Demand and Production Management with Uniform Guaranteed Lead Time

Uday S. Rao  Jayashankar M. Swaminathan ‡  Jun Zhang †

College of Business, University of Cincinnati, Cincinnati, OH 45221 *

Email: uday.rao@uc.edu, Tel: (513)556-7138, Fax:(513)556-5499 *

Kenan-Flaglar School of Business, University of North Carolina, Chapel Hill, NC 27599 ‡

Email: msj@unc.edu, Tel: (919)843-8341, Fax: (919)962-6949 ‡

A.B. Freeman School of Business, Tulane University, New Orleans, LA 70118 †

Email: jzhang4@tulane.edu, Tel: (504)865-5075, Fax: (504)865-6751 †

Submission History:
March 2003, Revised August, December 2004, March 2005

Abstract

Recently, innovation-oriented firms have been competing along dimensions other than price - lead time being one such dimension. Increasingly, customers are favoring lead time guarantees as a means to hedge supply chain risks. For a make-to-order environment, we explicitly model the impact of a lead time guarantee on customer demands and production planning. We study how a firm can integrate demand and production decisions to optimize expected profits by quoting a uniform guaranteed maximum lead time to all customers. Our analysis highlights the
increasing importance of lead time for customers, as well as the tradeoffs in achieving a proper balance between revenue and cost drivers associated with lead-time guarantees. We show that the optimal lead time has a closed-form solution with a newsvendor-like structure. We prove comparative statics results for the change in optimal lead time with changes in capacity and cost parameters and illustrate the insights using numerical experimentation.

**Keywords:** Manufacturing Strategy, Marketing-Manufacturing Interface, Make to Order, Uniform Lead Time, Lead-Time Sensitive Demand.
1 Introduction

As firms compete more fiercely in a global economy, response time has become an increasingly important strategic weapon (Stalk and Hout 1990, Blackburn 1991, Skinner 1996a, Skinner 1996b, Hayes and Pisano 1996, Boyaci and Ray 2003, Swaminathan and Lee 2003). With the internet enabling customers to easily compare prices, experts predict that firms will have to differentiate themselves in alternative ways such as offering better quality of service, faster lead times and fully customized products and services (Shapiro and Varian 1998). One of the important changes in the production environment over the last few years has been the prevalence of manufacturing systems which allow significant product customization often through use of a build- or assemble-to-order system with little or no finished goods inventory. Lead times are key factors determining the success of such systems (Swaminathan and Tayur 1998, Swaminathan and Tayur 2003). One way that firms are trying to differentiate themselves in such an environment is by quoting a guaranteed lead time. For example, Titleist/Foot-Joy, a leading manufacturer of customized golf balls, geared its manufacturing to achieve guaranteed lead times (Casey 1997). Further, the ability to set clear customer expectations and to be able to deliver those efficiently has prompted some firms to provide uniform lead times (across customers) in addition to guaranteeing delivery times. For example, LeatherTECH, a manufacturer of leather furniture, guarantees retailers across the country a two-week delivery on any of its 32 styles of leather sofas, love seats, chairs and ottomans (Hensell 1996). Our experience with construction and telecommunications equipment providers also indicated that their recent move to a uniform guaranteed service for field service problems, instead of the traditional gold, silver and bronze service categories, was im-
plemented to decrease the gap between customer expectations and actual performance, thereby improving customer satisfaction.

A competitive strategy such as offering a uniform guaranteed lead time has to be fulfilled with careful production planning (Hayes and Pisano 1996). When a firm quotes a shorter lead time in such an environment, it is likely to attract more customers and as a result, bring in more revenues. However, a shorter lead time also imposes greater expectations and requirements on the operations of the firm as well as its suppliers. Obtaining lead time guarantee is also an effective means for the customer to hedge supply chain risks (caused by supply uncertainty, which is different from supply chain disruption risk studied by Hendricks and Singhal 2005, Kleindorfer and Saad 2005, and Sodhi 2005). Firms usually consider a priori decisions such as quoting uniform and short lead times as higher level tactical issues (affecting market demand, capacity/production planning, and cash-flows) whereas order fulfillment is considered a lower level operational issue. To further exacerbate the problem, the first set of decisions are typically made by the firm’s marketing department whereas the second set of decisions are made by the production or operations department. As articulated by Skinner (1996a), there has been a missing link between a manufacturing task (such as guaranteeing a uniform lead time) and the production system design to fit the task. As a result, it is no surprise that firms find it very difficult to coordinate demand and production decisions.

In this paper, we address the “linkage problem” (Skinner 1996a) by presenting a model that combines the above two issues related to production and demand management in a make-to-order environment with no finished goods inventory. We assume that the firm quotes a uniform maximum lead time to all customers and guarantees delivery of the product within that lead time. The price for different variants of the product is
fixed and is assumed to be exogenously set by the market (we relax this assumption in § 5). The firm has limited capacity and could potentially outsource some part of the production at a higher cost. Alternatively, unmet demand can be assumed to be lost as in the new product environment of Kumar and Swaminathan (2003). The demand rate for a given lead time is stochastic and strictly decreasing in the lead time quoted. So, a longer quoted lead time implies that a smaller number of people would want to buy the product (in expectation). We assume that the firm updates its production schedule periodically and synchronizes it with the lead time quote. For instance, if the firm quoted a uniform lead time of 2 weeks and updated its production plan every week, it would satisfy orders placed in week 1 by production in week 2. Such a periodic system permits coordinated “just-in-time” delivery of components by suppliers. Under the setting described above, the firm optimizes the quoted lead time (or equivalently, the production planning interval) to maximize expected profits per unit time. We show that the optimal profit function is quasi-concave in the lead time. We also prove analytical results on how the optimal lead time changes with changes in demand parameters, production capacity, outsourcing cost, and unit selling price. We develop a closed-form newsvendor type solution for the optimal lead time, and provide insights on the effect of production capacity. We extend our model to jointly optimize price and lead times as well as consider other forms of demand distributions.

The main contributions of our paper may be summarized as follows:

1. We provide a model that captures three elements of a make-to-order system – demand generation, in-house production, and outsourcing. We accomplish this by integrating the dependency between the uniform, guaranteed quoted lead time and
the resulting customer demand.

2. We obtain a closed-form newsvendor type solution for the optimal lead time when the firm sells variants of a single product. This enables us to provide interesting insights on the behavior of the optimal lead times with changes in problem parameters.

3. Our analysis highlights the increasing importance of lead time for customers, as well as the tradeoffs in achieving a proper balance between revenue and cost drivers associated with lead-time guarantees. Our computational study demonstrates that (i) higher capacity creates more value when the quoted lead time is smaller, (ii) to deal with high demand uncertainty, one should quote a longer lead time, (iii) a higher outsourcing cost hurts a firm’s expected profit more when its quoted lead time is short, and (iv) if a firm can update its production schedule more often, it can quote a shorter guaranteed lead time and accrue a higher expected profit.

The rest of the paper is organized as follows. In § 2, we discuss relevant literature. In § 3, we explain the model and assumptions therein. In § 4 and 5, we present a problem analysis that yields various properties. In § 6, we illustrate insights using numerical experimentation. We conclude in § 7 with a discussion of model extensions.

