A Model for Optimal Delivery Time Guarantees

Arthur V. Hill
Julie M. Hays
Eitan Naveh
University of Minnesota

This article formulates a model for finding the optimal delivery time performance guarantee. The expected profit model is solved to find a closed-form expression for the optimal delivery time promise. The simple, yet powerful model gives new insights into performance service guarantees in general and delivery time guarantees in particular.

Many manufacturing and distribution firms guarantee to meet a standard delivery time promise and pay significant compensation to customers when deliveries are late. If the firm promises its customers a delivery time that is too short, it will frequently not make the promise, have to pay significant compensation, and possibly lose market share over time. If the delivery time guarantee is too long, customers will find the delivery time unattractive and will buy elsewhere. Hart (1993) and others (Hill 1995) call this a "performance service guarantee."

Although a delivery time performance-guarantee scenario will be used as the context for this article, other performance service guarantee contexts could have been used as well. Other similar performance service guarantees include no-stockout guarantees (Hart 1993), waiting time guarantees (Friedman and Friedman 1997; Kumar, Kalwani, and Dada 1997), and up-time maintenance guarantees (Hill 1992). It should be noted, however, that an unconditional satisfaction guarantee (Hart 1988) is more complex and is not addressed in this article.

According to data collected by the Center for Advanced Purchasing Studies, delivery promises are far from perfect in many industries in the United States. On-time delivery benchmarks for several industries are summarized in Table 1.

In their book on time-based competition, Stalk and Hout (1990) state that

> Generally, if a time-based competitor can establish a response (that is) three or four times faster than its competitors, it will grow at least three times faster than the market and be at least twice as profitable as the typical industry competitor. (p. 98)

They also argue that firms with shorter customer lead times can segment the market to capture higher margin customers, leaving price-sensitive customers to the competition (Stalk 1988; Stalk and Hout 1990). For example, Socket Express, Inc., a custom socket manufacturer, captured 2% of a $400 million market in just 4 years based on the delivery time performance guarantee that "Every order is shipped on time with zero defects or the order is free" (Benson 1990).

Li and Lee (1994) and Lederer and Li (1997) developed queuing models that support Stalk and Hout's (1990) basic contention. Other authors also argue that firms should fo-

This research was supported by grants from the National Science Foundation Transformations to Quality Organizations (Grant SBR-9811047), Radisson Hotels Worldwide, the Juran Center for Leadership in Quality at the University of Minnesota, and the Institute for Management Development (Lausanne, Switzerland). The authors wish to thank Dr. David Fischer of Alliant Tech Systems for his helpful comments.

Journal of Service Research, Volume 2, No. 3, February 2000 254-264

TABLE 1
On-Time Delivery Benchmarks

Industry	*Industry Average*
Aerospace/ defense	Of all purchased lots, 84.1% were received on or before their contracted due date. Of all requisitions, 71.7% were received by the requirements planning organization in time to meet the need date of the user; 67.1% were received by the buyers in time to meet the need date of the user (Langdon et al. 1997).
Semiconductors	Percentage of deliveries received on-time was 80.9% for direct materials, 91.3% for indirect materials, 89.3% for capital equipment, and 81.7% for facilities/construction (Langdon and Richelsoph 1997).
Electronics	Of all deliveries, 91% were received on time (Langdon, Schwerman, and Seiter 1996).
Personal care products	Of parts purchased, 88.7% were received on or before their contracted/scheduled due date (Langdon and Lawson 1997).
Pharmaceuticals	Of all deliveries, 90.2% were received on time (Langdon, Richelsoph, and Schultz 1998).

cus on delivery time to stimulate demand (Blackburn 1991; Bockerstette and Shell 1993; Meyer 1993). Treacy and Wiersema (1995) describe the "operationally excellent" firm as one with short customer lead times. Much of the literature on "mass customization" deals with approaches to avoiding tradeoffs between customization and delivery time (Goldman, Nagel, and Preiss 1995; Kotha 1995; Pine 1993; Victor and Boynton 1998). Grout (1996) takes the buyers' perspective and considers penalties when the seller does not satisfy a "guaranteed" delivery time set by the buyer. He argues that 100% on-time delivery is neither achievable nor optimal.

Much research attention has been devoted to setting due dates for job shops (Baker and Bertrand 1981; Cheng and Gupta 1989; Markland, Fry, and Philipoom 1989; Philipoom, Rees, and Wiegmann 1994). These researchers assume that orders can have different promise times based on customer requirements and shop load and that demand is independent of the delivery promise time. Other researchers have developed methods for determining a "common delivery time" for a given set of orders (Conway 1965; De, Ghosh, and Wells 1991, 1992).

This article will focus on the situation in which the firm offers a standard delivery time guarantee to an entire market segment and in which the demand is influenced by the delivery time guarantee. Actual delivery time, T, is treated as a random variable. The firm's delivery time guarantee, t, is the decision variable. The goal is to find the optimal delivery time guarantee, t^*, to maximize expected profit per period, which is the expected gross margin per period less the expected compensation per period paid to customers for late deliveries.

The article is organized as follows. Section 2 develops the basic model followed by a discussion of several implementation and extension issues in Section 3. Section 4 presents a hypothetical example. Section 5 discusses the limitations of the model. The last section concludes the article by summarizing the contributions of the research.

THE MODEL FOR OPTIMAL DELIVERY TIME GUARANTEES

The Objective

The objective is to find the optimal delivery time guarantee, t^*, that maximizes the total expected profit per period for the firm:

$$\Pi(t) = \mu D(t) - L(t), \tag{1}$$

where $\Pi(t)$ is the total expected profit per period, μ is the average gross margin per order, $D(t)$ is the expected demand per period (in orders) for a delivery time guarantee of t periods, and $L(t)$ is the expected tardiness cost per period for a delivery time guarantee of t periods.

