

---

## 26 Strategic pricing response and optimization in operations management

*Teck H. Ho and Xuanming Su*

---

### **Abstract**

This chapter reviews how rational customers or buyers should respond to firms' pricing decisions and how firms should optimize prices as a consequence of these strategic responses. The key departure from standard economic and marketing pricing research is that either customers or firms are simultaneously faced with other operational choices, such as capacity sizing and inventory control, which are of central interest to operations management researchers. The chapter covers four broad areas and it is intended to serve as a selective rather than comprehensive review of the extensive literature.

### **1. Introduction**

The main purpose of this chapter is to provide a selective review of pricing models, with an emphasis on issues that are of interest to operations management researchers. Apart from pricing decisions, these models tend to explicitly incorporate supply-side considerations that reflect physical characteristics of production processes, such as inventory control, capacity constraints and demand uncertainty. In such settings, there are two broad questions: how should rational customers respond to firms' pricing decisions, and how should firms optimize prices to maximize profits?

In this review, we focus on the following four broad areas. The first two areas cover pricing and inventory decisions, and the last two areas cover pricing in the presence of capacity constraints.

1. *EOQ inventory models* The classic economic ordering quantity (EOQ) model is an inventory model that is typically applied to products with a relatively stable consumption pattern over time. One main advantage is that this model applies to the buyer as both a consumer and producer. The model shows how the buyer should react operationally in response to the seller's pricing decisions. On the consumer side, the model addresses questions such as optimal shopping frequency and stockpiling decisions. On the producer side, the model addresses issues such as fixed costs (e.g. due to batch production or transportation costs) and inventory carrying costs. This class of models studies producers' pricing and inventory decisions as well as consumers' purchase and inventory decisions in response to sellers' pricing decisions.
2. *Newsvendor inventory models* The newsvendor model is another classic inventory model in operations management. It is ideally suited for analyzing a business-to-business setting where a retailer must cope with demand uncertainty by ordering from a manufacturer that has long production lead times before the short selling season begins. The central dilemma captured by newsvendor models is the tradeoff between excess inventory that remains unsold and profit losses due to insufficient orders. Here, we explore how pricing (demand-side control) can complement inventory ordering decisions (supply-side control) to improve the retailer's profits.

3. *Dynamic pricing models* For many products, the firm has a finite capacity that cannot be replenished. Examples include airplane tickets, hotel rooms, and even fashion apparel. In these settings, revenue management tactics are commonly used to optimize prices over time in response to sales performance. Apart from dynamic pricing by firms, we also review models that consider consumers' dynamic responses to these prices, i.e. whether they should buy or wait for discounts. Here, we focus on the effect of finite capacity, which creates additional considerations for firms (since unsold units have little to no value) as well as for consumers (since the item may be sold out if the consumer waits too long).
4. *Queueing models* In operations management, queueing models are typically used to capture capacity constraints in service settings. Unlike the dynamic pricing models above (where a customer may either get the item or not), queueing models admit intermediate outcomes. Here, the waiting time dimension is used to reflect various forms of service degradation associated with excess demand, relative to available capacity. Pricing can then be used to influence demand to improve profits for the firm. These models study firms' pricing decisions in queueing contexts, and also address consumers' decisions (e.g. whether to enter the queue given the price, how much 'work' to send to the queue, which priority level to choose, and so on).

In each of the areas listed above, we shall highlight the significance of operational considerations. One common theme across all areas is that consumers are active decision-makers and respond strategically to these operational issues (so we refer to them as rational or strategic consumers). For example, in EOQ models, consumers choose purchase quantities and make stockpiling decisions, and in the dynamic pricing models, consumers choose the timing of their purchases. As a result, firms' pricing decisions serve as an important strategic lever to shape consumer behavior and optimize profits. To draw a stark comparison, we occasionally compare the above models with their counterparts that do not have the corresponding operational issues. The next four sections review the four areas listed above. We provide some closing remarks and suggest broad directions for future research in the concluding section.

