Policies Utilizing Tactical Inventory for Service-Differentiated Customers

Serhan Duran[§], Tieming Liu[†], David Simchi-Levi[‡], Julie L. Swann[§*]

May 2007

Abstract

We consider a manufacturer serving two customer classes where one wants the item immediately and the second receives a discount to accept a delay. We show that a (S, R, B) base stock policy is optimal under differentiation and non-differentiation where S, R, and B are the order-up-to, reserve-up-to, and backlog-up-to amounts.

1 Introduction

Some manufacturing or retail companies now segment their customers according to service and price, since it is not uncommon for customers to have different service and price preferences and the differentiation may benefit the firms. For example, one class of customers may be given immediate fulfillment while another class might receive delayed fulfillment for a discount. For instance, if an executive's laptop has been stolen he may pay a premium for immediate delivery, while someone ordering a computer to go to college may order in advance for a discount. Amazon.com also offers price and delivery time options where paying a price premium gives a customer immediate fulfillment while receiving a Super Saver Shipping discount gives Amazon the opportunity for delayed fulfillment. This may provide greater customer utility (either increased service or decreased price as desired by different customers), while offering greater flexibility to the firm in managing the production system.

Though this can increase utility to the customer or the firm, it is necessary to analyze how to manage the system, which may be more complicated due to the service differentiation. One method to manage this kind of system is to use *tactical inventory*, where current inventory may be set aside to satisfy future demand, and delayed fulfillment of current customers (or "backlogs")

[§]H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology
[†]School of Industrial Engineering and Management, Oklahoma State University
[‡]Dept. of Civil and Environmental Engineering and the Engineering Systems Division, MIT
[*]Corresponding author, jswann@isye.gatech.edu

may be planned. Tactical inventory may increase profits while ensuring that service of both kinds of customers is met.

The use of tactical inventory is considered in Scarf [8], in which the idea of protecting inventory from being sold to current customers or "discretionary sales" is introduced. Scarf showed that a base-stock policy is optimal for a single-class problem when production setup costs are fixed, but that the optimal discretionary sales decision may be different for different demand realizations. Chan et al. [1] also incorporated the idea of tactical inventory decisions for a single-class stochastic inventory model with multi-period pricing and production decisions under limited capacity when demand is a general stochastic function. They show that when the fixed production cost is zero, then the optimal discretionary sales (or reserve inventory) is independent of the demand realization. However, when pricing is a decision, then the discretionary sales decision does depend on the demand realization.

The use of tactical inventory was extended in Duran et al. [4] to allow the reserving of inventory as well as planned backlogging of current customers as a second kind of tactical inventory decision. In that paper, the authors consider multiple customer classes differentiated by their priority level, where the first-class customers receive complete priority over the second-class customers in the use of current resources and future backlogging. A key feature of that paper is that customers in both of the classes behave homogeneously in terms of the delivery time (all customers are willing to wait for fulfillment). The main result is that policies of the (S, R, B) form are optimal, where S is the order-up-to quantity, R is the reserve-up-to amount to protect from selling to current customers, and B is the backlog-up-to amount. Since first and second-class customers can receive delayed fulfillment, the R and B decisions may further be nested by customer class.

A fundamental difference in the current paper compared to [4] is that in this paper the customer classes are differentiated according to their tolerance for delay fulfillment, or "patience". In the current paper, customer classes are not ordered by priority on resources. Although the proof techniques in the current paper are similar, the results do not immediately follow from the models and analysis in [4] because the use of patient and impatient customers changes the form of the models. Customers differentiated according to their service preferences may be more applicable in certain settings, such as when some customer types may have an immediate need for some products.

Other papers that consider serving multiple customer classes in a production system include Deshpande et al. [2], Frank et al. [6], Gupta and Wang [7], and Sobel and Zhang [9]. In most of these, the customer classes differ according to their priority or fulfillment, and the authors look for policies to manage the system. However, a significant difference from ours is that tactical inventory decisions are not considered in these papers. Although allowing tactical inventory may complicate the decision, previous work has shown that it can add to the profits in a manufacturing environment by providing the ability to shift demand ([1, 4]).

