). Stanford University Press, Stanford, CA.

[12] Duenyas, I., W. J. Hopp, and Y. Bassok. 1997. Production Quotas as Bounds on Interplant JIT Contracts. *Management Science*, **43**, 1372-1386.

[13] Duenyas, I., W. J. Hopp, and M. L. Spearman. 1993. Characterizing the Output Process of a CONWIP Line with Deterministic Processing and Random Outages. *Management Science*, **39**, 975-988.

[14] Fukuda, Y. 1964. Optimal Policies for the Inventory Problem with Negotiable Leadtime. *Management Science*, **10**, 690-708.

[15] Gallego, G., and A. Scheller-Wolf. 2000. Capacitated inventory problems with fixed order costs: Some optimal policy structure. *European Journal of Operational Research*, **126**, 603-613.

[16] Groenevelt, H., and N. Rudi. 2002. A Base Stock Inventory Model with Possibility of Rushing Part of Order. *Working Paper,* University of Rochester, Rochester NY.

[17] Hopp, W. J., M. L. Spearman, and I. Duenyas. 1993. Economic Production Quotas for Pull Manufacturing Systems. *IIE Transactions*, **25**, 71-79.

[18] Huggins, E. 2002. *Supply Chain Management with Overtime and Premium Freight.* Ph.D. Thesis, University of Michigan, Ann Arbor, MI.

[19] Huggins, E. L., and T. L. Olsen. 2003. Supply Chain Management with Guaranteed Delivery. *Management Science*, **49**, 1154-1167.

[20] Huggins, E. L., and T. L. Olsen. 2005. Supply Chain Coordination with Guaranteed Delivery. Submitted to *Annals of Operations Research.*

[21] Ibragimov, I. 1956. On the Composition of Unimodal Distributions. *Theory of Probability and Its Advances.* **1** 255-260.

[22] Ishigaki, T., and K. Sawaki. 1991. On the $(s, S)$ Policy with Fixed Inventory Holding and Shortage Costs. *Journal of the Operations Research Society of Japan*, **34**, 47-57.

[23] Johansen, S., and A. Thorstenson. 1998. An Inventory Model with Poisson Demands and Emergency Orders. *International Journal of Production Economics*, **56**, 275-289.

[24] Lee, H., and S. Nahmias. 1993. Single-Product, Single-Location Models. In *Handbooks in Operations Research and Management Science.* Nemhauser and Kan, eds., 3-50.

[25] Lovejoy, W., and K. Sethuraman. 2000. Congestion and Complexity Costs in a Plant with Fixed Resources that Strives to Make a Schedule. *Manufacturing & Service Operations Management*, **2**, 221-239.

[26] Mohebbi, E., and M. Posner. 1999. A Lost-Sales Continuous Review Inventory System with Emergency Ordering. *International Journal of Production Economics*, **58**, 93-112.

[27] Moinzadeh, K. and C. Schmidt. 1991. An $(S-1, S)$ Inventory System with Emergency Orders. *Management Science*, **39**, 308-321.

[28] Porteus, E. L. 1971. On the Optimality of Generalized $(s, S)$ Policies. *Management Science*, **17**, 411-426.

[29] Porteus, E. L. 1985. Numerical Comparisons of Inventory Policies for Periodic Review Systems. *Operations Research*, **33**, 134-152.

[30] Rosling, K. 2002. Inventory Cost Rate Functions with Nonlinear Shortage Costs. *Operations Research*, **50**, 1007-1017.

[31] Scarf, H. 1960. The Optimality of $(s, S)$ Policies in the Dynamic Inventory Problem. In *Mathematical Methods in the Social Sciences*. 196-202. K. Arrow, S. Karlin and P. Suppes (eds.). Stanford University Press, Stanford, CA.

[32] Sethi, S. P., H. Yan, and H. Zhang. 2003. Inventory Models with Fixed Costs, Forecast Updates and Two Delivery Modes. *Operations Research*, **51**, 321-328.

[33] Smith, S. 1977. Optimal Inventories for an $(S-1, S)$ System with No Backorders. *Management Science*, **23**, 522-533.

[34] Tagaras, G., and D. Vlachos. 2001. A Periodic Review Inventory System with Emergency Replenishments. *Management Science*, **47**, 415-429.

[35] Toktay, L. B., and L. M. Wein. 2001. Analysis of a Forecasting-Production-Inventory system with stationary demand. *Management Science*, **47**, 12681281.

[36] Van Mieghem, J. A. 1999. Coordinating Investment, Production and Subcontracting. *Management Science*, 45, 954-971.

[37] Veinott, A. 1966. On the Optimality of $(s, S)$ Inventory Policies: New Conditions and a New Proof. *SIAM Journal of Applied Mathematics*, **14**, 1067-1083.

[38] Whittmore, A., and S. Saunders. 1977. Optimal Inventory Under Stochastic Demand with Two Supply Options. *SIAM Journal of Applied Mathematics*, **32**, 293-305.

[39] Zheng, Y-S. 1991. A Simple Proof for Optimality of $(s, S)$ Policies in Infinite-Horizon Inventory Systems. *Journal of Applied Probability*, **28**, 802-810.

[40] Zheng, Y-S. 1994. Optimal Control Policy for Stochastic Inventory Systems with Markovian Discount Opportunities. *Operations Research*, **42**, 721-738.