## 2. EOQ-based pricing models

The standard economic ordering quantity (EOQ) model is perhaps one of the most fundamental models in operations management. In the standard EOQ setup, a seller charges a time-invariant price  $p$ . A rational customer who has a constant consumption  $r$  per unit time must decide how much ( $Q$ ) to order to minimize her total costs over time.<sup>1</sup> The ordering policy must consider three cost components: the purchase cost ( $p$  per unit); the fixed cost in placing an order, such as traveling cost (denoted by  $K$ ); and the cost associated with inventory holding ( $h$  per unit per time period). This implicitly assumes that holding cost  $h$  does not depend on price or quantity purchased. As a result, this model does not account for the cost of capital associated with holding inventory (e.g. financial holding cost).

---

<sup>1</sup> There is an equivalent utility maximization problem. Since the consumption rate is exogenous and fixed over time, the ordering policy that minimizes the total cost also maximizes the utility.

The rational customer's decision is to choose an optimal ordering quantity  $Q$  to minimize the total cost ( $TC$ ) per unit time. Note that the rational customer places orders at an interval of  $Q/r$  and at the time when the inventory is zero. The objective function of the optimization problem is

$$TC(Q) = p \cdot r + \frac{Kr}{Q} + h \cdot \frac{Q}{2} \tag{26.1}$$

where the first term on the left-hand side gives the purchase cost per unit time, the second term is the fixed cost per unit time and the last term is the average holding cost per unit time. Solving the problem yields the optimal economic ordering quantity as follows:

$$Q^* = \sqrt{\frac{2Kr}{h}} \tag{26.2}$$

Note that the optimal economic ordering quantity does not depend on the firm's pricing decision  $p$  even though the optimal total cost increases linearly with  $p$ . Let us illustrate the above formula with a numerical example. Let the consumption rate ( $r$ ) be 5 units per month, the setup cost per order ( $K$ ) be \$10, and the holding cost be \$1 per unit per month. Using equation (26.2), we compute the optimal ordering quantity  $Q^*$  as 10 units per order. That is, the consumer must place an order of 10 units every two months.

Ho et al. (1998) extend the EOQ model to investigate how a rational customer should strategically respond when the firm's pricing policy fluctuates over time (i.e.  $p$  is a random variable). The fluctuation of the price is described by a time-invariant probability distribution that consists of  $S$  scenarios (i.e. a general discrete distribution with pricing scenario  $p_s$  occurs with a probability  $\pi_s$ ). The rational customer has knowledge of the price distribution but is unaware of the price realization before incurring the fixed cost (or in their context a shopper traveling to visit a grocery store). Once the fixed cost is sunk, the customer observes the price  $p_s$  and then chooses purchase quantity  $Q_s$ . Let  $\mu_p = \sum_{s=1}^S \pi_s \cdot p_s$  and  $\sigma_p^2 = \sum_{s=1}^S \pi_s \cdot (p_s - \mu_p)^2$  be the mean and variance of the distribution respectively. The customer's ordering policy is to decide how much to order under each pricing scenario  $s$ ,  $Q_s$ . As before, the total cost per unit time under pricing scenario  $s$  is given by

$$TC(Q_s) = \frac{Kr}{Q_s} + p_s \cdot r + \frac{h \cdot Q_s}{2} \tag{26.3}$$

It can be shown that the long-run average cost per unit time is given by

$$\mu_{TC}(Q_1, \dots, Q_S) = \frac{K + \sum_{s=1}^S [\pi_s \cdot p_s \cdot Q_s + \pi_s \cdot h \cdot Q_s^2 / (2 \cdot r)]}{\sum_{s=1}^S \pi_s \cdot Q_s / r} \tag{26.4}$$

The customer must select a purchase quantity under each scenario  $s$  in order to minimize the above long-run average cost per unit time.