We focus on a single product sold at a single manufacturer over a multi-period time horizon, where the manufacturer has limited production capacity. First-class customers claim the item immediately and never accept a delayed fulfilment and are willing to pay a premium over the market price. Second-class customers are sensitive to price, always pay the market price, and accept a delay fulfillment. We analyze the system where the manufacturer can differentiate between the customer classes and service-discriminates according to their preferences, and we also study a system where the manufacturer cannot discriminate and does not serve the classes differently though customers accept or not according to their service preferences. For both systems, the manufacturer decides in each period the amount of inventory to protect from being sold to the current period's demand and saved for future demand, and the amount of demand to backlog as well as the overall production quantity. We show that a modified base-stock policy in the form of (S, R, B) is optimal, whether the manufacturer can or cannot differentiate between the customer classes.

2 Models and Results

2.1 Assumptions and Notation

We study a multi-period time horizon with periodic review where the periods are denoted as $t = 1, 2, \ldots, T$, with T being the end of the horizon. The production in each period t is limited by the capacity, q_t, and the manufacturer pays a production cost per unit of c_t in period t. The salvage value of any units left at the end of the horizon is v. The inventory holding cost per unit in period t, h_t, is assessed to carry inventory from period t to $t+1$.

The first-class customers (index of 1) are willing to pay a premium over the market price for immediate delivery of the item and do not accept delayed fulfillment. The second-class customers (index of 2) pay the market price, and they accept fulfillment delayed up to one period; they can also be served before that deadline if resources are available. We assume that the firm has predetermined prices p_t^1 and p_t^2 to charge customer class 1 and 2 in period t, respectively; the prices may be varying from period to period and are unknown to the customers until the beginning of the period t, since in some companies pricing decisions are made by the marketing department before the start of a selling season while production decisions are made by the operations department. In period t, for $i = 1, 2$, ℓ_t^i is the penalty per unit for demand in class i that is not satisfied and lost, and β_t^2 is the penalty per unit for items that are backlogged for delayed fulfillment. Penalty terms for the first class are assumed to be higher than the second class.

We assume the demand of each class i in time period t, D_t^i, is a non-stationary stochastic function; the probability and cumulative distribution functions $(\phi_t^i, \mathbf{\Phi}_t^i)$ are known, continuous and differentiable; and that the customer classes are independent. We do not assume particular forms of the demand functions.

The net inventory (*on-hand inventory* $-$ *backlogs*) at the beginning of period t is I_t. At the beginning of a period, the manufacturer checks the inventory level and decides the production quantity; let S_t represent the inventory plus production in period t. We assume products arrive immediately, and the manufacturer fulfills the backorders carried from the previous period with the available inventory. (We allow backordered items to be delivered no more than one period later, and we restrict backorders in each period to be no more than the capacity in the next period; therefore, previously accepted orders are fulfilled before new orders are accepted.) Then the demand is realized during the period, and at the end of the period the manufacturer decides the amount of inventory to reserve for future sales and the amount of backorders to be promised in the current period for future fulfillment. Then the current demand is satisfied according to the inventory and backlogging decisions.

Let $J_t(I_t)$ be the expected profit from period t forward to the end of the horizon when starting at period t with I_t units in inventory, or the *profit-to-go* function. Let $G_t(S_t)$ be the expected profit-to-go from period t forward to the end of the horizon with S_t units of product available (after production). The first and second derivatives of $J_t(I_t)$ are denoted by $J_t'(I_t)$ and $J_t''(I_t)$, respectively. When the expected profit functions are specifically defined for a strategy, they will have an additional superscript indicating the strategy.

2.2 Time Differentiation Strategy

In the Time Differentiation Strategy (TDS), we assume that the manufacturer can differentiate the customer classes by offering two time-differentiated services: selling the item for p_t^1 and delivering the item immediately, or selling the item for the discounted price p_t^2 and delivering the item no later than one period later.

Let B_t be the maximum planned backorders in period t to fulfill from future capacity. In period

S_t^2	$= (S_t - R_t^1 - D_t^1)^+$	available inventory after first-class demand is satisfied
A_t	$= \min(B_t, [D_t^2 - [S_t^2 - R_t^2]^+]^+)$	actual backlogged orders
$I_{t+1}^{R^1}$	$= \min(S_t, R_t^1)$	inventory carried forward due to R^1 decision
$I_{t+1}^{R^2}$	$= \min(S_t^2, R_t^2)$	inventory carried forward due to R^2 decision
I_{t+1}^{low}	$= [S_t^2 - R_t^2 - D_t^2]^+$	inventory carried forward due to low demand

