November, 2011

An optimal policy for joint dynamic price and lead-time quotation

Jiejian FENG
Liming LIU, *Lingnan University, Hong Kong*
Xianming LIU

OPERATIONS RESEARCH
Vol. 59, No. 6, November–December 2011, pp. 1523–1527
ISSN 0030-364X | EISSN 1526-5463 | 11 | 5906 | 1523
http://dx.doi.org/10.1287/opre.1110.0969

TECHNICAL NOTE

An Optimal Policy for Joint Dynamic Price and Lead-Time Quotation

Jiejian Feng
School of Business, Saint Mary's University, Halifax, Nova Scotia B3H 3C3, Canada, jiejian.feng@smu.ca

Liming Liu
Faculty of Business, Lingnan University, Kowloon, Hong Kong, China, limingliu@ln.edu.hk

Xiaoming Liu
Faculty of Business Administration, University of Macao, Macao, China, xmliu@umac.mo

For a dynamic joint price and lead-time quotation problem with a fairly general demand function, we show that the policy consisting of a threshold and a reward-maximizing lead-time is optimal. This policy offers some interesting managerial insights. Under this policy, finding the exact optimal quotation can be accomplished by single-variable policy iterations of unimodal value functions.

Subject classifications: dynamic pricing; lead-time quotation; threshold policy.
Area of review: Manufacturing, Service, and Supply Chain Operations.
History: Received July 2008; revisions received July 2009, February 2010, June 2010; accepted December 2010.

1. Introduction

For make-to-order (MTO) suppliers, offering product delivery lead times together with prices to customers is a standard way of doing business. Because competition is increasingly time based, combining a fast and reliable lead-time commitment with flexible pricing for each individual order has been adopted by more MTO suppliers as a competitive strategy. Because customers often need to schedule their own production or equipment installations, business and industrial equipment suppliers, such as Siemens and General Electric, have to offer precise delivery lead times based on their workload status and set flexible prices strategically to secure customer orders and maintain profitability.

Although the actual decision on the price and lead time for an order is usually made by a sales person or middle management, how this decision should be made is of strategic importance to the supplier because it affects the use of a firm's limited resources and impacts the firm's market share and profitability directly. In other words, the front-line staff who work with customers directly need to follow a policy or rule that guides them to make the right price and leadtime decisions for each customer order. Such a policy is of strategic importance to a supplier and should be carefully formulated with top management involvement.

Whereas the literature on static lead time or joint price and lead-time decision problems is extensive (e.g., Baker 1984, Boyaci and Ray 2003, Hopp and Spearman 1993, Li 1992, Liu et al. 2007, Lutze and Özer 2008, Palaka et al. 1998, So and Song 1998, Wein 1991, Yano 1987), the dynamic version has been studied by only a few. Duenyas and Hopp (1995) considered the dynamic lead-time quotation problem in an $M/M/1$-queue setting. They showed that the optimal quoted lead time is increasing in the queue length or the number of outstanding orders. Duenyas (1995) extended this result to the case with multiple customer classes. Assuming that the processing time becomes known exactly when a job arrives, Slotnick and Sobel (2005) showed that according to the values of the known processing time of the new order and the remaining workload, there are four cases for the optimal quotation: rejection; zero lead time; maximum lead time; and general lead time. Plambeck (2004) showed that it is asymptotically optimal, under heavy traffic, to quote the more patient customer a delivery time proportional to the current queue length and give the more time-sensitive customer a higher priority. Ata and Olsen (2009) provided near-optimal dynamic lead-time quotations when the waiting cost is convex, concave, or convex-concave.

The purpose of this work is to find a prototype dynamic price and lead-time quotation policy that takes into consideration a firm's production capability, cost structure, and customers' purchasing behaviors. Such a policy can be helpful to firms for developing rules for workload-dependent price and lead-time quotations in real time. We present the system and model definition in §2. In

§3, we show the optimality of the (T, RML) policy and express the optimal lead time analytically. We prove that the value function is unimodal in price for every policy iteration so that the optimal price can be easily computed. In §4, we summarize the results and discuss some possible extensions.