The optimal ordering quantity under pricing scenario  $s$  shown to be

$$Q_s^* = \sqrt{\frac{2K^*r}{h}} - \frac{r}{h} \cdot (p_s - \mu_p) \tag{26.5}$$

where  $K^* = K - (r/2h) \cdot \sigma_p^2$ . Consequently, the expected economic ordering quantity is given as follows:

$$\mu_Q^* = \sqrt{\frac{2K^*r}{h}} \quad (26.6)$$

Note that the optimal ordering quantity is no longer independent of price once the latter is random. It is now linear in price  $p_s$ . This linear ordering rule states that the ordering quantity in scenario  $s$  is proportional to the difference between the reference price (the average price  $\mu_p$ ) and the price of scenario  $p_s$ . Interestingly, the expected ordering quantity is decreasing in the variance of the price ( $\sigma_p^2$ ). So a higher price fluctuation induces the customer to place more but smaller orders by providing an option value (or ordering flexibility) that effectively reduces the fixed cost of placing an order. Consequently, the rational customer shops more often and places a smaller order when variance of the price increases. The authors test these predictions on an extensive dataset from a grocery chain and find strong support for them. The authors also extend the model by allowing the customer to adjust her consumption rate  $r$  in response to price fluctuation. They show that it is optimal for the customer to increase her consumption rate if variance of the price increases.<sup>2</sup>

Let us use the same example above to examine the influence of price variability. Assume there are two pricing scenarios (i.e. regular (\$10) and discounted (\$8)), each occurring with equal probability of 0.5. Consequently, the price variance is  $\sigma_p^2 = 1$  and the revised setup cost is  $K^* = 10 - (5/2 \cdot 1) \cdot 1 = 7.5$ . Using equation (26.6), the average ordering quantity becomes 8.67 units, which is smaller than the average ordering quantity of ten units under no price variability. Given the same consumption rate, the consumer must shop more frequently.

Assunção and Meyer (1993) consider a similar problem but in a shopping context where there is no fixed cost associated with placing an order (e.g.  $K = 0$ ). In their setting, the customer pays periodic visits to the store (and hence the travel costs are sunk) and must decide in each period how many units to order and consume given a current inventory holding level ( $Z$ ) and observed price ( $p$ ). The price fluctuation is assumed to follow a Markov process (i.e. the immediate future price is only a function of the most recent observed price). Formally, the customer must solve the following dynamic programming problem:

$$V(Z, p) = \max_{Q, r} \{U(r) - p \cdot Q - h(Z + Q - r) + \beta \cdot E[V(Z + Q - r, p_{t+1}) | p]\} \quad (26.7)$$

where  $U(\cdot)$  is the utility from consumption and is concave in  $r$  and  $h(\cdot)$  is the holding cost of inventory. The authors provide structural results on the optimal purchase and consumption policies. They show that the customer should order in a period only if the current inventory level  $Z$  is below a threshold level  $I(p)$  (which is a function of price only).

---

<sup>2</sup> It is assumed that utility is concave in consumption and the customer maximizes the net utility, which is the difference between utility from consumption per unit time and the long-run average cost per unit time.

The optimal ordering quantity is given by  $I(p) - Z$  (i.e. order up to  $I(p)$ ). The optimal consumption is shown to be always increasing in the current inventory ( $Z$ ) level independent of the observed price  $p$ .<sup>3</sup> The customer buys less and consumes more if the holding cost increases. Also, both purchase and consumption are decreasing in the observed price  $p$  as long as some sensible assumptions about future price expectations hold.