Table 1: Additional notation

t, R_t^1 is the inventory to protect from being sold to first and second-class customers, and R_t^2 is the additional inventory to protect from being sold to second-class customers; thus, the total amount to protect from class 2 in period t is $R_t^1 + R_t^2$. For convenience, define S_t^2 to be the amount of inventory available to the second-class customers in period t, the actual backlogged orders from second-class customers after all demand is satisfied in period t from available inventory as A_t, and inventory carried forward to period $t+1$ due to reserved amounts R^1 and R^2 as $I_{t+1}^{R^1}$ and $I_{t+1}^{R^2}$, respectively. If total demand is sufficiently *low* so that there is leftover inventory at the end of period t, we denote this inventory as I_{t+1}^{low}. See Table 1 for a summary of the additional notation.

In each period, after the demand for both classes of customers are revealed, the maximum number of first-class customers is satisfied from the available inventory on hand immediately, and the second-class customers are satisfied from the available inventory left over after the first-class demand is satisfied and from the available backlog amounts. The usage of on-hand inventory for a second-class demand instead of a first-class one is obviously sub-optimal, and the second-class customers may be satisfied immediately if there is inventory, since it avoids the inventory holding cost and backlogging penalty for those customers.

We model this resource allocation problem with service-differentiated customers as a Markov decision process, where the state of the system is represented by the net inventory. We can now write the optimal expected profit in period t for the Time Differentiation Strategy as the following recursive equation:

$$J_t^{TDS}(I_t) = \max_{S_t:\max(0,I_t)\le S_t\le I_t+q_t} \Big\{-c_t(S_t - I_t) + G_t^{TDS}(S_t)\Big\}, \quad \text{and} \tag{1}$$

$$G_t^{TDS}(S_t) = \max_{B_t:B_t\le q_{t+1};\, R_t^1, R_t^2: R_t^1+R_t^2\le S_t} g_t^{TDS}(S_t, R_t^1, R_t^2, B_t), \quad \text{where} \tag{2}$$

$$
\begin{aligned}
g_t^{TDS}(S_t, R_t^1, R_t^2, B_t) = \iint \quad & \Big\{ p_t^1 \min\big(D_t^1, S_t - R_t^1\big) + p_t^2 \min\big(D_t^2, [S_t^2 - R_t^2]^+ + B_t\big) \\
& -h_t(S_t^2 - R_t^2 - D_t^2)^+ - h_t \min(S_t^2, R_t^2) - h_t R_t^1 \\
& -\ell_t^1\big(D_t^1 - S_t + R_t^1\big)^+ - \ell_t^2\big(D_t^2 - [S_t^2 - R_t^2]^+ - B_t\big)^+ \\
& -\beta_t^2 \min\big([D_t^2 - [S_t^2 - R_t^2]^+]^+, B_t\big) \\
& +J_{t+1}^{TDS}\big((I_{t+1}^{low} + I_{t+1}^{R^1} + I_{t+1}^{R^2}) - A_t\big)\Big\} d\mathbf{\Phi}_t^1(D_t^1) d\mathbf{\Phi}_t^2(D_t^2). \qquad (3)
\end{aligned}
$$

Equation (1) includes production cost and the remaining profit-to-go after production (Equation (2)), which is maximized over the tactical inventory (R_t^1, R_t^2) and backlogging (B_t) decisions. The first terms in Equation (3) include the revenue from first-class customers from the available inventory and the revenue from the second-class demand from available inventory and planned backlogging. The second line (third, fourth and fifth terms) includes the holding cost for leftover inventory and for the two reserving inventory decisions. The sixth and seventh terms, respectively, are the rejection penalties for unsatisfied first and second-class demand, and the eighth term is the delay penalty for backlogged demand. The last term in the equation is the profit-to-go in future periods, as a function of any leftover physical inventory and backlogged orders. For the last period of the horizon (T), the final term is replaced by $\upsilon(S_T - D_T^1 - D_T^2)^+$, which includes the salvage cost for the leftover inventory. The constraints ensure that the manufacturer does not backlog more future capacity than he has in the next period or reserve more inventory than is available.

To simplify the TDS problem, we show that in an optimal policy, in every period either the amount of inventory protected from the second-class customers or the amount of backlogged demand of the second-class demand must equal zero (or both).