2. Problem Definition and the Model

2.1. System Setting

We consider a firm that supplies a product or service on order. Customers arrive according to a renewal process with an interarrival time distribution $A(x)$ and arrival rate λ. The processing time follows an exponential distribution with rate μ. Orders are fulfilled (processed) one at a time in a first-in-first-out sequence. When an existing order is being processed, the newly confirmed order joins a job queue. Thus, the firm's supply system and its potential demands are modeled by the $GI/M/1$ queue.

Let s denote the job queue length, which includes the job currently being processed. The firm quotes a price and a delivery lead time to a new customer according to s. When a quotation is rejected by the customer, the firm incurs a loss-of-goodwill cost c_g. When the quotation is accepted, a revenue p will be generated from this order at a cost of c (for materials and production). The firm experiences a delay penalty of c_t per unit time when the order completion time is longer than the quoted lead time.

2.2. Demand Function

Waiting can be costly to customers. Although waiting costs may be different for different customers, a firm may not be able to distinguish the types of individual customers in real time. We assume that what the firm can infer from market research is a homogeneous waiting cost $x(l)$ as a function of the lead time l for all customers, and a customer will make the purchase when his/her valuation of the product is greater than the total cost $p + x(l)$ for a given quotation $a = (p, l)$. The firm also knows the cumulative distribution function $1 - q(p + x(l))$ of an arbitrary customer's valuation of the product. Clearly, $q(p + x(l))$ is our demand function and should be nonincreasing in both p and l. The marginal waiting cost $x'(l)$ is nonnegative, and we make two more assumptions on the waiting cost.

Assumption 1. *For analytic tractability, we assume convex delay cost, i.e.,* $x''(l) \geqslant 0$.

Assumption 2. *Anticipated waiting cost is smaller than the late delivery penalty cost, i.e.,* $x'(0) < c_t$.

Note that the convexity assumption is also used in Ata and Olsen (2009). Assumption below is for technicality.

Assumption 3. *There exist a minimum price* $p_{\min} \geqslant 0$, *a maximum price* $p_{\max}$, *and a maximum lead time* $l_{\max}$ *such that* $q(p_{\min} + x(0)) = 1$, $q(p_{\max} + x(l)) = 0$ *for any* $l \geqslant 0$, *and* $q(p + x(l_{\max})) = 0$ *for any* $p \geqslant p_{\min}$; *and either* p *is bounded or* $\lim_{p\to\infty} q(p + x(0))p = 0$.

With these assumptions, we define a demand model that covers the linear and exponential models that are widely used in the literature, e.g., Boyaci and Ray 2003, Liu et al. 2007, Palaka et al. 1998, Plambeck 2004, Duenyas and Hopp 1995, Slotnick and Sobel 2005. For example, let $q(p + x(l)) = -\tau_1(p + x(l)) + C$ and $x'(l) = \tau_2/\tau_1$. We have $x(l) = (\tau_2/\tau_1)l + C_1$ and $q(p + x(l)) = -\tau_1(p + (\tau_2/\tau_1)l) + C_2$, where $C_2 = C + C_1$. Thus, in this case, our demand model reduces to the commonly used linear model. By Assumption 3, we have $C_2 = 1 + \tau_1 p_{\min}$, and hence $q(p + x(l)) = 1 - \tau_1(p - p_{\min}) - \tau_2 l$. By Assumption 3 again, we also have $p_{\max} = 1/\tau_1 + p_{\min}$ and $l_{\max} = 1/\tau_2$. Similarly, we can construct the exponential demand function.

A limitation of our demand model lies in the assumption that the waiting cost is independent of price. Although this assumption is often reasonable in practice, there are situations when a customer's tolerance to waiting is also affected by the price.