2 Related Literature

In the recent past, several researchers have started addressing issues related to coordinating lead time management with production planning decisions. Most of the research relates to make-to-order environments where this coordination issue is critical. For lead
time quotation, researchers have mainly focused on quoting lead times in real-time for each customer order. Weng (1996) explores integrating lead time quotation and production planning with two classes of customers: lead time sensitive and lead time insensitive. He determines the optimal (profit maximizing) lead time to quote in a queuing environment. In his model, although quoted lead time affects the manufacturing lead time thereby incurring higher production costs for lower lead times, it does not affect the arrival rate of demand. Duenyas and Hopp (1995) consider the problem of dynamically quoting lead times to customers based on the congestion in the manufacturing environment. They model market response to quoted lead times by assuming that a customer will accept a lead time with a probability which is decreasing in the lead time quoted. Using queuing and semi-Markov decision processes, they analyze several versions of this problem. Cheng and Gupta (1989) survey the related problem of assigning due-dates in job shop scheduling environments.

Our paper is closely related to the above papers in that we also consider lead time quotation as well as production in an integrated model and determine the optimal lead time to quote as well as the optimal production policy. However, we assume a uniform guaranteed lead time for all customers, allow for outsourcing of additional demand and do not consider those congestion effects on the shop-floor that result from uncertainty in production parameters. Further, we use a stochastic programming framework instead of a non-linear programming approach based on queuing approximations. Although the effect of quoting longer lead times is similar in the models, our model is more tactical (not intended to be real-time since our uniform lead time quote is made before demand materializes) whereas the models referred to above are more operational and could be utilized in real-time. Keeping some of the order level details away from our
model helps us get simpler and intuitive solutions to the problem. So and Song (1998) also consider a fixed lead time model and study the interrelationships between pricing, delivery time guarantees, demand and overall profitability. In a G/G/1 queuing setting, they model demand rate as a Cobb-Douglas function of price, with a service constraint on the delivery time. They also allow the firm to increase capacity over time. In a single product setting with constant variable unit production costs, they characterize the price, lead time and capacity expansion choices that maximize profit. Our model is similar to theirs in that we focus on the single product problem; however, we guarantee delivery within lead time with 100% reliability by allowing an outsourcing option and we consider several model extensions. Boyaci and Ray (2003) and Ray and Jewkes (2004) use a queuing model to study pricing and lead-time decisions for two substitutable services with dedicated capacity. Their mean demand depends on the prices charged and the lead time quoted. Our paper is similar to theirs because we also consider the tradeoff between capacity and the service (lead time) guarantee. However, our modeling approach is different and we study the impact of the production schedule-update frequency on the choice of the quoted lead time and on the firm’s expected profit.

Other work in this area has assumed a deterministic (typically linear) relationship between price, lead time, and demand (Hill and Khosla 1992, Palaka, Erlebacher, and Kropp 1998, and Yano and Dobson 2002). This simplifies simultaneous determination of the optimal price and lead times to maximize overall profits. Use of a customer delay cost for larger lead times (instead of reduced demand) is considered in Dewan and Mendelson (1990), in the context of a service facility, and in Mendelson and Whang (1990), using a multi-class M/M/1 queuing model. Li (1992) studies the tradeoff between inventory and capacity using a game-theoretic model in which firms compete for
customers who are homogeneous and have a utility function that is decreasing function in lead time. Li and Lee (1994) study market competition where customer preferences (and utilities) depend on quality, delivery time and price, focusing on demand while ignoring the issue of operational production planning under limited capacity. Agrawal and Seshadri (2000) develop a related newsvendor model to simultaneously determine the retailer’s purchasing-order-quantity and selling-price when demand is a function of price and the retailer is risk-averse.

3 Model Description and Formulation

We consider a make-to-order system with one product. The demand for variants of the product is received continuously by the sales department, which guarantees a maximum delivery lead time of $2T$ (we generalize this to $MT$ for any integral $M$ in §5). The concept of uniform guaranteed worst case lead time can be attractive to the manufacturing firm in at least two ways. First, it helps in clearly articulating the service to be delivered (to people within the firm) and in refining customer expectations. This is likely to help the firm monitor and control the gap between customer expectations and actual performance in a timely and accurate manner. Second, it allows the firm to utilize the guaranteed lead time as a powerful marketing punchline. Clearly, in following a strategy of quoting a uniform guaranteed lead time to all customers, the firm does lose the opportunity of extracting the maximum surplus from each customer class. However, some firms perceive that such a strategy could help them differentiate themselves in the marketplace in terms of quality and reliability of service. In our model, when customers place an order for a unit of the product, they pay the firm $p$ $(\text{unit product selling price})$. This is common
practice in direct marketing channels such as telemarketing, telephone and on-line sales, and is not a restrictive assumption in our model.

Every $T$ time periods, the sales department summarizes the accumulated customer orders and sends this information to the manufacturing department. The production department then generates a production schedule based on the aggregated orders, procures raw materials and components, assembles and ships the finished products. It is possible that the production department has instantaneous access to the customer order information database at all epochs, however the production schedule is only changed once every $T$ time periods. Such a periodic planning model reduces the amount of decision-making effort required by the production personnel. Further, once $T$ is advertised to customers and suppliers and it is known that production plans are updated every $T$ periods (say, every Saturday morning), the sales department can allow customers to make minor changes to their order until the end of the preceding $T$-interval (say, Friday evening) and the component suppliers are given a reasonably accurate look-ahead into the next period’s production plan which facilitates efficient delivery. Note that we will ignore the time value of money, since we expect the production planning interval $T$ to be small, i.e., of the order of one week.

The demand rate, given a planning interval of $T$ periods, is denoted by $\eta(T,p)$. Occasionally, for notational simplicity, $\eta(T,p)$ will be simply written as $\eta$; further in the production problem, $\eta$ will also denote the realization of demand rate. Details of the demand process (revenue inflow) and the production plan (cost control) are provided below.
3.1 Demand Process

We make the following assumptions on $\eta(T,p)$, the demand rate or the average per period demand (over $T$ periods):

A1. *Additive Demand*: $\eta(T,p) = f(T,p) + \epsilon$, where $\epsilon$ is a continuous random variable, independent of $T$ and $p$, with cumulative distribution function, $\Phi(\cdot)$, and probability density function, $\phi(\cdot)$.

Note that we assume the above for ease of notation in illustrating our basic framework. In § 7.2 we show how our results can be extended to $\eta(T,p) = a_1(\epsilon)f(T,p) + a_0(\epsilon)$, where $a_0(\cdot)$ and $a_1(\cdot) \geq 0$ are functions representing respectively the additive and multiplicative components of demand.

Assumption A1 says that the demand rate is uncertain and comprised of two components: the first component, $f(T,p)$, is deterministic and depends on the quoted lead time and the product price, whereas the second component is stochastic and independent of $T$ and $p$. For instance, the mean demand rate $\eta$ may be affected by (1) lead time (based on the time-sensitive customer segment), (2) price, and (3) product quality, marketing effort, or some other physical attribute (based on the time-insensitive customer segment). Admittedly, assuming deterministic $f(T,p)$ is restrictive, however our demand model is still more general than many of those previously considered and is itself analytically challenging. The stochastic part, $\epsilon$, may be viewed as the error in use of a deterministic model of demand rate. Further, because $\epsilon$ represents the combined effect of several unobserved factors affecting demand, it is natural to assume that $\epsilon$ is normally distributed. Although we do not require any distributional assumption on $\epsilon$, a
normal distribution is likely to simplify estimation of $f(T,p)$ from historical or survey sampling data.