Lemma 1. *In any optimal policy under the Time Differentiation Strategy, we have:*

$$B_t \cdot (R_t^1 + R_t^2) = 0 \qquad t = 1, 2, \ldots, T.$$

See [3] or [5] for proofs of results. To see the result intuitively, suppose that $(R_t^1 + R_t^2) > 0$, which means that the manufacturer may reject some current second-class demand in period t in order to reserve some inventory for period $t+1$. Then it is intuitive that it would not be optimal for the manufacturer to use the inventory in period $t+1$ to fulfill any current second-class demand in period t, thus we will have $B_t = 0$. The intuitive explanation for the case with $B_t > 0$ is similar.

Lemma 1 implies that the structure of the optimal policies can be simplified as follows. In each period, the manufacturer can choose one of two policies: either the Reserve-Inventory policy with

$R_t^1 + R_t^2 \geq 0$, or the Backlog-Demand policy with $B_t \geq 0$. Thus,

$$G_t^{TDS}(S_t) = \max\Big\{G_t^{TDS-R}(S_t), G_t^{TDS-B}(S_t)\Big\},$$

where $G_t^{TDS-R}(S_t)$, and $G_t^{TDS-B}(S_t)$ represent the profit-to-go with S_t units of products available after production under the Reserve-Inventory policy and the Backlog-Demand policy, respectively. These policies are defined by

$$G_t^{TDS-R}(S_t) = \max_{R_t^1, R_t^2 : R_t^1 + R_t^2 \leq S_t} \Big\{g_t^{TDS}(S_t, R_t^1, R_t^2, 0)\Big\} \quad and \quad G_t^{TDS-B}(S_t) = \max_{B_t : B_t \leq q_{t+1}} \Big\{g_t^{TDS}(S_t, 0, 0, B_t)\Big\}.$$

We will address the structural results and corresponding policies for these models in Section 3, after introducing the non-differentiating strategy.

2.3 Common Service Strategy

In some cases, even though the manufacturer knows the existence of multiple classes of customers, he may not be able or willing to treat customers differently. In such environments, the manufacturer manages the customers as a single class, and attempts to serve each customer with the same service strategy. We model the problem of a manufacturer who does not differentiate between two classes of customers with the *Common Service Strategy* (CSS), where the manufacturer serves customers with a first-come-first-serve rule and offers all customers a one-period backlog for the item if the on-hand inventory is depleted. The second-class customers will accept the delayed fulfillment, but the first-class demand is lost if it is not fulfilled immediately; all customers are willing to accept immediate fulfillment.

We assume the manufacturer takes the second-class customers' reservation price, p_t^2, as the selling price to all customers, although our results also hold under other prices. The total amount of demand (class 1 and class 2 together) in period t is $D_t^{1,2} = D_t^1 + D_t^2$ and the total demand has the probability and cumulative distribution functions $(\phi_t^{1,2}, \mathbf{\Phi}_t^{1,2})$. We let α_t be the average proportion of demand from the second class in period t, i.e., $\alpha_t = E[D_t^2]/E[D_t^1 + D_t^2]$. We assume that the customer classes are distributed homogeneously across a time period in accordance with α_t. (Note that if this assumption does not hold, the model becomes an approximation of the true situation.) Let ℓ_t be rejection penalty in a period t; e.g., in our calculations we use ℓ_t as the weighted average rejection penalty ($\ell_t = (1 - \alpha_t)\ell_t^1 + \alpha_t \ell_t^2$), but other values can also be used. In each period t, the manufacturer decides B_t, the amount of planned backlogging in the current period; R_t, the inventory to protect from being sold in the current period; and S_t, the target level of inventory.

The optimal decisions are found by solving the profit-to-go function under the Common Service Strategy:

$$J_t^{CSS}(I_t) = \max_{S_t:\max(0,I_t)\le S_t\le I_t+q_t} \Big\{-c_t(S_t - I_t) + G_t^{CSS}(S_t)\Big\}, \text{ and} \tag{4}$$

$$G_t^{CSS}(S_t) = \max_{R_t:R_t\le S_t;\ B_t:B_t\le q_{t+1}} g_t^{CSS}(S_t, R_t, B_t) \quad \text{where,} \tag{5}$$