Kunreuther and Richard (1971) extend the EOQ model to consider the situation where the customer is a retailer who must simultaneously decide how many units to order ( $Q$ ) from the distributor and how much to charge the product to a group of end customers. The end-customer demand per unit time and the retail price have a one-to-one mapping so that the retailer's pricing decision is mathematically equivalent to its consumption or sales rate decision ( $r$ ). Formally, the retailer's profit function is given by

$$\Pi(Q, r) = R(r) - \frac{K \cdot r}{Q} - \frac{h \cdot Q}{2} - p \cdot r \tag{26.8}$$

where  $R(r)$  is the retailer's revenue when it sells  $r$  units per period. Differentiating the profit function with respect to  $Q$  and  $r$  yields the following two simultaneous equations:

$$Q^* = \sqrt{\frac{2 \cdot K \cdot r^*}{h}} \tag{26.9}$$

$$\frac{dR(r)}{dr} = p + \sqrt{\frac{K \cdot h}{2 \cdot r^*}} \tag{26.10}$$

where  $(R(r)/dr)$  is the marginal revenue per unit time from the last unit sold when the sales rate is  $r^*$ . The authors use the above results to investigate the costs of sequential decision-making (or lack of coordination). They consider an environment where the marketing department first chooses  $r^*$  ignoring the fixed and inventory holding costs. The purchase department then chooses an ordering quantity taking the sales rate  $r^*$  as given (i.e. solving the standard EOQ problem). They find that the costs of sequential decision-making can be high if the product of fixed cost and holding cost  $K \cdot h$  is high and the optimal sales rate  $r^*$  is small.

Blattberg et al. (1981) show that an economic reason that a retailer might offer special sales or deals on its products is the transfer of inventory holding costs from the firm to its customers. Their model framework consists of two sub-models, one for retailer and one for the customer. The idea here is to solve for an equilibrium with each party seeking to minimize its total costs. In their customer model (and with usual notations), customer  $i$  chooses a order quantity  $Q_i$  in order to minimize the total cost over a purchasing cycle of  $Q_i/r_i$  given below:

$$TC(Q_i) = \frac{h_i \cdot Q_i}{2} \cdot \frac{Q_i}{r_i} - d \cdot Q_i \tag{26.11}$$

where the second term of the left-hand side is the total cost saving due to price deals  $d$  over the purchase cycle. Note that there is no fixed cost associated with placing an order (i.e.

---

<sup>3</sup> This positive relationship between consumption and inventory is used by Bell et al. (2002) in their model of customer behavior.

$K = 0$ ). Solving, we obtain the optimal ordering quantity ( $Q^*$ ) and the optimal purchase period ( $t^*$ ) as follows:

$$Q_i^* = \frac{d \cdot r_i}{h_i} \quad (26.12)$$

$$t^* = \frac{d}{h_i} \quad (26.13)$$

Customers are segmented into two groups: the first with high holding costs ( $h_H$ ) and the second with lower holding costs ( $h_L$ ). All customers have the same consumption rate  $r$ . There is a total of  $N$  customers of which an  $\alpha$  ( $0 < \alpha < 1$ ) fraction has low holding costs. It is assumed that only customers with low holding costs buy on deals. Consequently, the aggregate quantity bought on deal is

$$Q_D = \alpha \cdot N \cdot Q_2^* = \alpha \cdot N \cdot \frac{d \cdot r}{h_L} \quad (26.14)$$

In their retailer model setup, the retailer must choose a deal amount  $d$  and the length of reordering period  $T$  (i.e. retailer's ordering quantity divided by its sales rate) in order to minimize its total cost per unit time. The total cost consists of a fixed cost per order  $K_R$ , a holding cost ( $h_R$  per unit per unit time), and costs associated with sales. They show that the optimal deal amount and the reorder period are given as follows:

$$d^* = \left[ \frac{K_R \cdot h_L}{N \cdot r(\alpha + (1 - \alpha) \cdot \frac{h_R}{2 \cdot h_L})} \right]^2 \quad (26.15)$$