A2. Dependence on $T$: $f'_T(T,p) < 0$, that is, $f(T,p)$ is strictly decreasing with respect to $T$ for given $p$ and differentiable over $T \in [T^{lb}, T^{ub}]$.

Because customer orders over $T$ time periods are collected by sales and the corresponding products are then assembled and shipped within the next $T$ periods, customer lead time can be no more than $2T$. The general case with a lead time guarantee of $MT$ can also be handled, see § 5. The longer the $T$, the longer will be the delivery lead time observed by the customers. Once the firm quotes a lead time guarantee of $2T$, those time-sensitive customers who find this guarantee inadequate will not place orders, larger $T$ will correspond to more customers not placing orders. Consequently, the firm’s demand will decrease in $T$, which we model by assuming that $f(T,p)$ is decreasing in $T$. This is consistent with the customer utility model of demand commonly used in marketing and economics (Li and Lee 1994). Note that by assumption A2, $f(T,p)$ is strictly decreasing in $T$. We assume this to simplify our analysis. However, this assumption is not restrictive since every decreasing instance of $f(T,p)$ over $[T^{lb}, T^{ub}]$ may be solved by separately considering each strictly decreasing portion of $f(T,p)$ and picking the best solution identified from the different intervals. The lower limit $T^{lb}$ may be determined by information processing limitations (as discussed in Lovejoy and Whang 1995) or by production planners who want to avoid “system nervousness” and the additional effort resulting from too frequent production planning. The upper bound $T^{ub}$ may be set by marketing’s sales targets or other competitive factors.
3.2 Production Process

Every $T$ time periods, the production department must make operational decisions based on the demand accumulated during the previous $T$ time periods. Products can be made either in-house (within the plant) at a per unit production cost of $c$ or by outsourcing at a per unit cost of $c^o$. Given the selling price vector, $p$, we assume that $c \leq c^o$ and $c \leq p$. When $c^o \leq p$, no matter which production mode is used (in-house or outsourcing), it is always profitable to satisfy as much demand as possible. Consequently, the firm will choose the $T$ to maximize expected demand rate, $\eta(T, p)$, that is, the optimal $T$ is $T^{lb}$, the minimum possible $T$. Hence, to exclude such trivial cases, we assume $p \leq c^o$. In order to keep the model simple, the in-house production cost as well as the outsourcing cost have been assumed to be constant. It is reasonable to assume that these costs will be constant within a range of lead times. In cases where the costs vary a lot over the possible values of lead time, one could solve our problem for several smaller intervals of lead time where these costs are constant and pick the best solution. The in-house production rate$^1$ is $\rho$ and there is no restriction on the volume of outsourcing. The uncapacitated outsourcing assumption may be justified by the fact that $c^o > p$, thus the subcontractor who satisfies the outsourced demand is being generously compensated for their delivery speed (at a cost greater than the market selling price). Alternatively, the additional cost incurred by outsourcing, $c^o - c$, may be viewed as a penalty for violating the delivery guarantee. This is consistent with similar guarantees by firms such as Dominos pizza and Caterpillar parts logistics which have sometimes advertised “guaranteed delivery within the quoted lead time or the part is free.”

$^1$Although we treat $\rho$ as given, our model may be used to select $\rho$ when capacity costs can be modeled as a well-behaved function, e.g., $C(\rho) = c_\rho \rho$. 

13
Table 1: Table of Notation

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>Unit in-house production cost</td>
</tr>
<tr>
<td>$c^o$</td>
<td>Unit outsourcing cost</td>
</tr>
<tr>
<td>$p$</td>
<td>Unit selling price</td>
</tr>
<tr>
<td>$\rho$</td>
<td>In-house production rate</td>
</tr>
<tr>
<td>$\xi(T)$</td>
<td>Demand during $T$ time periods</td>
</tr>
<tr>
<td>$f(T, p)$</td>
<td>Deterministic demand rate (due to lead time quote $2T$)</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>Random demand rate (due to unobserved factors)</td>
</tr>
<tr>
<td>$\Phi(\cdot), \phi(\cdot)$</td>
<td>CDF, pdf of $\epsilon$</td>
</tr>
</tbody>
</table>

The notation used throughout this paper is summarized in Table 1. Note that we will sometimes use $\xi(T)/T$ instead of $\eta(T, p)$, where $\xi(T)$ is the random realized demand over $T$ periods, given $p$. We suppress the dependence of demand on $p$ whenever the selling price is assumed to be exogenously specified. The total production cost over $T$ time periods will depend on the production process (in-house vs. outsourcing) and the magnitude of realized demand during the previous $T$ time periods. Given that the production schedule is updated two times within the quoted lead time, we see that the production problem may be easily solved by satisfying as much demand as possible using in-house production. When in-house capacity is insufficient use outsourcing. We denote the optimal production cost over time $T$, for demand $\xi$, by $Q(T, \xi)$. Then $Q(T, \xi) = c\xi$ when $\xi \leq \rho T$, $Q(T, \xi) = c\rho T + c^o(\xi - \rho T)^+$ when $\xi \geq \rho T$. Therefore $Q(T, \xi)$ may be written as,

$$Q(T, \xi) = c\xi + (c^o - c)(\xi - \rho T)^+.$$ (1)

Note that production decisions are made after demand materializes, whereas $T$ is chosen
under uncertainty in demand. The production cost does not include the cost for work-in-process (WIP) inventory, since in our make-to-order environment, we assume that all components are paid for up-front (beginning of the $T$-interval) and all products are paid for by the customer at the time the order is placed. Further, products are never stored since they are shipped as soon as completed, possibly before the guaranteed lead time of $2T$. Thus we obviate the combinatorial problem of sequencing the products in real-time, which simplifies analysis of our tactical model.

To efficiently manage the system, the firm must determine an optimal order processing cycle $T$ and an optimal price level $p$. To facilitate our analysis, we first characterize the optimal $T$ for any given value of $p$, which we call the $T$ problem. Every $T$ time periods the company builds a production schedule in response to demand and incurs costs $Q(T, \xi)$. The overall goal of the $T$ problem is to maximize the long-run average expected profit per period. In this paper we restrict attention to the $T$ problem with fixed price $p$; we briefly discuss joint optimization of $T$ and $p$ in § 7.1.

In each interval of $T$ time periods, the firm’s expected profit is $E[p\xi - Q(T, \xi)]$. Thus, the $T$ problem may be formulated as:

$$
T \text{ Problem:} \quad \max_{T \in [T^b, T^u]} \left( J(T) = E_{\xi} \left[ J(T, \xi(T)) \equiv \left\{ p\frac{\xi}{T} - \frac{Q(T, \xi)}{T} \right\} \right] \right). \quad (2)
$$

For notational simplicity, we define $g = f(T) \equiv f(T, p)$ and re-write $J(T)$ as follows:

$$
\mathcal{J}(g) = E_{\epsilon}[p(g + \epsilon) - q(g, \epsilon)], \quad (3)
$$

where

$$
q = \frac{Q(T, \xi)}{T} = c\xi/T + (c^p - c)(\xi/T - \rho)^+ = c(g + \epsilon) + (c^p - c)(g + \epsilon - \rho)^+. \quad (4)
$$
4 Analytical Results for $M = 2$

In this section, we show that expected profit $J(T)$ is quasi-concave (see Bazaraa, Sherali, and Shetty (1993) for a definition of quasi-concavity), develop a solution approach to compute $T^*$ (the optimal $T$), and present comparative statics on how $T^*$ changes with different parameters. We begin our analysis by exploring the shape of $J(T)$ and how profit changes with capacity, $\rho$.