2.3. Decision Model

We formulate an infinite-horizon discounted profit-maximization model for our decision problem. Let $\mathbf{a} = (a_0, a_1, \ldots, a_s, \ldots, a_S)$ be the action vector of the firm for the corresponding system states, where S is a (large) finite cut-off (upper bound) system state at which new orders will be rejected. Let $v_s(a_s)$ be the total discounted expected profit over an infinite horizon starting from the current arrival epoch when the system state is s and action a_s is taken, and $\mathbf{v}(\mathbf{a}) = (v_0(a_0), v_1(a_1), \ldots, v_s(a_s), \ldots, v_S(a_S))$. Similarly, $r_s(a)$ denotes the discounted expected reward generated from the current order. We have

$$r_s(a) = q(p + x(l))(p - D_s(l) - c) - (1 - q(p + x(l)))c_g, \tag{1}$$

where $D_s(l) = c_t \int_l^\infty \int_l^w \mathrm{e}^{-\alpha y}\,\mathrm{d}y\,\mathrm{d}F_s(w)$ is the expected discounted delivery cost, α is the discount rate, and $F_s(w)$ is the distribution function of the time from the confirmation to the completion of this order. We will drop subscript s from a and use a and (p, l) interchangeably.

Let $u_{sj}(a \mid y)$ be the probability that the system state changes from s prior to taking the action $a = (p, l)$ to j, y time units after taking action a but before the arrival of the next customer. Let $N(y)$ be the number of jobs that can be finished within time y, given that there are ample jobs to process. It follows a Poisson distribution with rate μy. Hence, $u_{sj}(a \mid y) = q(p + x(l))P([s + 1 - N(y)]^+ = j) + (1 - q(p + x(l)))P([s - N(y)]^+ = j) \equiv q(p + x(l))u_{sj1}(y) + u_{sj2}(y)$, where

$$u_{sj1}(y) = \begin{cases} -\dfrac{(\mu y)^s}{s!}\mathrm{e}^{-\mu y}, & j = 0 \\[2ex] \dfrac{(\mu y)^{s-j}}{(s-j)!}\left(\dfrac{\mu y}{s+1-j} - 1\right)\mathrm{e}^{-\mu y}, & 0 < j < s+1 \\[2ex] \mathrm{e}^{-\mu y}, & j = s+1 \\[1ex] 0, & j > s+1 \end{cases}$$

$$u_{sj2}(y) = \begin{cases} \mathrm{e}^{-\mu y} \sum_{i=s}^{\infty} \frac{(\mu y)^s}{s!}, & j = 0 \\ \frac{(\mu y)^{s-j}}{(s-j)!} \mathrm{e}^{-\mu y}, & 0 < j < s+1 \\ 0, & j = s+1 \\ 0, & j > s+1. \end{cases}$$

Let $m_{sj}(\mathbf{a})$ be the discounted transition probability from state s prior to the current action a to state j right before the next action. Then, $m_{sj}(\mathbf{a}) = \int_0^\infty \mathrm{e}^{-\alpha y} u_{sj}(a \mid y)\, dA(y) = q(p + x(l))\, m_{sj1} + m_{sj2}$, where m_{sj1} and m_{sj2} are independent of the action vector and can be computed once the order interarrival time distribution is given. Because for any given action $\mathbf{a}$ the evolution of the system state satisfies the Markovian property, our decision problem of finding the optimal price and lead time that maximize $\mathbf{v}$ is an SMDP with state variable s, decision variables p and l, value vector $\mathbf{v}(\mathbf{a})$, and discounted probability matrix $\mathbf{M}(\mathbf{a}) = (m_{sj}(\mathbf{a}))_{(S+1)\times(S+1)}$.

Because $0 \leqslant q(p + x(l)) \leqslant 1$ and by Assumption 3, we have $D_s(l) \leqslant c_t/\alpha \Rightarrow |r_s(a)| \leqslant q(p + x(l))p + c_t/\alpha + c + c_g$, i.e., $\mathbf{r}(\mathbf{a})$ is bounded. By the standard theory on SMDP (e.g., Puterman 1994, Theorems 11.3.1 and 11.3.2), we have $\mathbf{v}(\mathbf{a}) = \mathbf{r}(\mathbf{a}) + \mathbf{M}(\mathbf{a})\mathbf{v}(\mathbf{a})$, and there exist a unique optimal policy $\mathbf{a}^*$ and a unique optimal reward vector $\mathbf{v}^*$ that satisfy the following optimality equation,

$$\mathbf{v} = \max_{\mathbf{a}} \{\mathbf{r}(\mathbf{a}) + \mathbf{M}(\mathbf{a})\mathbf{v}\}. \tag{2}$$

3. The Optimal (T, RML) Policy

For a given acceptance probability q, the price $p_q(l)$ is concave decreasing in l by Assumption 1. For state s, let $R_{q,s}(l) \equiv p_q(l) - D_s(l)$, $l \geqslant 0$. $R_{q,s}(l)$ is the expected reward from serving one customer with price p and lead time l that give an acceptance probability q. Different from $r_s(a)$, this is the reward when the service has actually taken place.