$$T^* = \frac{d^*}{h_L} \quad (26.16)$$

Besides the optimal deal amount  $d^*$ , the optimal dealing frequency ( $f^*$ ) is the number of deals offered in any given time interval  $\tau$  and is simply  $(\tau/t^*) = \tau \cdot (h_L/d^*)$ . The predictions of the overall model are: (1) the deal amount increases when the holding cost to the customer ( $h_L$ ) and the fixed cost ( $K_R$ ) cost increase. It decreases as the the total consumption rate ( $N \cdot r$ ) increases; and (2) the deal frequency increases when the holding cost ( $h_L$ ) and the total consumption rate increase and when the fixed cost decreases. The authors find support for these predictions using a panel dataset and hence establish the transfer of inventory explanation as a plausible rationale for price promotion.

Jeuland and Narasimhan (1985) consider a similar problem and study how a monopolist firm should set its price when it serves two groups of customers with different consumption rates. Each group of customers  $i$  has a consumption rate  $r_i$  conditioned on the firm's price  $p$  given as follows:

$$r_i = \alpha_i - p \quad (26.17)$$

where  $\alpha_1 > \alpha_2 > 0$ . That is, given a price  $p$ , group 1 customers consume more than group 2 customers. A key assumption is that the high-demand (i.e. group 1) customers have a higher inventory holding cost so that the two groups shop differently when faced with price promotion. Customers are assumed to make periodic visits to the firm so that

travel costs are sunk (i.e. customers have zero fixed costs,  $K = 0$ ). The firm gives a discount  $d$  once every  $T$  periods.

Because of high inventory holding cost, group 1 customers never stockpile or forward-buy, so they purchase and consume in a deal period a quantity given by  $r_{1d} = \alpha_1 - p + d$ . During nondeal periods, these customers purchase and consume  $r_{1n} = \alpha_1 - p$ . Group 2 customers always forward-buy during the deal period for consumption in all periods. These customers consume at a rate of  $r_2 = \alpha_2 - p + d - (hT/2)$  in every period, where  $h$  is the inventory cost per unit per period.<sup>4</sup> The authors establish that it is indeed possible and profitable for the firm to price-discriminate between the two groups of customers by offering promotion deals occasionally. This work provides a theoretical reason why a firm might want to give discount deals in practice. It highlights a necessary condition (i.e. high consumption rate must be accompanied by high inventory cost) for such a promotion strategy to be successful.

Bell et al. (2002) extend Jeuland and Narasimhan's model and study how homogeneous firms should engage in price promotion in a competitive setting where customers might increase their consumption as a result of inventory holding. They consider a two-period model in which rational customers must decide how much to buy in each period. A customer has two choices, buy either one unit or two units in the first period depending on the observed price. If the customer buys one unit, she consumes one unit and must incur a fixed cost  $K$  to visit the store again in the second period. In she buys two units, there are two possible consumption scenarios. In the first consumption scenario (which occurs with probability  $(1 - \theta)$ ), the customer consumes one unit and must incur a cost  $h$  to hold the second unit for consumption in the second period. The customer does not visit the store in the second period. In the second consumption scenario (which occurs with probability  $\theta$ ), the customer consumes two units and must incur a cost  $K$  to visit the store in the second period. The authors show that the symmetric equilibrium profits for each firm are

$$\Pi^* = \frac{v \cdot [v + h(1 - \theta)] \cdot N}{[v - h(1 - \theta)] \cdot n} \tag{26.18}$$

where  $v$  is the per unit utility of the product to a customer,  $N$  is the total number of customers and  $n$  is the total number of firms in the industry.

It can be readily shown that equilibrium profits decrease as  $\theta$  increases. That is, increased consumption effect due to price fluctuation intensifies price competition. This phenomenon occurs because the increased consumption leads to deeper price discounts and an increase in the frequency of promotions. They examine both predictions using purchase data of eight categories from a grocery chain. Some categories (e.g. bacon, soft drinks etc.) are likely to have a higher consumption effect (i.e. higher  $\theta$ ) while others (e.g. bathroom tissue, detergent etc.) are likely to have a smaller consumption effect (a lower  $\theta$ ). Overall, they find support for their predictions.