**Proposition 1** $J(T)$ is quasi-concave in $T$ and the optimal value of $T$ satisfies $T^* = \min\{T : \frac{dJ}{dT} \leq 0\}$.

**Proof.** By definition, $J(T) = J(g)$. From (4), it is easy to see that $q(g, \epsilon)$ is convex in $g$ for each $\epsilon$. Hence, using (3), we see that $J(g) \equiv E_\epsilon[p(g + \epsilon) - q(g, \epsilon)]$ is concave in $g$. Now the required result follows from the fact that $g = f(T)$ is decreasing in $T$ and that any concave function of a monotone function is quasi-concave. □

**Proposition 2** The profit function $J(T, \rho)$ is jointly concave in $(g, \rho)$, where $g = f(T)$; $J(T, \rho)$ is increasing in $\rho$.

**Proof.** Clearly, $J(T, \rho)$ is increasing in $\rho$ for each $g$. The concavity result follows directly from the fact that the average cost function $\frac{Q\xi(T)}{T} \equiv q(g, \epsilon)$ is jointly convex in $(g, \rho)$ for each realization of $\epsilon$. □

**Corollary 3** $J(T^*(\rho), \rho)$, the optimal profit function corresponding to $T^*(\rho)$, is increasing and concave in $\rho$.

**Proof.** It is well known (Rockafellar 1970) that maximizing a jointly concave function over a closed convex subset of the variables maintains concavity in the remaining variables. By Proposition 2, $J(T, \rho) = J(g, \rho)$ is jointly concave in $(g, \rho)$ and, by Assumption
A2, the feasible region for \( g \) is a closed interval. Hence, \( \max_g J(g, \rho) = \max_T J(T, \rho) \) is concave in \( \rho \).

By the above corollary, we conclude that, although increasing \( \rho \) will increase the optimal profit, the marginal growth in profit is decreasing in \( \rho \). This is an interesting observation from a managerial perspective particularly while making decisions regarding capacity expansion or negotiating outsourcing costs.

### 4.1 The Newsboy Formula for the Optimal \( T \)

From (3), we have \( J(g) = E_e[(p - c)(g + \epsilon) - (c^o - c)(g + \epsilon - \rho^+)] \). Determining the optimal \( T \) is equivalent to finding the \( g^* \) that maximizes \( J(g) \). Since the form of \( J(g) \) is identical to a newsvendor problem with underage cost \( (c^o - p) \), overage cost \( (p - c) \), and “demand” \( \epsilon \), we know that the optimal solution satisfies

\[
\text{Prob}\left\{ \frac{\xi(T)}{T} \leq \rho \right\} = \frac{c^o - p}{c^o - c} \quad \text{or} \quad \rho - g^* = \Phi^{-1}\left( \frac{c^o - p}{c^o - c} \right).
\]

This \( g^* \) yields \( T^* \) as follows: Let \( g^{ub} = f(T^{lb}) \) and \( g^{lb} = f(T^{ub}) \). If \( g^* < g^{lb} \), set \( T^* = T^{lb} \).

If \( g^* > g^{ub} \), then \( T^* = T^{lb} \), else \( T^* = f^{-1}(g^*) \). Henceforth, we will focus on the case where \( g^* \) does not take on a boundary value (i.e., \( g^* \) is strictly between \( g^{lb} \) and \( g^{ub} \)).

**An Economic Interpretation of the Optimality Condition for \( T^* \):** When \( T^* = f^{-1}(g^*) \), we note that smaller \( T \) corresponds to higher demand rate and lower in-house production capacity within the planning interval, while larger \( T \) results in lower demand rate with higher in-house capacity. Intuitively, we would like to choose a \( T \) value that matches demand to production capacity \( \rho \), while taking into consideration the tradeoff between the unit loss due to outsourcing, \( (c^o - p) \), and the unit profit from in-house production, \( (p - c) \).

For any choice of \( T \), loss due to outsourcing occurs with probability \( \text{Prob}\{\frac{\xi(T)}{T} \geq \rho\} \), and
profit from in-house production occurs with probability \( \Pr(\frac{\xi(T)}{T} \leq \rho) \). The marginal expected outsourcing loss is \((c^o - p)\Pr(\frac{\xi(T)}{T} \geq \rho)\), which decreases as \( T \) increases, and the marginal expected profit from in-house production is \((p - c)\Pr(\frac{\xi(T)}{T} \leq \rho)\), which increases as \( T \) increases. Clearly, at the optimal \( T \), the marginal expected loss must equal the marginal expected profit or \((c^o - p)\Pr(\frac{\xi(T)}{T} \geq \rho) = (p - c)\Pr(\frac{\xi(T)}{T} \leq \rho)\), which yields our newsvendor-like optimality equation.

Equation (5) reflects a fundamental tradeoff in managing a make-to-order system, that is, the tradeoff between cost increases \((c^o - p)\) when demand is high and the lost potential profit \((p - c)\) when demand is low. This is similar to the newsvendor model in a make-to-stock environment, where order quantities are determined before demand materializes and excess demand incurs shortage costs while low demand results in inventory holding / salvage costs. Whereas in the traditional newsvendor we set order quantities given demand, in the make-to-order environment, given capacity, we set lead time to capture the desired amount of demand. For the single product case, equation (5) provides an efficient, closed-form expression for computing the optimal \( T \). The above result is similar in spirit to the inverse newsvendor result obtained by Carr and Lovejoy (2000) where they decide which demands to accept (thereby determining the demand distribution) for a capacitated firm. In our model, the changes in parameters of the demand distribution occur due to different quoted lead times.

### 4.2 Comparative Statics

In this section, we develop sensitivity results to changes in problem parameters. We use \( T(\alpha) \), for \( \alpha = \epsilon, \rho, p, c \) or \( c^o \), to make explicit the dependence of \( T \) on parameter \( \alpha \);
$T^*(\alpha)$ denotes the corresponding optimal $T(\alpha)$, assuming all parameters other than $\alpha$ remain unchanged. As is common in stochastic comparisons (Stoyan 1983), we say that a random variable $X$ is stochastically smaller than $Y$, denoted by $X \leq_{st} Y$, if $\text{Prob}\{X \leq a\} \geq \text{Prob}\{Y \leq a\}$, for all $a$. Proposition 4, proven in Appendix A, shows that, as the uncertainty in demand rate increases, the corresponding optimal $T^*$ increases.

**Proposition 4** Let $\xi^k_k = f(T) + \epsilon^k$, for $k = 1, 2$. If $\epsilon^1 \leq_{st} \epsilon^2$, then $T^*(\xi^1) \leq T^*(\xi^2)$.