Lemma 1. *The reward function $R_{q,s}(l)$ is unimodal in l.*

Proof. Taking derivative on $R_{q,s}(l)$, we have $R'_{q,s}(l) = -x'(l) - dD_s(l)/dl = c_t(1 - F_s(l))\mathrm{e}^{-\alpha l} - x'(l)$. $R'_{q,s}(l)$ is strictly decreasing in l, and it will remain negative once it falls through zero. Because $R'_{q,s}(0) = c_t - x'(0) > 0$ by Assumption 2, $R_{q,s}(l)$ is unimodal. □

Let $l_{r,s}$ denote the root of $c_t(1 - F_s(l))\mathrm{e}^{-\alpha l} - x'(l) = 0$. We can show that $l_{r,s}$ is increasing in s because $F_{s_2}(l) < F_{s_1}(l)$ for all $l > 0$ and $s_1 < s_2$. Without considering the constraint imposed by $l_{\max}$, $l_{r,s}$ is the maximizer of $R_{q,s}(l)$, and we call it the *reward-maximizing lead time* or RML. Consider inequality

$$c_t(1 - F_s(l_{\max}))\mathrm{e}^{-\alpha l_{\max}} - x'(l_{\max}) \geqslant 0. \tag{3}$$

Proposition 1. *There exists a finite s such that* (3) *holds if and only if $c_t > x'(l_{\max})\mathrm{e}^{\alpha l_{\max}}$.*

Proof. If $l_{\max} = \infty$, neither (3) nor $c_t > x'(l_{\max})\mathrm{e}^{\alpha l_{\max}}$ can hold. For $l_{\max} < \infty$, the result can be derived by the fact that $F_s(l_{\max})$ is strictly decreasing in s and

$\lim_{s \to +\infty} F_s(l_{\max}) = 0.$ □

Definition 1. Let $T \equiv \min\{s\colon (3) \text{ holds}\}$ if (3) holds for some $s \leqslant S - 1$; otherwise, let $T \equiv S$.

We will show that T is an admission threshold, but first, we need the following monotone property.

Lemma 2. *v_s^* is a nonincreasing function of s.*

Proof. The delivery cost $D_s(l)$ is clearly increasing in s for any given value of l. Thus, for any given $a = (p, l)$, $r_s(a)$ is nonincreasing in s. We now construct two sample paths: Sample Path A (SPA) starts from state s_1 and Sample Path B (SPB) starts from state s_2, where $s_1 < s_2$. We compare the summations of discounted rewards $v_{s_1}^0$ and $v_{s_2}^0$ from every arrival in the future for the two sample paths. Let the future customer arrival processes and service times for SPA and for SPB be identical, and the same optimal action for SPB is applied to both SPA and SPB (for example, if at the next customer arrival epoch SPA and SPB have two and four unfinished jobs, respectively, the two new customers in SPA and SPB would both be quoted (p_4^*, l_4^*)). Clearly, the system state at every customer arrival epoch in SPA is always lower than or equal to that in SPB. Hence, the corresponding reward r generated in SPA is greater than or equal to that in SPB at every customer arrival epoch in future. Thus, the summation of time-discounted rewards in future for SPA is greater than that for SPB, i.e., $v_{s_1}^0 \geqslant v_{s_2}^0$. Because SPB follows the optimal policy and SPA does not, we have, $v_{s_1}^* \geqslant v_{s_1}^0 \geqslant v_{s_2}^0 = v_{s_2}^*$. □