$$\begin{aligned} g_t^{CSS}(S_t, R_t, B_t) = \int \Big\{ & p_t^2 \min(D_t^{1,2}, S_t - R_t + \min(B_t, \alpha_t(D_t^{1,2} - S_t + R_t)^+)) \\ & -h_t \max(R_t, S_t - D_t^{1,2}) - \beta_t^2 \min(B_t, \alpha_t(D_t^{1,2} - S_t + R_t)^+) \\ & -\ell_t^1(1-\alpha_t)\min(B_t/\alpha_t, (D_t^{1,2} - S_t + R_t)^+) - \ell_t(D_t^{1,2} - S_t + R_t - B_t/\alpha_t)^+ \\ & +J_{t+1}^{CSS}(\max(R_t, S_t - D_t^{1,2}) - \min(\alpha_t(D_t^{1,2} - S_t + R_t)^+, B_t))\Big\} d\mathbf{\Phi}_t^{1,2}(D_t^{1,2}) \end{aligned} \tag{6}$$

Equations (4) and (5) are as described before, except in the CSS strategy the latter is optimized over fewer reserving decisions; other differences are as below. The selling revenue includes items from both classes sold immediately, and any items backlogged from the second class only. Delay penalties are charged for backlogged second-class demand, and penalties are paid for first-class demand not satisfied immediately. The fourth term is the penalty associated with the lost first-class demand who are offered delayed fulfillment but are not willing to accept it, and the fifth term is the rejection penalty for demand beyond the acceptance level for both classes. The last term in Equation (6) is again the profit-to-go, and constraints are as before.

Next we show that in an optimal policy, in any period, either the amount of reserved inventory equals zero or the amount of backlogged demand equals zero, i.e., they cannot both be positive.

Lemma 2. *In any optimal policy under the Common Service Strategy, we have $R_t \cdot B_t = 0$, for $t = 1, 2, ..., T$.*

This is similar to the result for TDS, except now it applies to the reserving decision that is common to the two customer classes.

As before, with Lemma 2, the structure of the optimal policies can be simplified. Under the Common Service Strategy, in any period the manufacturer chooses one of two policies: either he protects inventory for the future and does not backlog current demand ($R_t \ge 0, B_t = 0$), called the Reserve-Inventory policy, or he backlogs current demand but does not save items for the future ($B_t \ge 0, R_t = 0$), called the Backlog-Demand policy. Thus,

$$G_t^{CSS}(S_t) = \max\Big\{G_t^{CSS-R}(S_t), G_t^{CSS-B}(S_t)\Big\}, \qquad \text{where}$$

$$G_t^{CSS-R}(S_t) = \max_{R_t: 0 \le R_t \le S_t} \left\{ g_t^{CSS}(S_t, R_t, 0) \right\} \quad \textit{and} \quad G_t^{CSS-B}(S_t) = \max_{B_t: 0 \le B_t \le q_{t+1}} \left\{ g_t^{CSS}(S_t, 0, B_t) \right\}.$$

There are similarities in the structure of the results for TDS and CSS, although the models have several important differences. In the next section we further analyze similarities in the structure.

3 Results

Under both the Time Differentiation Strategy and the Common Service Strategy, we can show that the four profit-to-go functions $g_t^{TDS}(S_t, R_t^1, R_t^2, 0)$, $g_t^{TDS}(S_t, 0, 0, B_t)$, $g_t^{CSS}(S_t, R_t, 0)$, and $g_t^{CSS}(S_t, 0, B_t)$ are quasi-concave, each of them has a unique unconstrained optimizer that is independent of the inventory level S_t, and the expected profit $J_t(I_t)$ and $G_t(S_t)$ are concave functions of inventory I_t and S_t respectively. These results are summarized in the following theorem:

Theorem 3. *For all $t = 1, ..., T$,*

- $g_t^{TDS}(S_t, R_t^1, R_t^2, 0)$ *is a jointly quasi-concave function of* R_t^1 *and* R_t^2*, and* $g_t^{CSS}(S_t, R_t, 0)$ *is a quasi-concave function of* R_t*,*
- $g_t^{TDS}(S_t, 0, 0, B_t)$ *and* $g_t^{CSS}(S_t, 0, B_t)$ *are quasi-concave functions of* B_t*,*
- $G_t^{TDS}(S_t)$ *and* $G_t^{CSS}(S_t)$ *are concave functions of* S_t*,*
- $J_t^{TDS}(I_t)$ *and* $J_t^{CSS}(I_t)$ *are concave functions of* I_t*,*
- *The unconstrained optimizers ($R_t^{1^*}$, $R_t^{2^*}$, and B_t^*) for functions* $g_t^{TDS}(S_t, R_t^1, R_t^2, 0)$*, and* $g_t^{TDS}(S_t, 0, 0, B_t)$*, are independent of inventory level* S_t*, where for* $i = 1, 2$*,*