Proposition 5, below, demonstrates that when demand variability decreases, a firm’s expected profit does not decrease if its quoted lead time is kept fixed or adjusted optimally. We conduct these demand variability comparisons using the stochastic convex order comparison of demand noise ($\epsilon$). By definition, a random variable $X$ is smaller than $Y$ in convex order, denoted by $X \leq_{cx} Y$, if for all convex functions $h(\cdot)$ we have $E[h(X)] \leq E[h(Y)]$. It is well known that $X \leq_{cx} Y$ implies $E[X] = E[Y]$ and $\text{Var}[X] \leq \text{Var}[Y]$ (see, for example, Stoyan 1983).

**Proposition 5** Let $\xi^k/T = f(T) + \epsilon^k$, for $k = 1, 2$. If $\epsilon^1 \leq_{cx} \epsilon^2$, then $\mathcal{J}^1 \geq \mathcal{J}^2$ when $T$ is either held fixed or adjusted optimally, where $\mathcal{J}^k$ is the expected profit corresponding to $\epsilon^k$ for $k = 1, 2$.

**Proof.** We only show that for any given $T$, $\epsilon^1 \leq_{cx} \epsilon^2$ implies $\mathcal{J}^1 \geq \mathcal{J}^2$. This follows from the definition of stochastic convexity and the fact that $\mathcal{J} = E_c[(p - c)(g + \epsilon) - (c^0 - c)(g + \epsilon - \rho)^+]$, with $(p - c)(g + \epsilon) - (c^0 - c)(g + \epsilon - \rho)^+$ being concave in $\epsilon$.

**Proposition 6** Other parameters remaining unchanged,

1. If $\rho^1 \leq \rho^2$, $T^*(\rho^1) \geq T^*(\rho^2)$;
2. If \( p^1 \leq p^2 \), \( T^*(p^1) \geq T^*(p^2) \);
3. If \( c^{o1} \leq c^{o2} \), \( T^*(c^{o1}) \leq T^*(c^{o2}) \).
4. If \( c^1 \leq c^2 \), \( T^*(c^1) \leq T^*(c^2) \).

\[ T^*(\rho_1) \geq T^*(\rho_2). \]

**Proof.** To prove the first result, we only need to show that when \( \rho^1 \leq \rho^2 \), \( \frac{dJ}{dT}|_{\rho=\rho^1} \geq \frac{dJ}{dT}|_{\rho=\rho^2} \). Note that

\[
\frac{dJ}{dT} = \frac{dJ}{dg} \frac{dg}{dT} = \left[ (p - c) - (c^o - c)(1 - \Phi(\rho - g)) \right] f'_T(T).
\]

Since \( f'_T(T) < 0 \), we have \( \frac{dJ}{dT} \) is decreasing in \( \rho \), which proves the first result. The remaining results follow analogously. 

**Illustration of Managerial Insights from Comparative Statics:** Let \( J(T, \rho^k) \) denote the profit function for \( \rho = \rho^k, k = 1, 2 \). Let \( \rho^1 \leq \rho^2 \). By Proposition 2, we know that \( J(T, \rho^1) \leq J(T, \rho^2) \). From the proof of Proposition 6, we see that

\[
\frac{\partial J}{\partial T}|_{\rho=\rho^1} \geq \frac{\partial J}{\partial T}|_{\rho=\rho^2}.
\]

To present the implications of the above, in Figure 1 we have illustrated \( J(T, \rho^k) \), for \( k = 1, 2 \), taking care to ensure that the slopes satisfy the relation in expression (6), with \( T^{*k} \equiv T^*(\rho^k) \) denoting the profit maximizing \( T \) corresponding to \( \rho^k \). Consistent with the result in Proposition 6, we see that \( T^{*2} \leq T^{*1} \). The graph can be broken up into three regions: (i) \( T \leq T^{*2} \), (ii) \( T^{*2} \leq T \leq T^{*1} \), and (iii) \( T \geq T^{*1} \).

Consider region (i). In this case, the slope of both curves are non-negative, and by (6), the slope is higher for \( \rho^1 \). This implies that, as \( T \) decreases, \( J(T) \) decreases faster for lower \( \rho \). Consequently, a firm with lower \( \rho \) has a smaller incentive to reduce \( T \) below \( T^{*2} \) compared to a firm with higher \( \rho \). Similarly, in region (iii), the firm with higher \( \rho \) has a
smaller incentive to increase $T$ beyond $T^{*1}$ (as compared to a firm with smaller $\rho$), since its profit will be more suboptimal and will decrease at a faster rate. A similar argument applies over region (ii) where we see that the firm with smaller $\rho$ has a greater incentive to increase $T$ above $T^{*2}$. In summary, we see that higher capacity firms are less adversely affected by lower-than-optimal lead times and lower capacity firms are less affected by higher-than-optimal lead times. This could have an important bearing on how firms may choose to react in a dynamic, competitive environment. An analogous reasoning applies to a single firm which is contemplating capacity increases. Effectively, capacity increases favor reductions in $T^{*}$ (and in the optimal quoted lead time). Further, our model may be used within a what-if analyses framework to determine the profit maximizing value of $T$ and $\rho$. For instance, when the cost of capacity is linear (say $c\rho$), since $J(T, \rho)$ is increasing, concave in $\rho$, the optimal $\rho$ corresponds the $\rho$ value at which $\frac{\partial J(T, \rho)}{\partial \rho} = c\rho$. 

Figure 1: Effects of different capacity
From the proof of Proposition 6, we can derive the following relations which are similar to (6):

\[ \frac{\partial J}{\partial T} \bigg|_{p=p^1} \geq \frac{\partial J}{\partial T} \bigg|_{p=p^2} \quad \text{if } p^1 \leq p^2, \]

\[ \frac{\partial J}{\partial T} \bigg|_{c^o=c^o_1} \leq \frac{\partial J}{\partial T} \bigg|_{c^o=c^o_2} \quad \text{if } c^o_1 \leq c^o_2, \]

These may be interpreted in a manner similar to our analysis on the effects of capacity.

To summarize, we see that when a company can charge a premium price on its product (or when its outsourcing cost is lower), it can afford to compete on a shorter $T$. Such a company has a greater incentive to reduce its $T$, compared to a company with a lower selling price or higher production cost (for the same product variants).

5 Quoted Lead Time $= MT$, $M > 2$

So far, we have studied one possible way of coordinating lead time quotation with production management: Within a uniformly quoted maximum lead time, the production schedule is updated twice, once at the beginning of the lead time and once in the middle of the lead time. However, it is possible that within the quoted lead time, the production schedule is updated more than twice, say $M$ times. For example, a company may quote a four week lead time and update the production schedule once a week.

5.1 Production Problem when Lead Time $= MT$

Given that the production schedule is updated $M$ times within the quoted lead time, we see that the production problem now faces two types of demand, depending on when the demand is received. We call an order a high-priority demand if it was received $M - 1$ production runs ago, since if there is such an order in the system, it must be satisfied by
production or outsourcing in the current production run (interval) else the guaranteed lead time can not be satisfied. All other orders are called low-priority demand, since they may be backlogged into future production intervals. For this problem environment, the production problem can be solved in a greedy manner\(^2\). That is, first satisfy the high-priority demand using either in-house production or outsourcing, if the in-house production capacity is insufficient. If extra in-house production capacity is available after high-priority demand is satisfied, the residual capacity is used to satisfy low-priority demand. Note that no low-priority demand is satisfied by outsourcing in this production interval since, in the absence of capacity limits on outsourcing, it is always possible to postpone outsourcing to a future interval without increasing costs.