The Expected Demand Model

Define $D(t)$ as the expected demand in orders per period and $D(t)/M$ as the corresponding expected market share, where M is the size of the total "time-sensitive demand" market in orders per period. $D(t)$ decreases monotonically with t but at a decreasing rate. If the firm promises a zero order delivery time, it captures 100% of the time-sensitive market; if the firm promises a very long delivery time ($t \approx \infty$), the firm captures almost none of the market. The expected demand in orders per period, therefore, can be modeled as a function of t:

$$D(t) = e^{-\beta t}M, \text{ where } \beta > 0,\ M > 0,\ t > 0. \tag{2}$$

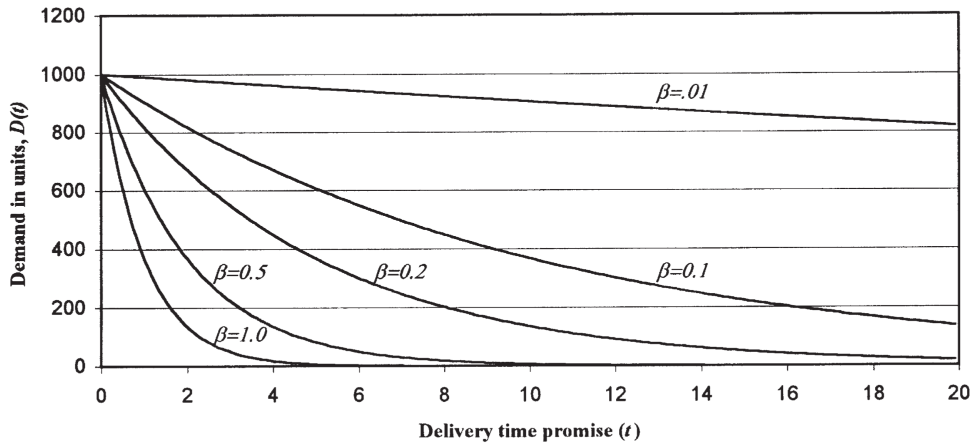
If the market is not time sensitive (i.e., $\beta = 0$), the optimization problem is analogous to the well-known "newsvendor" problem (Khouja 1995, 1996) that finds the optimal balance between the cost of earliness and the cost of tardiness.

From the above expression, the delivery time elasticity of demand, β, can easily be estimated from the firm's current delivery time guarantee, $t_0 > 0$, and current market share, $S_0 = D(t_0)/M > 0$, as

$$\beta = -ln(S_0)/t_0. \tag{3}$$

Figure 1 shows the demand function for a range of β parameters with a market size of $M = 1{,}000$ orders per period.

FIGURE 1
Graph of Demand Versus Delivery Time With M = 1,000 Orders/Period



The firm captures the entire time-sensitive market with a zero delivery time.

Hill and Khosla (1992) proposed an alternative approach when they developed a model for the relationship between customer lead time and demand in the context of evaluating the benefits of lead time reduction for a manufacturing firm. They modeled the demand function as $D(t) = at^{-b}$, where $b > 0$ was called the "lead time elasticity of demand." Although both approaches have some appeal, the approach used in this article is more intuitive for this context because demand is bounded at M. In this article, β is the "delivery time promise elasticity of demand," or, more generally, the "performance elasticity of demand" for a performance service guarantee.

Rust and Metters (1996) present a number of mathematical models that relate the consumer behavior literature to service management variables such as delivery time. Wirtz (1998) also connects service guarantee design decisions to the consumer behavior literature and supports Berry and Yadav's (1996) hypothesis that a well-designed service guarantee can increase demand by decreasing perceived customer risk. Tucci and Talaga (1997) provide empirical support for this viewpoint in a simulated study of service guarantees in the restaurant industry. These articles support the basic models presented in this article.

The Tardiness Cost Model

The actual delivery time, T, is a random variable with a distribution function, $F(T)$. Table 1 suggests that most firms set their delivery time promises in the tail of the delivery time distribution. Several authors (Bratley, Fox, and Schrage 1987; Law and Kelton 1991) recommend the exponential distribution for estimating probabilities in the tail of a distribution. With the exponential distribution, the probability that the delivery time for a customer, T, is greater than the delivery time guarantee, t, is

$$P(T \geq t) = 1 - F(t) = 1 - (1 - e^{-t/\tau} = e^{-t/\tau} \qquad (4)$$

where τ is the mean of the delivery time distribution. Other distributional forms for the tail of the delivery time distribution are considered in the extensions and limitations section.

The firm compensates the customer c_0 dollars per late delivery, incurs an additional g dollars of lost goodwill per late delivery, and, therefore, incurs a tardiness cost of $c = c_0 + g$ per late delivery. It is likely that c is larger for more time-sensitive markets (i.e., markets with a larger β value). For a given delivery time guarantee, t, the tardiness cost per period is

$$L(t) = cP(T \geq t)\, D(t) = c\, e^{-t(\beta + 1/\tau)} M. \qquad (5)$$

The basic model assumes that c_0, the compensation paid to customers, is sufficient to fully recover (retain) customers who experience late deliveries and that the time-sensitive market size, M, is not affected by late deliveries. Hays and Hill (1999) developed a different mathematical model that explicitly considers situations in which market share is affected by service failures and by service recovery policies.

The Total Expected Profit Model

The total expected profit function for a delivery time guarantee, t, is, therefore,

$$\Pi(t) = \mu\, D(t) - L(t) = M\,[\mu e^{-\beta t} - ce^{-t(1/\tau + \beta)}]. \qquad (6)$$

Taking the derivative of the expected profit function with respect to the decision variable, t, and setting it to zero yields

$$\delta\Pi(t)/\delta t = M[-\beta\mu e^{-\beta t} + c\,(1/\tau + \beta)\, e^{-t(1/\tau+\beta)}] = 0$$

$$\Rightarrow t^* = ln(\alpha c/\mu)\tau. \tag{7}$$

If τ is constant, then $\alpha = (1 + \beta\tau)/(\beta\tau) = [ln(S_0) - t_0/\tau]/ln(S_0)$ is also constant. This model for t^* is called the "basic model" for the remainder of the article.

The second derivative of the expected profit function with respect to t is

$$\delta^2\Pi(t)/\delta t^2 = M[\beta^2\mu e^{-\beta t} - c\,(1/\tau + \beta)^2 e^{-t(1/\tau+\beta)}]. \tag{8}$$

Evaluating the second derivative at $t^* = ln(\alpha\ c/\mu)\tau$ yields

$$\delta^2\Pi(t^*)/\delta t^2 = M[\beta^2\mu\varepsilon^{-\beta\ ln(\alpha\ c/\mu)\tau}$$

$$-c\,(1/\tau + \beta)^2 e^{-ln(\alpha\ c/\mu)\tau\,(1/\tau+\beta)}]$$

$$\Rightarrow \delta^2\Pi(t^*)/\delta t^2 = -M[(1/(\beta\tau) + 1)\ c/\mu]^{-\beta\tau}\ \beta\mu/\tau. \tag{9}$$

Given that $M > 0$, $\beta > 0$, $\tau > 0$, $c > 0$, and $\mu > 0$, the second derivative evaluated at t^* is less than 0, which confirms that t^* is the global maximum for the expected profit function.