$$(R_t^{1^*}(S_t), R_t^{2^*}(S_t)) = \underset{(R_t^1, R_t^2): 0 \le R_t^1, 0 \le R_t^2}{\operatorname{argmax}} \left\{ g_t^{TDS}(S_t, R_t^1, R_t^2, 0) \right\} \textit{ and } B_t^*(S_t) = \underset{B_t: 0 \le B_t}{\operatorname{argmax}} \left\{ g_t^{TDS}(S_t, 0, 0, B_t) \right\}.$$

- *The unconstrained optimizers (R_t^* and B_t^*) for* $g_t^{CSS}(S_t, R_t, 0)$ *and* $g_t^{CSS}(S_t, 0, B_t)$ *are independent of inventory level* S_t*, where*

$$R_t^*(S_t) = \underset{R_t: 0 \le R_t}{\operatorname{argmax}} \left\{ g_t^{CSS}(S_t, R_t, 0) \right\}, B_t^*(S_t) = \underset{B_t: 0 \le B_t}{\operatorname{argmax}} \left\{ g_t^{CSS}(S_t, 0, B_t) \right\}.$$

Theorem 3 implies an optimal policy for both the Time Differentiation Strategy and the Common Service Strategy that has a similar form, and thus we have the following corollary.

Corollary 4. *Given a vector of prices, there exists an optimal policy for*

- *the Time Differentiation Strategy with an optimal order-up-to level (S_t^*), optimal reserve-up-to-levels ($R_t^{1^*}$ and $R_t^{2^*}$), and an optimal backlog-up-to level (B_t^*),*

- *the Common Service Strategy with an optimal order-up-to level (S_t^*), an optimal reserve-up-to-level (R_t^*) and an optimal backlog-up-to level (B_t^*).*

Note that for CSS there is a single reserve inventory decision that non-discriminatingly applies to both classes of customers, and similarly for the planned backlogging decisions, while for TDS there are separate values for reserving that apply to each class and the planned backlogging only applies to the second-class demand. However, in both cases the form of the optimal policy is (S, R, B). In both cases the optimal policies are considered to be modified base stock ones, because the realized values may be limited by capacity or available inventory. The results also show that the optimal inventory decisions are independent of the realized demand, which implies that the decisions could also have been made before the exact demand realization.

Examining the decisions in more detail provides more information on their meaning. The optimal decisions for CSS are defined by the following:

$$\begin{aligned} S_t^* &= \max\{S : c_t \leq G_t^{'CSS}(S)\} && if\; c_t \leq G_t^{'CSS}(0) \\ R_t^* &= \max\{I : p_t^2 + \ell_t + h_t \leq J_{t+1}^{'CSS}(I)\} && if\; p_t^2 + \ell_t + h_t < J_{t+1}^{'CSS}(0) \\ B_t^* &= \min\{I : J_{t+1}^{'CSS}(-I) \geq p_t^2 + \ell_t^2 - \beta_t^2\} && if\; p_t^2 + \ell_t^2 - \beta_t^2 > J_{t+1}^{'CSS}(0), \end{aligned} \tag{7}$$

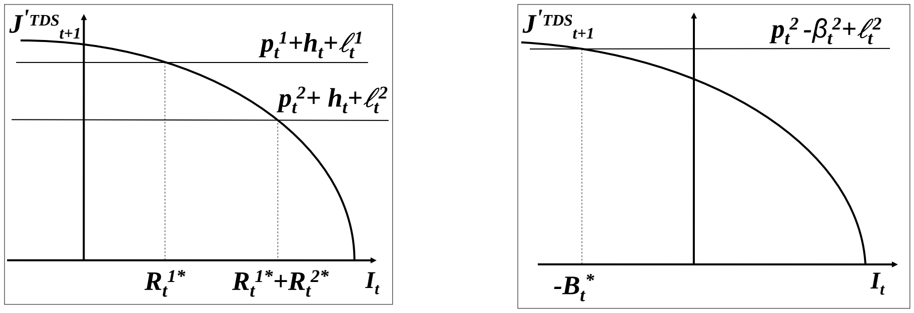
and the optimal decisions for TDS that are different from CSS are given by:

$$\begin{aligned} R_t^{1^*} &= \max\{I : p_t^1 + \ell_t^1 + h_t \leq J_{t+1}^{'TDS}(I)\} && if\; p_t^1 + \ell_t^1 + h_t < J_{t+1}^{'TDS}(0) \\ R_t^{1^*} + R_t^{2^*} &= \max\{I : p_t^2 + \ell_t^2 + h_t \leq J_{t+1}^{'TDS}(I)\} && if\; p_t^2 + \ell_t^2 + h_t < J_{t+1}^{'TDS}(0). \end{aligned} \tag{8}$$