### 5.2 T Problem when Lead Time = MT

Note that the extension of our results to the T problem is non-trivial since the steady state distribution of high-priority demand in each production interval is not easy to characterize. However, by defining certain new variables, we can still show that the profit function is quasi-concave in T.

Let \(Y_t^i\) be the residual demand received in period \(t-i\) for \(i = 1, \ldots, M-1\). Clearly at the beginning of period \(t\), the residual demand can be represented by \(Y_t = (Y_t^{M-1}, \ldots, Y_t^1)\), where \(Y_t^{M-1}\) represents high-priority residual demand and \(Y_t^1 = \xi_{t-1}\) represents demand in period \(t-1\).

As stated previously, the production planning problem can be solved optimally by producing as much as possible to satisfy \(Y_t^{M-1}\). If there is insufficient capacity (that is

\[^2\text{The proof of this result has been omitted and can be obtained from the authors.}\]
\( \rho T \leq Y_t^{M-1} \), use outsourcing. If there is extra capacity, produce to satisfy other residual demands, starting with \( Y_t^{M-2} \). Hence, residual demands at the beginning of period \( t+1 \), \( Y_{t+1} = (Y_{t+1}^{M-1}, \ldots, Y_{t+1}^1) \), satisfy:

\[
\begin{align*}
Y_{t+1}^{M-1} &= (Y_t^{M-2} - (\rho T - Y_t^{M-1})^+) \\
\vdots \\
Y_{t+1}^2 &= (Y_t^1 - (\rho T - Y_t^2 - \cdots - Y_t^{M-1})^+) \\
Y_{t+1}^1 &= \xi_t.
\end{align*}
\]

From the above recursive relations, we see that \( Y_{t+1} \) may be written as a function of \( Y_t \), that is, \( Y_{t+1} = \psi(Y_t, \xi_t) \) where \( \psi(Y_t, \xi_t) \) is increasing and continuous in \( Y_t \) for any \( \xi_t \). Without loss of generality, we assume that the initial residual demand \( Y_1 \) is a zero vector. Using the method of Loynes (1962), we can show that as \( t \) goes to infinity, \( Y_t \) converges to \( Y \) in distribution, where \( Y \) satisfies \( Y = \psi(Y, \xi) \). Since \( Y_t^i \leq_{st} \xi_{t-i} \), we may assume that \( Y \) is indeed an “honest” distribution (Loynes 1962). From now on, we will replace \( Y_t \) with the steady state residual demand \( Y = (Y^{M-1}, \ldots, Y^1) \). The following result characterizes the dependence of \( Y \) on \( g = f(T) \).

**Theorem 7** For each realization of \( \epsilon \), \( Y/T \) is increasing and convex (ICX) in \( g \).

**Proof.** We prove the result using induction by showing that \( Y_t/T \) is increasing and convex in \( g \). Clearly, the latter is true for \( t = 1 \).

**Induction Hypothesis:** Let \( Y_t/T \) be ICX for \( t = 1, \ldots, \tau \).

For each realization of \( \epsilon \), \( Y_{\tau+1} \) is defined by recursions (7). Consider the equation for \( Y_{\tau+1}^{M-1} \). There are two possible cases, based on the realization of \( \epsilon \). Either (i) \( \rho T \geq Y_{\tau}^{M-1} \)
or (ii) $\rho T \leq Y_r^{M-1}$. In case (i), $Y_{r+1}^{M-1}/T = (Y_r^{M-2}/T - \rho + Y_r^{M-1}/T)^+$, which is ICX by the Induction Hypothesis; in case (ii) $Y_{r+1}^{M-1}/T = (Y_r^{M-2}/T)^+$ which is also seen to be ICX. Similarly, we see that, the Induction Hypothesis implies that $Y_{r+1}^{i}/T$ is ICX for all $i$, which proves the required result.

**Theorem 8** For lead time $MT$ with integral $M$, the expected profit $J(T)$ is quasi-concave in $T$.

**Proof.** Given the above solution to the production problem, the steady state cost in each period may be written as:

$$Q(Y,T) = \begin{cases} 
  c\rho T + c^0(Y^{M-1} - \rho T) & \text{if } Y^{M-1} \geq \rho T, \\
  cY^{M-1} + c \min\{\rho T - Y^{M-1}, Y^{M-2} + \cdots + Y^1\} & \text{otherwise.}
\end{cases}$$

Thus, $Q(Y,T) = c\rho T - c(\rho T - Y^{M-1} - \cdots - Y^1)^+ + c^0(Y^{M-1} - \rho T)^+$. Since for all $i = 1, \ldots, M-1$, $Y^i/T$ is convex and increasing in $g = f(T)$, we see that, for each realization of $\epsilon$, $Q(Y,T)/T$ is convex in $g = f(T)$. Consequently, the expected profit $J(g)$ is concave in $g$. Thus, the profit function $J(T)$ is quasi-concave in $T$.

Obtaining a closed-form newsvendor type expression for $T^*$, the optimal $T$, is complicated by the fact that the steady state distribution of the residual demand $Y$ needs to be determined. In this case, we advocate that $T^*$ be computed numerically. That is, given quasi-concavity of $J(T)$, the optimal $T$ may be found using a one-dimensional golden section or Fibonacci search over $[T^{lb}, T^{ub}]$, in which $J(T)$ is estimated for each $T$ using simulation. Even though there will no longer be a simple closed-form expression for $T^*$, we can show that the comparative statics results from the $2T$ case may be adapted to the general $MT$ case. A more careful study of the $MT$ model, including optimization of $M$, is beyond the scope of this paper and would be a topic for future research.
6 Numerical Results

The main purpose of this computational study is to illustrate the analytical results developed earlier and to examine the impact of the production schedule-update frequency (the value of $M$). For simplicity, we only consider the case in which the demand rate is linear with respect to the guaranteed lead time, $MT$. Specifically, we assume $f(T) = a - bMT$, where $a$ is the maximum potential demand rate for the product (that is, the demand rate when the guaranteed lead time is zero) and $b$ reflects the sensitivity of demand to quoted lead time $MT$. We assume the demand noise $\epsilon$ is normal with a mean of zero and standard deviation of $\sigma$.

Our numerical experiments start with a base-case scenario. Subsequently, additional scenarios are generated by varying one of the problem parameters. For each test instance, we evaluated the expected profit by varying $T$ from 1 to 100 in steps of 1. As discussed in § 5.2, there is no simple closed-form expression for $T^*$ and the expected profit when $M > 2$. Hence, we use Monte Carlo methods, implemented in C++, to evaluate the expected profit of a problem instance for each $T$ value. The demand noise realizations were generated using the popular method of normal deviates by G.E.P. Box, M.E. Muller, and G. Marsaglai (see Knuth 1981, p. 117). We simulated the system for several thousand periods (151,000 in the reported experiments) with the first 1000 periods serving as "warm-up" periods. We used the batch means method to estimate performance measures (Bratley, Fox, and Schrage 1987), each batch comprised of 5,000 periods and 30 batches were used. For simplicity, we only report average performance values, e.g., expected profit, from the simulation. All standard deviations of the average simulation estimates were small, with a maximum standard deviation of one percent of the mean demand.
The base-case scenario parameters were chosen as follows: $a = 100$, $b = 1$, $c = 1$, $p = 3$, $c^o = 6$, yielding a critical ratio $\alpha = (c^o - p)/(c^o - c) = 0.6$; $\sigma = 10$; $\rho = 80$; $M = 2$.