The factor $ln(\alpha c/\mu)$ "inflates" or "deflates" the mean delivery time, τ, to compute the optimal delivery time guarantee. The optimal delivery time, t^*, increases slowly as the ratio c/μ increases. The optimal delivery time guarantee should be more conservative (longer) when the tardiness cost increases. Firms should offer a more aggressive (shorter) delivery time guarantee when the profit margin increases. The constant α reflects the sensitivity of the firm's optimal delivery time guarantee to changes in the c/μ ratio.

IMPLEMENTATION ISSUES AND EXTENSIONS

Estimating the Delivery Time Elasticity of Demand and Market Size Parameters

As mentioned above, the delivery time elasticity of demand can be estimated from the firm's current delivery time guarantee, t_0, and current market share, $S_0 = D(t_0)/M$, resulting in the expression $\beta = -ln(S_0)/t_0$. The current demand, $D(t_0)$, and market size, M, are only for the time-sensitive portion of the overall market. If the firm has no experience with changing its delivery time guarantee, M must be estimated from market data. However, if the delivery time guarantee has been changed (or a change could be predicted from market research), then β and M can be estimated by fitting the demand function to the two points, (t_0, D_0) and (t_1, D_1), using the following two equations:

$$\beta = ln(D_1/D_0)/(t_1 - t_0) \tag{10}$$

$$M = D_0\ exp(\beta t_0). \tag{11}$$

Direction of Change

The simplest implementation of the model is to determine only the direction of change in the delivery time guarantee, given that the firm currently has a delivery time guarantee of t_0. The delivery time guarantee should be increased when

$$t^* > t_0 \Rightarrow c/\mu > exp(t_0/\tau)/\alpha. \tag{12}$$

Therefore, whenever the ratio c/μ is greater than the "critical value" $exp(t_0/\tau)/\alpha$, the firm should increase its delivery time guarantee. Conversely, if ratio c/μ is less than the critical value, the firm should decrease its delivery time guarantee.

Inferred Tardiness Cost

An interesting alternative approach for understanding the model is to assume that t_0, the firm's current delivery time guarantee, is the optimal delivery time guarantee. The basic model can be rewritten to infer the tardiness cost parameter from the current delivery time guarantee:

$$t_0 = ln(\alpha c/\mu)\ \tau \Rightarrow c^* = exp(t_0/\tau)\mu/\alpha. \tag{13}$$

The firm should have consistency between the inferred tardiness cost parameter and its choice of delivery time and compensation parameters. If the actual compensation, c_0, is less than c^*, the firm should consider either raising its compensation or lowering its delivery time guarantee. Conversely, if c_0 is greater than c^*, the firm should consider lowering its compensation or increasing its delivery time guarantee.

Benefits of Improving the Mean Delivery Time

The basic model for t^* can be extended to measure the benefits of improving the mean delivery time, τ. Given the nonlinear form of the optimal delivery time guarantee function, small changes in τ yield large changes in expected profit. Define $\tau_0 > 0$ as the mean delivery time when no money is spent to shorten (expedite) the mean delivery time. Define $q \geq 0$ as the expediting expense per customer order. As q increases, the mean delivery time decreases but at a diminishing rate:

$$\tau(q) = \tau_0 e^{-\kappa q}, \tag{14}$$

where $\kappa \geq 0$ is the parameter of the model. The problem is to find the optimal values t^* and q^* that maximize the revised expected profit function:

$$\Pi(t, q) = (\mu - q)D(t, q) - L(t, q) = \tag{15}$$
$$(\mu - q)exp(-\beta t)M - c\ exp\{-t[\tau_0^{-1} exp(\kappa q)] + \beta\}M.$$

The total expected profit function does not yield a closed-form solution for the optimal expediting expense, q^*; however, the optimal q^* and t^* can be found by applying standard nonlinear optimization techniques.

Benefits of Improving Both the Mean and the Standard Deviation of the Delivery Time

The basic model assumes that the tail of the delivery time distribution is exponential, which is characterized by only one parameter, the mean of the distribution. Although the exponential is a reasonable approximation for the tail of many density functions, using a two-parameter distribution can provide greater accuracy and additional opportunities for analysis. Management might be able to employ methods such as "on-call" workers, subcontractors, or express mail to reduce both the mean and standard deviation of delivery time. This section develops an extension of the basic model to evaluate these types of investments.

Delivery times typically follow the gamma distribution (Law and Kelton 1991). This is because delivery time is the sum of several independent task times such as order entry time, order picking time, order packing time, travel time, and so forth, and the gamma distribution is the sum of independent exponential distributions (Hillier and Lieberman 1974). The gamma has the desirable characteristics that it is bounded at zero and can take on a wide variety of shapes.

Figure 2 shows the density functions for the gamma(k, τ/k) distribution for $k = \{1, 2, 5, 10\}$. Note that the gamma(1, τ/k) is the exponential(τ) and that the gamma(k, τ/k) with $k \in$ {integers} is the k-Erlang(τ/k). Figure 3 shows the probability of late delivery for these distributions. As k increases, the gamma(k, τ/k) approaches the normal(τ, σ) with $\sigma = \tau/k^{1/2}$. The difference between the gamma(k, τ/k) and normal(τ, $\tau/k^{1/2}$) at $k = 10$ is quite small. The financial analysis, therefore, should consider the benefits of reducing both the mean and the standard deviation of the delivery time guarantee (increasing k). For $k > 1$, delivery time is not exponential, and numeric methods can be used to find the optimal delivery time guarantee, t^*, and the optimal distribution parameters, τ^* and k^*, to optimize the expected profit $\Pi(t, \tau, k)$.