### 3. Newsvendor-based pricing models

Like the EOQ model, the newsvendor model is a celebrated and classic model in operations management. Here, we consider a retailer who must order a product with a short

---

<sup>4</sup> Note that the average inventory holding cost reduces the consumption rate.

life cycle from a distributor in order to serve a group of end customers. The end-customer demand is random, following a well-behaved cumulative distribution given by  $F(\cdot)$  and a density function and distribution given by  $f(\cdot)$  and  $F(\cdot)$  respectively. The retailer buys the product at price  $w$  and sells it at price  $p$ . It has a short and well-defined selling cycle so that any unsold product must be salvaged. The retailer must place an order of size  $Q$  before it knows the actual demand. The retailer faces two possible scenarios after the demand realization.<sup>5</sup> In the first scenario, the actual demand is higher than the order. Here the retailer experiences a foregone profit of  $(p - w)$  per unit of unfulfilled demand. In the second scenario, the retailer has an overstock of supply because the actual demand is smaller than the order. Consequently the firm incurs a cost of  $(w - s)$  per unit of leftover supply where  $s$  is the unit salvage value of the product. Let the unit underage cost  $C_u = (p - w)$  and the unit overage cost  $C_o = (w - s)$ . Formally, the retailer chooses an ordering quantity in order to minimize the total expected costs as follows:

$$EC(Q) = C_o \int_0^Q (Q - q) \cdot f(q) \cdot dq + C_u \int_Q^\infty (q - Q) \cdot f(q) \cdot dq \quad (26.19)$$

It can be shown that the optimal ordering quantity ( $Q^*$ ) is given by

$$F(Q^*) = P(D \leq Q^*) = \frac{C_u}{C_u + C_o} \quad (26.20)$$

That is, the retailer should order at a level  $Q^*$  that sets the probability of serving all customers to the ratio given by  $(C_u/C_u + C_o)$  (i.e. the relative underage cost). Readers are referred to Porteus (2002) for more details. Let us use a numerical example to illustrate the formula given in equation (26.20). Let  $w = \$2$ ,  $p = \$4$  and  $s = \$1$ . Then  $C_u = 4 - 2 = \$2$  and  $C_o = 2 - 1 = \$1$ . Consequently, we have  $F(Q^*) = (2/2 + 1) = (2/3)$ . If demand follows a normal distribution with mean 100 and standard deviation of 20, then the optimal ordering quantity becomes  $Q^* = 100 + 0.435 \cdot 20 = 108.7$  units.

Petruzzi and Dada (1999) extend the standard newsvendor problem by allowing the retailer to choose the stocking quantity and price simultaneously. Randomness in demand is captured in either an additive or a multiplicative form as follows:

$$D(p, \varepsilon) = y(p) + \varepsilon = a - b \cdot p + \varepsilon \quad (26.21)$$

$$D(p, \varepsilon) = y(p) \cdot \varepsilon = a \cdot p^{-b} \cdot \varepsilon \quad (26.22)$$

where  $\varepsilon$  has a support  $[L, H]$ , a density function  $f(\cdot)$ , and a distribution function  $F(\cdot)$ . It is assumed that leftovers are disposed at the unit cost  $\alpha$  and shortages experience a per-unit penalty cost of  $\beta$ . If  $\alpha$  is negative (i.e. leftovers have a salvage value), then  $|\alpha| = s$  (as defined above). Also, if  $\beta > 0$ , there is a loss in goodwill (i.e. the basic newsvendor problem assumes  $\beta = 0$ ). That is, the underage and overage costs are given as  $C_u = (p - w + \beta)$  and  $C_o = w + \alpha$  respectively.