For this data, the optimal solution is $T^* = (a - \rho + z\sigma)/(2b) = (100 - 80 + 0.253 \times 10)/2 = 11.3$, corresponding to an optimal expected profit of 140.5.

Impact of capacity (Figure 2). We study the impact of capacity on the choice of lead time and the firm’s expected profit by varying $\rho$ from 50 to 100 in steps of 10. From Figure 2, we see that consistent with Proposition 2 and Corollary 6, as the capacity increases from 50 to 100, the optimal integral $T$ decreases from 26 to 1 and the maximum expected profit increases from 80.6 to 180.6. Further, capacity creates more value for the firm when the quoted lead time is small. In fact, when $T$ is higher than 35, additional capacity does not create additional value.

Impact of demand variability (Figure 3). We chose $\sigma \in \{5, 10, 20, 30\}$ to study the impact of demand variability on the choice of lead time and the expected profit. As expected, when demand variability is smaller, the expected profit is higher and the
For different values of $N$, we see that if a firm updates its production schedule more often, the expected profit is less sensitive to the changes in $T$.

For the base-case, Figure 5 illustrates the firm's expected profit as a function of $N$. The expected profit by varying guaranteed lead time from one to 100 in steps of 1, $c = 1$, $p = 3$, $\sigma = 5$. For each value of the schedule-update frequency, $N$, we calculated $\alpha = (c - \sigma)/(d - \sigma)$ varied from 0.2 to 0.8 in steps of 0.2. From Figure 4, we see that the critical ratio $\alpha$ is similar to the impact of capacity. That is, higher $\alpha$ corresponds to higher profit.

Impact of $M$. Starting with the base-case scenario, we choose different values for $M$, e.g., $M \in \{2, 3, 4\}$. For each value of the schedule-update frequency, we calculated $\alpha = (c - \sigma)/(d - \sigma)$ varied from 0.2 to 0.8 in steps of 0.2. From Figure 4, we see that the critical ratio $\alpha$ is less sensitive to the changes in $T$. The optimal lead time is larger. Interestingly, when demand variability is higher, the expected profit is less sensitive to the changes in $T$. Figure 3: Impact of demand variability.
remained true for each of the three scenarios. Updating the production schedule more often, it can quote a shorter guaranteed lead time and accrue a higher expected profit. For example, when \( M \) changes from 2 to 3, the optimal guaranteed lead time decreases from 129 to 127. However, the marginal benefit of higher \( M \) decreases. From the figure, we see that when \( M \) is higher than 3, more frequent updates of the production schedule have a very small incremental impact on the optimal lead time and a firm’s expected profit. Thus the marginal benefit of more frequent schedule updates was decreasing in \( M \). Note that we also tested values of \( M \in \{6, \ldots, 10\} \) and observed similar results.

In the base-case, to capture high profits, the firm only needed to update its production schedule twice within the quoted lead time. Because we found this low update frequency result interesting and somewhat surprising, we tested its robustness under different problem parameters to identify when this observation would not hold true. In particular, we tested high demand variability: \( \sigma = 0.4 \) (the other parameters remain the same as the base-case scenario); lower capacity: \( \rho = 0.6 \) (the other parameters remain the same as the base-case scenario); and high outsourcing cost: \( c_o = 11 \) (the other parameters remain the same as the base-case scenario). The low update frequency result remained true for each of these scenarios.

Figure 4: Impact of critical ratio
setting should focus on expanding its capacity, reducing its outsourcing cost, or reducing in complex real-time scheduling systems, a firm operating in a guaranteed lead time does not add significant additional burden on production planning. Instead, of investing positive expected profits. Advocating a guaranteed lead time to attract more customers or our preliminary experiments suggest that for most reasonable cases with deterministic effects of the \( \omega, T \) parameters, and permits positive expected profit. The expected values of \( M > 3 \) created additional value. For this case with high demand variability, low capacity, and large outsourcing cost, Figure 6 illustrates the expected profit as a function of \( M \). In this figure, low values of \( M \) correspond to more negative expected profit (the curve for \( M = 2 \) has too little capacity and requires too much expensive outsourcing to be profitable, especially since the demand variability is high). Recall that for a fixed value of the lead time guarantee, \( M = \frac{L}{T} \), we have \( M = \frac{L}{T} \). As \( M \) increases, the production planning horizon decreases, so we collect demand over a smaller time interval and have a longer residual time interval, \( (N - 1)T \), to satisfy this demand. This reduces the detrimental effects of the \( \omega, T \) parameters, and permits positive expected profit. In summary, our preliminary experiments suggest that for most reasonable cases with

![Figure 5: Base-case Impact of M](image-url)
7 Model Extension and Variant

7.1 Extension: Joint Determination of Price and Lead Time

In our model, we assumed that the price is determined by the market and therefore cannot be controlled by the manufacturer. However, under certain circumstances, both price and lead time may be important decision variables. In such cases, common demand models for \( f(p, T) \) include the linear function \( f(p, T) = a - bp - q(T) \) and the Cobb-Douglas function \( f(p, T) = p - b_T T - q(T) \) (Palaka, Erlebacher, and Kropp 1998) and \( f(p, T) = p - b_T T - q_T - q(T) \) (So and Song 1998).

As shown below, in our model, joint determination of price and lead time is possible regardless of the functional form of \( f \).

**Proposition 9** When \( \ell \) is chosen as the optimal \( \ell^* \), the profit function is convex in \( p \).

![Figure 6: Worst-case Impact of \( M \)](image-url)
**Proof.** For any given $p$, the optimal $T^*$ satisfies $\rho - g^* = \Phi^{-1}(\gamma)$, where $\gamma = \frac{\rho - p}{\rho - c}$.

Therefore, the profit function may be written as

\[
J(T^*, p) = E\{(p - c)(g^* + \epsilon) - (c^o - c)(g^* + \epsilon - \rho)^+\} \\
= E\{(p - c)(\rho - \Phi^{-1}(\gamma) + \epsilon) - (c^o - c)[\epsilon - \Phi^{-1}(\gamma)]^+\} \\
= (p - c)(\rho + E\epsilon) - (p - c)\Phi^{-1}(\gamma) - (c^o - c)\int_{\Phi^{-1}(\gamma)}^{\infty} (\epsilon - \Phi^{-1}(\gamma))\phi(\epsilon)d\epsilon.
\]

Now take the first order derive of $J(T^*, p)$ with respect to $p$, we have

\[
\frac{dJ}{dp} = \rho + E\epsilon - \Phi^{-1}(\gamma) + \frac{p - c}{c^o - c}\Phi_p^{-1}(\gamma) - \int_{\Phi^{-1}(\gamma)}^{\infty} \Phi_p^{-1}(\gamma)\phi(\epsilon)d\epsilon \\
= \rho + E\epsilon - \Phi^{-1}(\gamma) + [1 - \gamma]\Phi_p^{-1}(\gamma) - \Phi_p^{-1}(\gamma)[1 - \Phi(\Phi^{-1}(\gamma))] \\
= p + E\epsilon - \Phi^{-1}(\gamma).
\]

Clearly $\frac{dJ}{dp}$ is monotone increasing in $p$, therefore when the optimal $T$ is chosen the profit is convex in $p$. 