Price Elasticity of Demand

When implementing a service guarantee, managers would like to know how much they could raise their price. The model can be extended to also consider price elasticity of demand:

$$D(t, p) = e^{-\beta t/p}M, \text{ where } \beta > 0, t > 0, p > 0, M > 0. \tag{16}$$

The expected profit function then becomes

$$\Pi(t, p) = (p - u)D(t, p) - L(t), \tag{17}$$

where p and u are the average order price and cost, respectively. This has no closed-form solution for t^* and p^*. However, nonlinear optimization methods can find the optimal price for a delivery time guarantee or both the optimal delivery time guarantee and the optimal price.

Multiple Competitors Model

Market demand can be allocated to competitors in the market based on both the delivery time guarantee and the price. The attractiveness of competitor i can be defined as $A_i(t_i, p_i) = exp(-\beta\ t_i, p_i)$. The expected demand for competitor i is then $D_i(t_i, p_i) = [A_i(t_i, p_i)/S]M$, where S is the sum of the A_j for all competitors in the market. This model assumes that β, the elasticity parameter, is a constant for all competitors in the market. If current prices, delivery time guarantees, and market shares are known, a nonlinear optimization can be used to find the least squares fit estimate for β. A firm could use this type of model to estimate the impact of changes in delivery time guarantees and prices on expected demand, market share, and profit.

Marketing Pressure Model

Fornell (1987) discusses several "marketing effort" variables such as advertising, sales effort, and promotion that affect demand. Define a as the expense per period for advertising or some other comparable marketing effort. The demand model with marketing effort is then

$$D(t, a) = e^{-\beta t/a}M, \text{ where } \beta > 0, t > 0, a > 0, M > 0 \tag{18}$$

The expected profit per period function then becomes

$$\Pi(t, a) = \mu D(t, a) - L(t) - a. \tag{19}$$

FIGURE 2
Density Function for the Gamma (k, τ/k) Distribution With a Mean τ = 3 Days

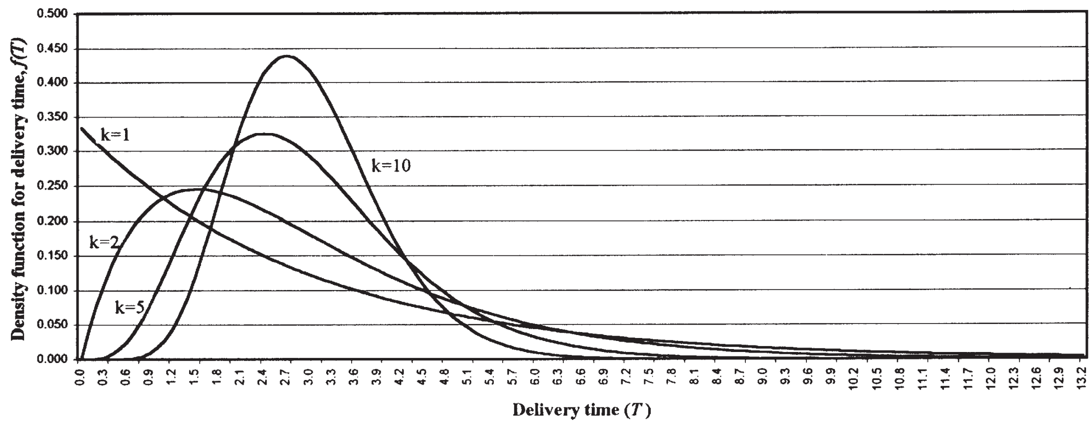
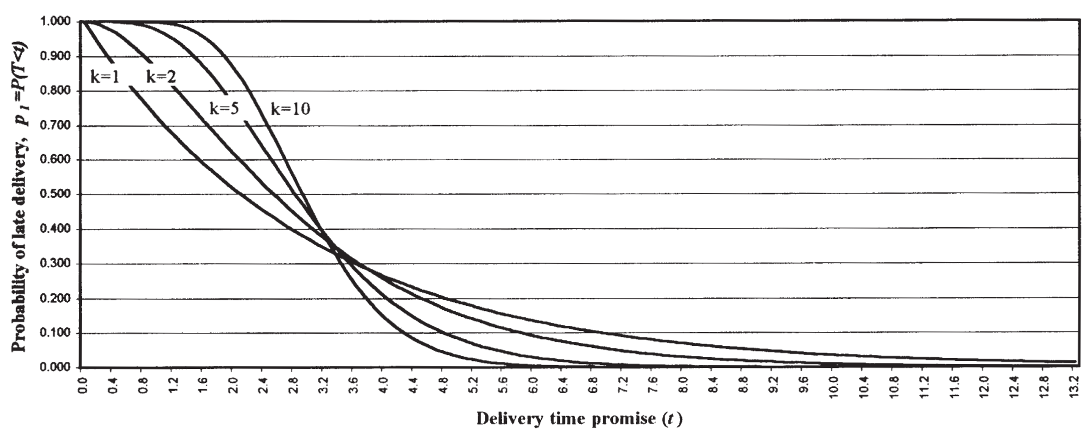


FIGURE 3
Probability of Late Delivery for the Gamma (k, τ/k) Distribution With Mean τ = 3 Days



Again, this function has no closed-form solution for t^* and a^* but can be solved numerically.

HYPOTHETICAL EXAMPLE

A custom door manufacturing firm has a delivery time distribution that is exponential with a mean of $\tau = 3$ days. The firm has an average sales price of $p = \$1{,}000$ and an average gross margin of $\mu = \$100$ per order. Doors are shipped one at a time and customers are compensated $c_0 = \$275$ per late delivery. Management believes that this compensation is sufficient to completely recover customers who experience late deliveries and that no goodwill is lost (i.e., $g = 0$ and $c = c_0$). The firm currently sells $D_0 = 135$ orders per day on average and offers a delivery time guarantee of $t_0 = 10$ days. The firm currently has approximately 4% of the overall custom door market. Management does not know the firm's market share of the time-sensitive market. From a market research survey, management finds that reducing the delivery time guarantee to $t_1 = 9$ days increases demand to $D_1 = 165$ orders per day. From this data, management finds that $\beta = ln(D_1/D_0)/(t_1 - t_0) \approx 0.2$ and $M = D_0\, exp(\beta t_0) \approx 1{,}000$ orders per day. The current time-sensitive market share, therefore, is approximately $S_0 = D_0/M \approx 13.5\%$. The α constant is $[ln(S_0) - t_0/\tau]/ln(S_0) \approx 2.66$.