For each decision, if the condition is not satisfied, then the decision variable equals zero.

In each case, a decision is found by comparing the net revenue from gaining or losing a customer with the marginal expected profit of an additional unit in the future ($p_t^i + \ell_t^i + h_t$ is the net revenue of selling to customer class i from inventory, $p_t^i + \ell_t^i - \beta_t^i$ is the net revenue from backlogging an i class customer, and $J'(I)$ is the marginal expected profit of an additional unit above I). This is also apparent from examining the decisions for TDS that are pictured in Figure 1 (the ones for CSS are similar in structure).

Previous work [4] examined the form of the optimal policies when customers are differentiated by priority level and behave homogeneously with respect to delayed fulfillment. In that case, the



(a) Reserve-Inventory Policy (b) Backlog-Demand Policy

Figure 1: Optimal Decisions for TDS under the Optimal Policies

form of the decisions was similar (i.e., (S, R, B)), but a different set of decisions and policies resulted from the model. An obvious difference is that in the current research only one backlogging decision results, since only the second-class customers are willing to accepted delayed fulfillment. Another difference is that in [4], one submodel that we found could be optimal was the Reserve-and-Backlog Policy, where both a backlogging and reserving decision could be positive in the current period, but that is not true for the TDS and CSS models.

Numerical experiments of the TDS and CSS policies show that significant profit improvement can be achieved with the tactical inventory, especially when production capacity is limited. See Duran [3] for details.

4 Conclusions

Many companies today provide differentiated service so that some customers are served immediately while others receive delayed fulfillment for a discount. Customers receive higher utility, and the company may gain flexibility to improve operations. This paper contributes to the literature on operational models to manage markets with segmented demand, and it shows the impact of one kind of flexibility in the production system. Additional research would also be helpful in showing how to manage systems with segmented demand. For instance, pricing can be used to determine the size of the customer classes in response to variability, and policies to manage systems with extended leadtimes and multiple classes is another. The area is rich and has many applications in modern manufacturing and e-tailing companies that may be experimenting with different business models.

5 Acknowledgments

Research supported in part by NSF grants DMI-0245352 and DMI-0348532 and General Motors.

References

[1] L. M. A. Chan, D. Simchi-Levi, and J. L. Swann. Pricing, production, and inventory policies for manufacturing with stochastic demand and discretionary sales. *Manufacturing and Service Operations Management*, 8(2):149–168, 2006.

[2] V. Deshpande, M. A. Cohen, and K. Donohue. A threshold inventory rationing policy for service-differentiated demand classes. *Management Science*, 49(6):683–703, 2003.

[3] S. Duran. *Optimizing demand management in stochastic systems to improve flexibility and performance*. PhD thesis, Georgia Institute of Technology, 2007.

[4] S. Duran, T. Liu, D. Simchi-Levi, and J. Swann. Optimal production and inventory policies of priority and price-differentiated customers. *Forthcoming in IIE Transactions*, 2006.

[5] S. Duran, T. Liu, D. Simchi-Levi, and J. Swann. Proofs for "Policies utilizing tactical inventory for service-differentiated customers". Available at http://www.isye.gatech.edu/∼jswann/, 2007.

[6] K. C. Frank, R. Q. Zhang, and I. Duenyas. Optimal policies for inventory systems with priority demand classes. *Operations Research*, 51(6):993–1002, 2003.

[7] D. Gupta and L. Wang. Manufacturing capacity revenue management. *Forthcoming in Operations Research*, 2006.

[8] H. E. Scarf. Optimal inventory policies when sales are discretionary. Cowles Foundation Discussion Paper No. 1270, 2000.

[9] M. J. Sobel and R. Q. Zhang. Inventory policies for systems with stochastic and deterministic demand. *Operations Research*, 49(1):157–162, 2001.