By Proposition 9, the optimal $p$ would correspond to an extreme point of the feasible set of values for $p$.

### 7.2 Variant: Additive + Multiplicative Demand Model

So far in this paper, we assumed that the effects of guaranteed lead time and the effects of other independent unobserved factors on the average demand per period were additive, i.e., $\xi(T) = f(T) + \epsilon$. However, all our results may be extended to the case where average demand is in a more general form with $\eta(T) = \frac{\xi(T)}{T}$ expressed as $\eta(T) = a_1(\epsilon)f(T) + a_0(\epsilon)$, where $f(T)$ is decreasing over $T \in [T^{lb}, T^{ub}]$ and $\epsilon$ is a nonnegative random variable representing the unobserved factors affecting demand, with $a_0(\cdot)$ and $a_1(\cdot)$ being
functions satisfying $a_1(\cdot) \geq 0$. For such a problem, we can prove\textsuperscript{3} quasi-concavity of the profit function and our comparative statics results. This follows from the fact that our previous proof techniques used sample path arguments which still apply since, for each realization $\hat{\epsilon} \geq 0$, $a_1(\hat{\epsilon})f(T) + a_0(\hat{\epsilon})$ behaves in a manner similar to $f(T) + \hat{\epsilon}$. Hence, results analogous to the additive case apply. For instance, $\mathcal{J}(g)$ is concave in $g = f(T)$ and the newsvendor-like formula for $T$ is

$$\int_{x:0 \leq a_1(x)g + a_0(x) \leq \rho} a_1(x)d\Phi_\epsilon(x) = \frac{(c^o - p)}{c^o - c}E[a_1(\epsilon)].$$

The above results may be further generalized to the case $\eta(T) = a_1(\epsilon_1)f(T) + a_0(\epsilon_0)$.

\section{Summary and Future Work}

To summarize, in this paper we have developed a model for integrating demand and production planning which permits us to determine an optimal production planning interval (and a corresponding optimal guaranteed maximum lead time). Several analytical properties of the profit and the optimal production interval were presented including comparative statics and numerical results on the effect of parameter changes on the optimal interval. This work is a first step in model-building to integrate production, outsourcing, and demand management. The managerial implications of our analysis are rather straightforward. We have highlighted in this analysis the increasing importance of lead time for customers, as well as the tradeoffs in achieving a proper balance between revenue and cost drivers associated with lead-time guarantees. In practice, achieving the increased profitability from properly structured lead-time guarantees is going to require careful analysis of costs, capacity commitments and due-date management systems, such

\footnote{The proof of this result has been omitted and can be obtained from the authors.}
as those embodied in Enterprise Resource Planning software. The underlying managerial implication of this work is the growing importance of coordinating cost and revenue drivers across marketing and production activities. In the case analyzed here, revenue increases from lead-time guarantees have to be carefully balanced against the cost of achieving these by production and logistic operators. Striking this balance will require companies to coordinate their sales and operations functions using S&OP procedures (Palmatier and Crum 2002). This will no doubt lead to further challenges, including segmenting customers according to the value attached to lead-time guarantees and other extensions to the framework developed here.

The limitations of our model are as follows. (i) We assume that the demand rate changes as a combination of additive and multiplicative effects of the quoted lead time (§ 7.2). Although this function does capture several simple relationships, there are more complicated relationships which our model does not generalize to easily; (ii) We assume that the same lead time is quoted to all customers for all products. Our main motivation for choosing the same lead time across the different products is that these products are often minor variants of each other belonging to the same family and hence quoting the same lead time is appropriate in that setting. While we recognize that our results only apply under some restrictions, we note that our demand model is more general than the deterministic demand functions common in the literature, and the model analyzed is applicable to a reasonably large set of environments. For instance, results from this paper may be directly applied to the following problem variants: (a) the case of seasonal demand rates, where the demand rate cycles through a set of values $\eta_1(T)$, $\eta_2(T)$, $\ldots$, $\eta_s(T)$ and we want to determine the optimal $T^*$ to quote in each season, and (b) the case where a fixed cost, $K$, is charged each time the production plan is updated (which
adds a cost $K/T$ to the cost function, while maintaining quasi-concavity of the profit function).

A Comparative Statics Results

Proof of Proposition 4. Since $\epsilon^1 \leq \epsilon^2$, we can construct $\hat{\epsilon}^1$ and $\hat{\epsilon}^2$ such that, $\hat{\epsilon}^1 \overset{st}{=} \epsilon^1$, $\hat{\epsilon}^2 \overset{st}{=} \epsilon^2$, and $\hat{\epsilon}^1 \leq \hat{\epsilon}^2$ a.s. (Stoyan 1983). Recall that, $J(T,\epsilon) = p(f(T) + \epsilon) - q(f(T),\epsilon)$. Consequently, for each $\epsilon$,

$$J'(T,\epsilon) = pf'(T) - q'(f(T),\epsilon)f'(T).$$

Note that, for $g = f(T)$, the per period cost function $q(g,\epsilon)$ may be written as $\tilde{q}(g + \epsilon)$, and that $\tilde{q}$ is convex (hence $q'$ is increasing). It follows that, since $f(T) + \hat{\epsilon}^1 \leq f(T) + \hat{\epsilon}^2$ a.s.,

$$\tilde{q}'(f(T) + \hat{\epsilon}^1) \leq \tilde{q}'(f(T) + \hat{\epsilon}^2) \text{ a.s.}$$

Since $f'(T) < 0$, we have, $\tilde{q}'(f(T) + \hat{\epsilon}^1)f'(T) \geq \tilde{q}'(f(T) + \hat{\epsilon}^2)f'(T)$ a.s. Therefore, $J'(T,\hat{\epsilon}^1) \leq J'(T,\hat{\epsilon}^2)$ a.s. Taking expected values, we get $E[J'(T,\hat{\xi}^1)] \leq E[J'(T,\hat{\xi}^2)]$. Since $\hat{\epsilon}^1 \overset{st}{=} \epsilon^1$ and $\hat{\epsilon}^2 \overset{st}{=} \epsilon^2$, we know that

$$E[J'(T,\hat{\epsilon}^1)] = E[J'(T,\hat{\epsilon}^1)] \leq E[J'(T,\hat{\epsilon}^2)] = E[J'(T,\epsilon^2)].$$

Exchanging the order of expectation and the derivative operators (which is valid by continuity), we have: $J'(T)|_{\hat{\epsilon}^1} \leq J'(T)|_{\hat{\epsilon}^2}$, which implies that $T^*(\hat{\xi}^1) \leq T^*(\hat{\xi}^2)$. ■

Acknowledgement: The authors thank the POM journal editors and referees, Dr. Suresh Chand, and several seminar/conference participants for their valuable comments and suggestions that helped us improve this paper.
References


Sodhi, M. S. (2005). Managing demand risk in tactical supply chain planning for
a global consumer electronics company. *Production and Operations Management* 14(1), XX–XX.