As a first estimate, the firm tests to see if it should increase or decrease its delivery time guarantee. The ratio $c/\mu = 275/100 = 2.75$ is less than the critical value $exp(t_0/\tau)/\alpha \approx 10.5$, which indicates that it should decrease its delivery time guarantee. From the basic model, the opti-

FIGURE 4
Expected Profit Versus Delivery Time Graph for the Example

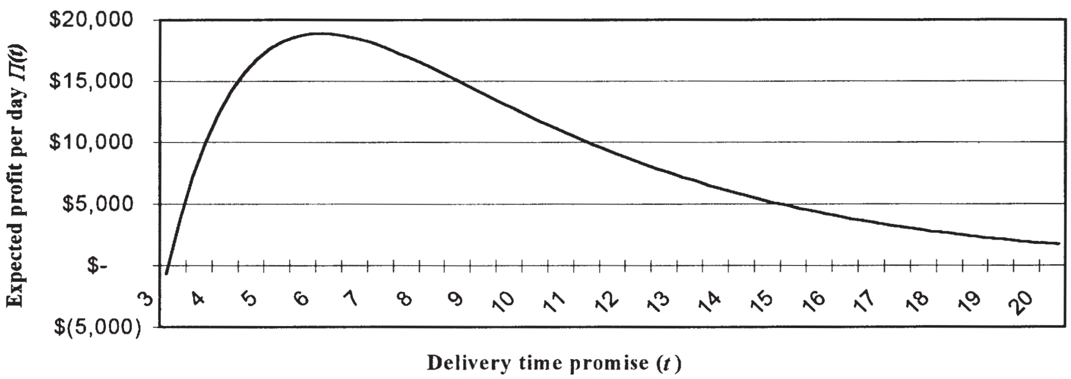


TABLE 2
Example Problem Comparison of Current and Optimal Delivery Time Guarantees

Performance Metric	*Current Delivery Time Guarantee*	*Optimal Delivery Time Guarantee*	*Change*
Delivery time guarantee	10 days	6 days	Down 4 days
Demand per day	135 orders	301 orders	Up 166 orders
Probability of a late delivery	4%	14%	Up 10%
Gross margin per day	$13,534	$30,119	Up $16,586
Tardiness cost per day	$966	$11,210	Up $9,882
Total profit per day	$12,568	$18,910	Up $6,704

mal delivery time guarantee is $t^* = ln(\alpha c/\mathrm{m})\tau \approx 6$ days with a daily expected profit $\Pi(6) = \$18{,}910$.

Figure 4 shows a graph of the expected profit function for this firm. If the firm promised a delivery time of less than 3 days, it would lose money. It is clearly better for this firm to give a longer promise time by a few days than to promise a day too early. When the firm promises a delivery time of 10 days, it would earn about one third less than the optimal expected profit.

Table 2 shows the implications of changing the delivery time guarantee from 10 days to the optimal value of 6 days. The shorter delivery time guarantee increases the firm's expected profit per period by more than 50%. The percentage of late deliveries increases from 4% to 14%, which increases tardiness cost per day from $966 to $11,210. This increase in tardiness should drive management to critically reevaluate the tardiness cost parameter. Customer compensation, c_0, needs to be large enough to recover nearly all customers who are inconvenienced by late deliveries; the goodwill parameter, g, should reflect the cost of negative word of mouth. The firm could increase the tardiness cost parameter to as much as $439 per late delivery and still increase its expected profit with the 6-day delivery time guarantee.

After considering analysis in Table 2, management increases the goodwill parameter (g) from $0 to $125 per late delivery. This changes the tardiness cost parameter to $c = c_0 + g = \$275 + \$125 = \$400$ per late delivery. The optimal delivery time guarantee is then $t^* = 7$ days, which yields a daily expected profit of $15,094, a 27% increase over the current situation. The probability of a late delivery for the 7-day delivery time guarantee is 7%.

Figure 5 shows that the optimal delivery time guarantee, t^*, increases slowly as the ratio c/μ increases. This suggests that $t = 6$ is nearly optimal for a fairly wide range of c/μ values. Assuming integer delivery promises, $t = 6$ is optimal for c/μ in the range (2.36, 3.27). With the gross margin, μ, fixed at $100, the optimal delivery time guarantee is 6 days for tardiness cost, c, in the range ($236, $327).

The tardiness cost parameter inferred by the current 10-day delivery time guarantee is $c^* = exp(t_0/\tau)\mu/\alpha \approx$ $1,051, which is far more than the $c_0 = \$275$ that the firm currently compensates customers for late deliveries. This suggests that the firm should either make much shorter delivery time guarantees to stimulate demand or consider increasing the compensation amount to be consistent with its long delivery time guarantee.

Improving the mean delivery time from $\tau = 3$ to $\tau = 2$ reduces t^* from 6 days to 4.5 days, increases the expected daily demand from 366 orders per day to 407 orders per day, and increases the optimal expected profit from $22,892 to $28,873 (a 26% improvement). The firm should be willing to spend up to $14.70 per order ($5,981 per day) to achieve this one third reduction in mean delivery time.