From the standard result for SMDP, we can show that if $a_s^* = (p_s^*, l_s^*)$ maximizes $\psi_{s,\mathbf{v}^*}(a)$ for every state s, $\mathbf{a}^*$ is a stationary optimal policy where $\psi_{s,\mathbf{v}}(a)$ is the sth element of the vector to be maximized on the right-hand side of (2), i.e.,

$$\begin{aligned}\psi_{s,\mathbf{v}}(a) &= r_s(a) + \mathbf{M}_s(a)\mathbf{v} \\ &= \sum_j m_{sj2} v_j - c_g + q(p + x(l)) \\ &\quad \cdot \Big(p - D_s(l) - c + c_g + \sum_j m_{sj1} v_j\Big).\end{aligned} \tag{4}$$

We note that, ignoring other factors, the firm is indifferent to selling at price $c - c_g$ or to losing sales. On the other hand, in a competitive market, there must be some customers whose valuations of the firm's product are below $c - c_g$. Thus, the total acquisition cost for these customers for their unconditional acceptance should not be greater than $c - c_g$, i.e., we should have $p_{\min} \leqslant c - c_g$.

Theorem 1. *Suppose $p_{\min} \leqslant c - c_g$. For any system state s, it is optimal to quote according to the following rules where T and $l_{r,s}$ are given earlier in this section*:

(1) *When $s \geqslant T$, reject the new customer*;

(2) *When $s < T$, $l_{r,s} < l_{\max}$ and $l_s^* = l_{r,s}$, i.e., the optimal lead time is the* RML, *and the optimal price can be obtained by single-variable policy iterations.*

Proof. For any given v, the maximization in (2) is equivalent to maximizing $\psi_{s,\mathbf{v}}(a)$ for every s, and we have $\max_a\{\psi_{s,\mathbf{v}}(a)\} = \max_{q_s}\{\max_{a:\, q(p+x(l))=q_s}\{\psi_{s,\mathbf{v}}(a)\}\}$. For any given positive q_s, all the terms in $\psi_{s,\mathbf{v}}(a)$ except $p - D_s(l)$ are independent of the decision variables according to (4). Thus, for given q_s, if l maximizes $R_{q,s}(l) = p - D_s(l)$, it also maximizes $\psi_{s,\mathbf{v}}(a)$ and vice versa.

By Lemma 1 and Definition 1, it is easy to show that $l_{r,s} < l_{\max}$ when $s < T$ and $l_{r,s} \geqslant l_{\max}$ when $s \geqslant T$. Clearly, $R_{q,s}(l)$ increases in l in $[0, l_{\max}]$ if $s \geqslant T$. Given q, l increases when p decreases and $R_{q,s}(l)$ is maximized at $(p_{\min}, l_q)$ for some $l_q \leqslant l_{\max}$. However, we can show that it is impossible to have $l_q < l_{\max}$. Suppose this is not the case and $l_q < l_{\max}$ for some $s \geqslant T$. Then, a new customer will accept the quote with a probability $q_a = q(p_{\min}, l_q)$. The reward from this customer is $q_a[p_{\min} - c - D_s(l_q)] + (1-q_a)(-c_g) < q_a(p_{\min} - c) + (1-q_a)(-c_g) \leqslant -c_g$, worse than rejecting the customer by quoting $l_{\max}$. Furthermore, with $l_q < l_{\max}$, the system state will increase to $s+1$ with $q_a > 0$, and the expected future reward will also be smaller or equal to that of rejecting the customer now by Lemma 2. Thus, we must have $l_q = l_{\max}$, i.e., it is optimal to reject the new customer when $s \geqslant T$.

Consider $s < T$. We may construct a probability frontier curve $q(p + x(l)) = q(p_{\min}, l_{r,s})$ that cuts the whole decision domain into two regions: $D_1 = \{a\colon q(p + x(l)) > q(p_{\min} + x(l_{r,s}))\}$ and $D_2 = \{a\colon q(p + x(l)) \leqslant q(p_{\min} + x(l_{r,s})\}$, such that

$$\max_a\{\psi_{s,\mathbf{v}^*}(a)\} = \max\left\{\max_{a\in D_1}\{\psi_{s,\mathbf{v}^*}(a)\},\ \max_{a\in D_2}\{\psi_{s,\mathbf{v}^*}(a)\}\right\}.$$