FIGURE 5
Optimal Delivery Time Guarantee, t^*, Versus the c/μ Ratio

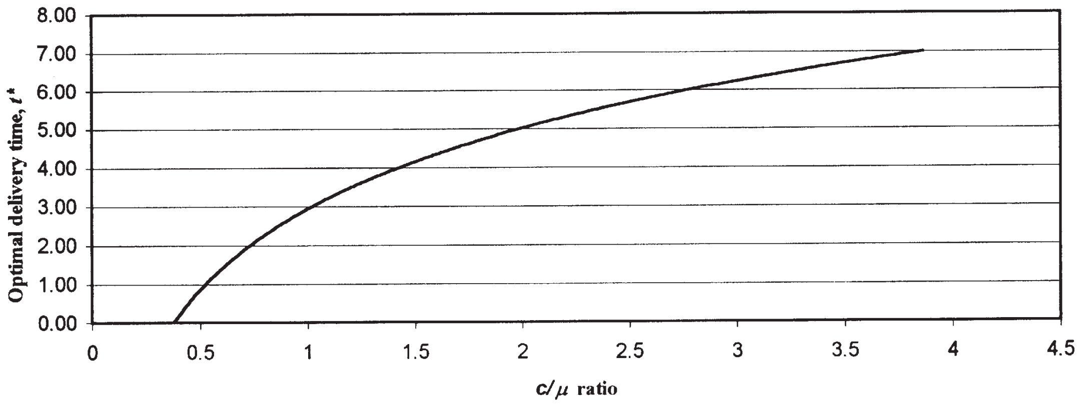


TABLE 3
Expected Profit per Day Versus the Standard Deviation of Delivery Time

k Parameter for the Gamma Distribution	*Standard Deviation of the Delivery Time in Days, $\tau/k^{1/2}$*	*Optimal Delivery Time Guarantee in Days, t^**	*Optimal Expected Profit per Day (\$), Π (t, k)*	*Probability of Late Delivery With Delivery Time Guarantee t = 6 days (%)*
1	3.00	6.0	18,910	14
2	2.12	5.8	22,571	9
3	1.73	5.6	25,238	6
4	1.50	5.5	27,253	4
5	1.34	5.3	28,849	3
6	1.22	5.2	30,156	2
7	1.13	5.1	31,257	1
8	1.06	5.0	32,201	1
9	1.00	4.9	33,025	0.7
10	0.95	4.9	33,752	0.5

FIGURE 6
Optimal Expected Profit per Day Versus the Standard Deviation of Delivery Time

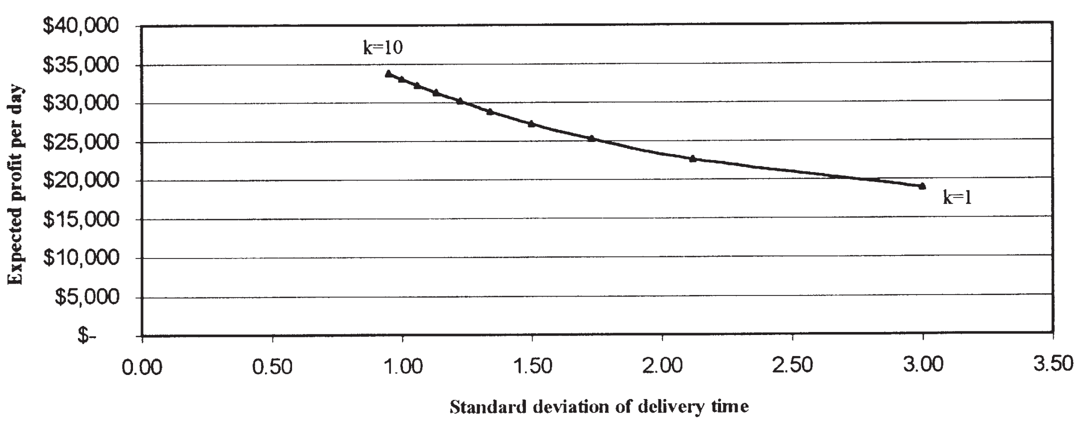


TABLE 4
Standard Deviation of Delivery Time Versus Probability of Late Delivery for t = 6 Days

k Parameter for the Gamma Distribution	*Standard Deviation of the Delivery Time in Days, $\tau/k^{1/2}$*	*Probability of Late Delivery With Delivery Time Guarantee t = 6 days (%)*
1	3.00	14
2	2.12	9
3	1.73	6
4	1.50	4
5	1.34	3
6	1.22	2
7	1.13	1
8	1.06	1
9	1.00	.7
10	0.95	.5

TABLE 5
Difference $P_g - P_e$ When Using the Exponential (τ) to Approximate the Gamma (k, τ/k) (in percentages)

	k									
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
.30	0	0	0	0	–1	–1	–1	–2	–2	–2
.25	0	–1	–2	–3	–4	–4	–4	–5	–5	–5
.20	0	–2	–4	–5	–6	–7	–7	–8	–8	–9
P_g .15	0	–4	–6	–7	–8	–9	–10	–11	–11	–12
.10	0	–4	–7	–9	–10	–11	–12	–13	–14	–14
.05	0	–4	–7	–9	–11	–12	–13	–14	–15	–16
.01	0	–3	–5	–7	–9	–10	–11	–13	–13	–14

Management wants to analyze the benefits of reducing the standard deviation of delivery time. Table 3 shows the optimal t^* and the corresponding expected profit as the standard deviation of the delivery time decreases (as k increases). For $k > 1$, delivery time is not exponential, and the optimal t^* values have to be found numerically.

Figure 6 displays the Table 3 information in a graph. This graph suggests that expected profit for the optimal delivery time guarantee increases roughly linearly with a decrease in the standard deviation of delivery time. Management also can use the expected profit model to analyze alternatives for simultaneously decreasing the mean and the standard deviation of the delivery time distribution.

Table 4 assumes a delivery time guarantee of 6 days and shows how the probability of late delivery decreases as the standard deviation of delivery time decreases. The firm could keep late deliveries at 4% (the current value) if it could increase k from 1 to 4 (decrease the delivery time standard deviation from 3 to 1.5). This would require that the firm better manage the "tail" of the delivery time distribution.

LIMITATIONS

The model assumes that the tardiness cost parameter, $c = c_0 + g$, includes all tardiness-related costs, such as lost gross margin and the cost of negative word of mouth. It is important to note that a well-managed service recovery process can mitigate these costs (Smith, Bolton, and Wagner 1998).

Following De, Ghosh, and Wells (1991, 1992), the basic model assumes that the tardiness cost for an order is zero if the delivery is on time and a constant ($c = c_0 + g$) if the delivery is late. The model could be extended to define tardiness cost as a linear function of tardiness, $c(t) = c_0 \, max(T - t, 0)$ and could be further extended to include an earliness cost. Unfortunately, such extensions make the model far less tractable.

The model assumes that the delivery time/demand relationship has no discontinuities. It is possible that once a firm has achieved a superior delivery time in its market, further reductions in delivery time may not result in significantly increased demand. The model also assumes no competitive response to a change in a firm's delivery time guarantee. A game theoretic model could be developed to explore such issues. Without considering a competitive response, the basic model might be used to estimate an upper bound on the increase in profit.

As shown in Figures 2 and 3 above, delivery times are likely to follow the gamma distribution rather than the exponential distribution (Law and Kelton 1991). Table 5 shows the difference, $P_g - P_e$, where $P_g = P_{gamma}(T > t)$ and $P_e = P_{expo}(T > t)$ are the probabilities of late delivery for the gamma(k, τ/k) and exponential(τ) distributions, respectively. The differences are independent of the mean delivery time. The important observation here is that when the delivery time promise is in the tail ($P_g < .3$) and the delivery time distribution is not exponential (e.g., $k > 1$), the exponential tends to overestimate the probability of tardiness and, therefore, tends to err on the "conservative" side. When the exponential has an unacceptably large error, numeric methods can be applied to find the optimal delivery time guarantee, t^*, for the gamma or any other delivery time distribution.

In some cases, the delivery time distribution follows a shifted exponential distribution in which the $P(T < t_{min}) = 0$ for the minimum possible delivery time t_{min}. For these situations, variable transformations can be performed to implement the basic model.

Although the model is built on a solid conceptual foundation, it has not yet been supported by empirical research.

We are pursuing empirical tests of the model to validate both the submodels and the overall model.

CONCLUSIONS

Service guarantees are an important tool for many firms to stimulate demand, increase prices, and improve profits. This article presents a simple but powerful model for finding the optimal performance service guarantee promise in a delivery time context. The model makes explicit tradeoffs between the benefits of short delivery time guarantees and the cost of late deliveries. With only definitional changes, the model can be applied to other performance service guarantee contexts in which the demand is influenced by the guarantee and the performance variable follows the exponential distribution. For example, the basic model can be applied to find the optimal parameter for a waiting time performance guarantee (Kumar, Kalwani, and Dada 1997; Friedman and Friedman 1997). (Waiting time for a customer is exactly exponential for the M/M/1 queue and gamma distributed in some other queuing situations.) The basic model can also be applied to set the "up-time" parameter for an "up-time" maintenance performance guarantee (Hill 1992). (Time between failure is often assumed to be exponential.) With some modification, the model also can be applied to the "no stockout" performance guarantee context (Hart 1993; Silver, Pyke, and Peterson 1998).

The model is surprisingly simple and, as a result, provides useful intuition for designing performance service guarantees. The model also provides insights into setting the compensation amount, estimating the size of the time-sensitive market, evaluating investments in improving process capabilities (the mean and standard deviation of the delivery time), and pricing.

The model offers several opportunities for empirical research. For example, it would be interesting to explore the relationship between customer retention and the difference between the inferred tardiness cost and actual compensation, $\Delta c = c^* - c_0$. We hypothesize that firms with lower Δc are paying a more consistent compensation amount and, therefore, will do a better job of retaining their time-sensitive customers. It would also be interesting to test if the difference between the firm's actual and optimal delivery time guarantees, $\Delta t = t - t^*$, is a useful metric for "time-based competition" (Stalk and Hout 1990). We hypothesize that firms with a lower Δt are more aggressively pursuing a time-based competition strategy.

The model gives researchers new insights into how performance service guarantees work and how they might be optimized. The simple, yet powerful, model may prove to be a useful tool for analyzing performance service guarantees for both service and manufacturing firms.

REFERENCES

Baker, Kenneth R., and J.W.M. Bertrand (1981), "A Comparison of Due Date Selection Rules," *AIIE Transactions*, 13, 123-31.

Benson, Tracy E. (1990), "Socket Express Connects with Guarantee," *Industry Week*, May 21, 39-42.

Berry, Leonard L. and Manjit S. Yadav (1996), "Capture and Communicate Value in the Pricing of Services," *Sloan Management Review*, 37 (4), 41-51.

Blackburn, Joseph D. (1991), *Time-Based Competition: The Next Battle Ground in American Manufacturing*. Business One Irwin/APICS Series in Manufacturing. Homewood, IL: Business One Irwin.

Bockerstette, Joseph A. and Richard L. Shell (1993), *Time Based Manufacturing*. New York: McGraw-Hill.

Bratley, Paul, Bennett L. Fox, and Linus E. E. Schrage (1987), *A Guide to Simulation*, 2nd ed. New York: Springer-Verlag.

Cheng, T.C.E. and M. C. Gupta (1989), "Survey of Scheduling Research Involving Due Date Determination Decisions," *European Journal of Operational Research*, 38, 156-66.

Conway, R. W. (1965), "Priority Dispatching and Job Lateness in a Job Shop," *Journal of Industrial Engineering*, 16, 228-37.

De, Prabuddha, Jay B. Ghosh, and Charles E. Wells (1991), "Optimal Delivery Time Quotation and Order Sequencing," *Decision Sciences*, 22, 379-90.

———, ———, and ——— (1992), "Optimal Due Date Assignment and Sequencing," *European Journal of Operational Research*, 57, 323-31.

Fornell, Claes (1987), "Defensive Marketing Strategy by Customer Complaint Management: A Theoretical Analysis," *Journal of Marketing Research*, 24 (November), 337-46.

Friedman, Hershey H and Linda W. Friedman (1997), "Reducing the 'Wait' in Waiting-line Systems, Waiting Line Segmentation," *Business Horizons*, 40 (July-August), 54-58.

Goldman, Steven L., Roger N. Nagel, and Kenneth Preiss (1995), *Agile Competitors and Virtual Organizations: Strategies for Enriching the Customer*. New York: Van Nostrand Reinhold.

Grout, John R. (1996), "A Model of Incentive Contracts for Just-in-Time Delivery," *European Journal of Operational Research*, 96, 139-47.

Hart, Chris (1988), "The Power of Unconditional Service Guarantees," *Harvard Business Review*, 88 (July/August), 54-62.

——— (1993), *Extraordinary Guarantees*. New York: Amacom.

Hays, Julie M. and Arthur V. Hill (1999), "The Market Share Impact of Service Failures," *Production Operations Management*, forthcoming.

Hill, Arthur V. (1992), *Field Service Management*. Homewood, IL: Business One Irwin.

——— (1995), "Service Guarantees: The Fast-Track to Service Quality," *IMD Perspectives*, 2 (1), 1-4.

——— and Inder Khosla (1992), "Models for Optimal Lead Time Reduction," *Production and Operations Management*, 1 (2), 185-97.