For region D_1, $l < l_{r,s}$ so that $R'_{q,s}(l) > 0$. Thus, $\psi_{s,\mathbf{v}^*}(a)$ is maximized on the line $p = p_{\min}$ in D_1. By the same argument as above, we conclude that the maximum reward achievable in this region is worse than the reward $-c_g + v_s^*$ from $(p_{\max}, l_{r,s})$. For region D_2, $R_{q,s}(l)$ increases when $l < l_{r,s}$ and decreases when $l > l_{r,s}$ by Lemma 1. Thus, $R_{q,s}(l)$ is maximized on $l = l_{r,s}$ in this region and so is $\psi_{s,\mathbf{v}^*}(a)$. We conclude that the optimal solution must be in D_2, and thus it is optimal to quote $l_{r,s}$ for $s < T$. $\square$

The (T, RML) policy offers some interesting managerial insights on the strategy for handling customer orders in a competitive and time-sensitive market. We discuss them in the following remarks.

Remark 1. The dynamics between the firm's outstanding workload and the customer's utility drive the admission decision: When there are no outstanding orders, accept all customers except those whose utility is too low, e.g., $\leqslant c - c_g$; as the outstanding workload increases, accept those customers with utilities on or above $p^* + x(l^*)$; Reject all customers when the outstanding workload is above T.

Admission control is often practised in the industry. The (T, RML) policy confirms the practice and provides a rule on what is a heavy workload when it is not beneficial to accept a new order.

Remark 2. It is optimal to quote the lead time that maximizes the reward of serving the new order without considering its congestion effect. Given the lead time $(l_{r,s})$, the optimal price maximizes the overall profit by balancing the market share and profit margin.

Intuitively, pricing is often considered more important and should be determined first. In fact, one may observe from practice that price is often static, whereas lead time is dynamic according to the state of the system for a given price. However, when both the price and lead time are to be set dynamically, the policy suggests determining the optimal lead time first to maximize the reward from the accepted customer. The optimal price is then found through the global optimization (policy iteration) to take care of the future congestion effect from accepting the current order. Using lead time to balance market share and profitability, with the price being fixed first, may not be effective because any adjustment in the lead time will affect the delivery cost, which in turn may lead to a change in the optimal price. Duenyas and Hopp (1995) showed that with a static lead time, there exists an admission threshold. Theorem 1 extends this result to the case of the joint dynamic price and lead-time quotation problem. When the new customer is not rejected outright, we need to solve a single-variable semi-Markov decision problem to obtain the optimal price. With Theorem 2 below, we can use the policy iteration algorithm to compute the optimal price exactly and efficiently. Let $\text{FR}(p)$ be the failure rate function of $1 - q(p + x(l))$ in p for fixed l. We have

$$\text{FR}(p) = \frac{1}{q(p+x(l))}\frac{\partial[1 - q(p+x(l))]}{\partial p} = -\frac{1}{q(p+x(l))}\frac{\partial q(p+x(l))}{\partial p} > 0.$$

Theorem 2. *When the failure rate* $\text{FR}(p)$ *is nondecreasing in p, $\psi_{s,\mathbf{v}}(p, l)$ is unimodal in p.*

Proof. Let $\Omega(p) \equiv \text{FR}(p)V - 1$, where $V = p - c - D_s(l) + \sum_j m_{sj1} v_j + c_g$. When $q(p + x(l)) \neq 0$,

$$\frac{\partial \psi_{s,\mathbf{v}}(p, l)}{\partial p} = \frac{\partial q(p+x(l))}{\partial p} V + q(p+x(l)) = -q(p+x(l))\Omega(p). \tag{5}$$

Hence, $\partial\psi_{s,\mathbf{v}}(p, l)/\partial p > 0$, $=0$, <0 if and only if $\Omega(p) < 0$, $=0$, >0, respectively. Clearly, V is strictly increasing in p. Thus, when $V \leqslant 0$, $\Omega(p) < 0$; when $V > 0$, $\Omega(p)$ is strictly increasing in p, and hence there exists at most one root for $\Omega(p) = 0$. Therefore, $\psi_{s,\mathbf{v}}(p, l)$ is unimodal in p. $\square$