Hillier, Frederick S. and Gerald J. Lieberman (1974), *Operations Research*, 2nd ed. San Francisco: Holden-Day.

Khouja, Moutaz (1995), "The Newsboy Problem under Progressive Multiple Discounts," *European Journal of Operational Research*, 84 (2), 458-66.

——— (1996), "The Newsboy Problem with Multiple Discounts Offered by Suppliers and Retailers," *Decision Sciences*, 27 (3), 589-99.

Kotha, Suresh (1995), "Mass Customization: Implementing the Emerging Paradigm for Competitive Advantage," *Strategic Management Journal*, 16 (Special Issue), 21-42.

Kumar, Piyush, Manohar U. Kalwani, and Maqbool Dada (1997), "The Impact of Waiting Time Guarantees on Customers' Waiting Experiences," *Marketing Science*, 16 (4), 295-314.

Langdon, Debra Seaman, Sherri Allen, David Schultz, and Charles Zheng (1997), *Benchmarks for the U.S Aerospace/Defense Contracting Industry: Purchasing Performance Benchmarks for Industry*. Tempe, AZ: Center for Advanced Purchasing Studies/National Association of Purchasing Management.

——— and Rosemary Lawson (1997), *Benchmarks for the Personal Care Products Industry: Purchasing Performance Benchmarks for Industry*. Tempe, AZ: Center for Advanced Purchasing Studies/ National Association of Purchasing Management.

——— and Jennifer Richelsoph (1997), *Benchmarks for the Semiconductors Industry: Purchasing Performance Benchmarks for Industry.* Tempe, AZ: Center for Advanced Purchasing Studies/National Association of Purchasing Management.

———, ———, and David Schultz (1998), *Benchmarks for the Pharmaceutical Industry: Purchasing Performance Benchmarks for Industry*. Tempe, AZ: Center for Advanced Purchasing Studies/ National Association of Purchasing Management.

———, Tammy Schwerman, and Bret Seiter (1996), *Benchmarks for the Electronics Industry: Purchasing Performance Benchmarks for Industry*. Tempe, AZ: Center for Advanced Purchasing Studies/National Association of Purchasing Management.

Law, Averill M. and W. David Kelton (1991), *Simulation Modeling and Analysis*, 2nd ed. McGraw-Hill Series in Industrial Engineering and Management Science. New York: McGraw-Hill.

Lederer, Phillip and Lode Li (1997), "Pricing, Production, Scheduling, and Delivery Time Competition," *Operations Research*, 45 (3), 407-20.

Li, Lode and Yew Sing Lee (1994), "Pricing and Delivery Time Performance in a Competitive Market," *Management Science*, 40 (5), 533-646.

Markland, Robert E., Timothy D. Fry, and Philip R. Philipoom (1989), "Due Date Assignment in a Multistage Job Shop," *IIE Transactions*, 21 (2), 153-62.

Meyer, Christopher (1993), *Fast Cycle Time*. New York: Free Press.

Philipoom, Patrick R., Loren P. Rees, and Lars Wiegmann (1994), "Using Neural Networks to Determine Internally-Set Due-Date Assignments for Job-Shop Scheduling," *Decision Sciences*, 25 (5/6), 825-52.

Pine, B. Joseph II (1993), *Mass Customization: The New Frontier in Business Competition*. Cambridge, MA: Harvard Business School Press.

Rust, Roland T. and Richard Metters (1996), "Mathematical Models of Service," *European Journal of Operational Research*, 91 (3), 427-39.

Silver, Edward A., David F. Pyke, and Rein Peterson (1998), *Inventory Management and Production Planning and Scheduling*, 3rd ed. New York: John Wiley.

Smith, Amy K., Ruth N. Bolton, and Janet Wagner (1998), "A Model of Customer Satisfaction with Service Encounters Involving Failure and Recovery," working paper, Marketing Science Institute, Report No. 98-100, February.

Stalk, George (1988), "Time: The Next Source of Competitive Advantage," *Harvard Business Review*, 66 (4), 41-52.

——— and Thomas M. Hout (1990), *Competing against Time: How Time-Based Competition Is Reshaping Global Markets*. New York: Free Press.

Treacy, Michael and Fred Wiersema (1995), *The Discipline of Market Leaders: Choose Your Customers, Narrow Your Focus, Dominate Your Market*. Reading, MA: Addison-Wesley.

Tucci, Louis A. and James Talaga (1997), "Service Guarantees and Customers' Evaluation of Services," *Journal of Services Marketing*, 11 (1), 10-18.

Victor, Bart and Andrew C. Boynton (1998), *Invented Here: Maximizing Your Organization's Internal Growth and Profitability.* Boston: Harvard Business School Press.

Wirtz, Jochen (1998), "Development of a Service Guarantee Model," *Asia Pacific Journal of Management*, 15 (1), 51-75.

Arthur V. Hill is a professor in the Curtis L. Carlson School of Management at the University of Minnesota. He earned a B.A. in mathematics from Indiana University, an M.S. in industrial administration, and a Ph.D. in management from the Krannert School at Purdue University. Professor Hill has extensive international experience, including work in Moscow, South Africa, Hungary, Austria, Switzerland, and Singapore. He has served as the editor-in-chief of the *Journal of Operations Management* and on the editorial review boards of *Decision Sciences*, *Production and Operations Management*, *Production and Inventory Management Journal*, and *Technology and Operations Review.* His current research interests include service guarantees as means to accelerate service quality improvement and manufacturing management for medical device manufacturing.

Julie M. Hays is a doctoral student in the Operations and Management Science Department in the Curtis L. Carlson School of Management at the University of Minnesota. She holds a B.S. in chemical engineering from the University of Minnesota and an M.B.A. from the College of St. Thomas. Her dissertation research is on service quality issues. She is a 1998 Juran Fellow and is heavily involved in a $140,000 grant from National Science Foundation/Transformations to Quality Organizations to study service guarantees.

Eitan Naveh is a postdoctoral fellow in the Center for the Development of Technological Leadership in the Institute of Technology at the University of Minnesota. He holds a B.Sc. in industrial engineering and management, an M.Sc., and a D.Sc. in quality assurance and reliability, all from the Technion–Israel Institute of Technology. His current research interests include development and implementation of quality systems, quality standardization, and returns on quality.