4. Summary

We constructed a discounted SMDP model to study the joint dynamic price and lead-time quotation problem with a fairly general demand function. We showed that a simple two-parameter (T, RML) policy is optimal. The two policy parameters T and RML can be determined exactly using the analytical expressions provided. Under this policy, the joint dynamic quotation problem is reduced to a sequential decision problem. Whereas the optimal lead time is RML, the optimal price can be easily computed through policy iterations based on the unimodal property of the value functions. The (T, RML) policy reveals some interesting managerial insights on how one should approach the dynamic price and lead-time decision problem: determine first whether to reject an order or not according to the number of outstanding orders; if not, set the lead time to maximize the gain from this order (i.e., RML), then set the price to maximize the long-run gain given the lead time.

The model studied here can be extended in a number of ways. We can generalize the service-time distribution to the phase-type distribution. What we need is to add a second state variable to track the phase of the current processing time. All the results remain valid. We may add an early delivery penalty. Lemma 1 and Proposition 1 still hold, but it is difficult to show the monotone property for $\mathbf{v}$. T is still a threshold, and the problem can still be reduced to a sequential SMDP. The exact optimal quotation can be obtained either explicitly or numerically. Duenyas (1995) considered joint job sequencing and lead-time quotation. His key results on sequencing can also be extended to our model.

Acknowledgments

The authors thank the associate editor, two referees, and the area editor for their constructive comments that helped improve the manuscript. This work is supported in part by Hong Kong Research Grant Council grant 6170/05E, National Science Foundation of China grant 70731003, and University of Macau grant RG031/04-05S/C88/LXM/FBA.

References

Ata, B., T. L. Olsen. 2009. Near-optimal dynamic lead-time quotation and scheduling under convex-concave customer delay costs. *Oper. Res.* **57**(3) 753–768.

Baker, K. R. 1984. Sequencing rules and due-date assignments in a job shop. *Management Sci.* **30**(9) 1093–1104.

Boyaci, T., S. Ray. 2003. Product differentiation and capacity cost integration in time and price sensitive markets. *Manufacturing Service Oper. Management* **5**(1) 18–36.

Duenyas, I. 1995. Single facility due date setting with multiple customer classes. *Management Sci.* **41**(4) 608–619.

Duenyas, I., W. J. Hopp. 1995. Quoting customer lead-times. *Management Sci.* **41**(1) 43–57.

Hopp, W. J., M. L. Spearman. 1993. Setting safety leadtimes for purchased components in assembly systems. *IIE Trans.* **25**(2) 2–11.

Li, L. 1992. The role of inventory in delivery-time competition. *Management Sci* **38**(2) 182–197.

Liu, L., M. Parlar, S. X. Zhu. 2007. Pricing and lead time decisions in decentralized supply chains. *Management Sci.* **53**(5) 713–725.

Lutze, H., Ö. Özer. 2008. Promised lead-time contracts under asymmetric information. *Oper. Res.* **56**(4) 898–915.

Palaka, K., S. Erlebacher, D. H. Kropp. 1998. Lead-time setting, capacity utilization, and pricing decisions under lead-time dependent demand. *IIE Trans.* **30**(2) 151–163.

Plambeck, E. L. 2004. Optimal lead-time differentiation via diffusion approximations. *Oper. Res.* **52**(2) 213–228.

Puterman, M. L. 1994. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., New York.

Slotnick, S. A., M. J. Sobel. 2005. Manufacturing lead-time rules: Customer retention versus tardiness costs. *Eur. J. Oper. Res.* **163**(3) 825–856.

So, K. C., J. S. Song. 1998. Price, delivery guarantee and capacity selection. *Eur. J. Oper. Res.* **111**(1) 28–49.

Wein, L. M. 1991. Due-date setting and priority sequencing in a multiclass $M/G/1$ queue. *Management Sci.* **37**(7) 834–850.

Yano, C. A. 1987. Setting planned leadtimes in serial production systems with tardiness costs. *Management Sci.* **33**(1) 95–106.