The profit  $\Pi(Q, P)$  can be written as

---

<sup>5</sup> We assume both order and demand are continuous so that the probability that the order is identical to demand is zero.

$$\Pi(Q, p) = \begin{cases} p \cdot D(p, \varepsilon) - w \cdot Q - \alpha \cdot [Q - D(p, \varepsilon)] & \text{if } D(p, \varepsilon) \leq Q \\ p \cdot Q - w \cdot Q - \beta \cdot [D(p, \varepsilon) - Q] & \text{if } D(p, \varepsilon) > Q \end{cases} \quad (26.23)$$

Consider the additive case and define the stocking factor  $z = Q - y(p)$ . Then the profit function can be rewritten as a function of  $(z, p)$  as

$$\Pi(z, p) = \begin{cases} P \cdot [y(p) + \varepsilon] - w \cdot [y(p) + z] - \alpha \cdot [z - \varepsilon] & \text{if } \varepsilon \leq z \\ p \cdot [y(p) + z] - w \cdot Q[y(p) + z] - \beta \cdot [\varepsilon - z] & \text{if } \varepsilon > z \end{cases} \quad (26.24)$$

Expected profit is:

$$E[\Pi(z, p)] = \int_L^z (p \cdot [y(p) + x] - \alpha \cdot [z - x]) \cdot f(x) \cdot dx \quad (26.25)$$

$$+ \int_z^H (p \cdot [y(p) + z] - \beta \cdot [x - z]) \cdot f(x) \cdot dx - w \cdot [y(p) + z]$$

$$= \Phi(p) - L(z, p) \quad (26.26)$$

where  $\Phi(p) = (p - c) \cdot [y(p) + E(\varepsilon)]$  and  $L(z, p) = C_o \cdot \Theta(z) + C_u \cdot \Omega(z)$  where  $\Theta(z) = \int_L^z (z - x) \cdot f(x) \cdot dx$  and  $\Omega(z) = \int_z^H (x - z) \cdot f(x) \cdot dx$ .

Let  $p^o = (a + b \cdot w + E(\varepsilon)/2b)$ . If  $y(p) = a - b \cdot p$ , then the optimization problem can be solved sequentially as follows:

1. For a fixed  $z$ , the optimal price is determined uniquely as  $p^* = p(z) = p^o - (\Omega(z)/2b)$ .
2. The optimal stocking quantity  $Q^*$  is given by  $Q^* = y(p^*) + z^*$ , where  $p^*$  is determined as above and  $z^*$  is determined as follows:
  - (a) If  $F(\cdot)$  is an arbitrary distribution, an exhaustive search over all possible values of  $z$  in the region  $[L, H]$  will determine  $z^*$ .
  - (b) If  $F(\cdot)$  satisfies the condition that  $2r(z)^2 + (dr(z)/dz) > 0$ , where  $r(z) = (f(z)/[1 - F(\cdot)])$  is the hazard rate, then  $z^*$  is the largest  $z$  in the region  $[L, H]$  that satisfies  $(dE(\Pi(z, p(z))))/dz = 0$ .
  - (c) If the condition for (b) is met and  $a - b(w - 2\alpha) + L > 0$ , then  $z^*$  is the unique  $z$  that satisfies  $(dE(\Pi(z, p(z))))/dz = 0$ .

In the multiplicative case, we have  $D(p, \varepsilon) = y(p) \cdot \varepsilon = a \cdot p^{-b} \cdot \varepsilon$ . Let  $z = (Q/y(p))$ , then the expected profit function can be rewritten like before as a sum of a deterministic profit term and an expected loss function term:

$$E[\Pi(z, p)] = \Phi(p) - L(z, p) \quad (26.27)$$

where  $\Phi(p) = (p - c) \cdot y(p) \cdot E(\varepsilon)$  and  $L(z, p) = y(p) \cdot [C_o \cdot \Theta(z) + C_u \cdot \Omega(z)]$ .

Let  $p^1 = (b \cdot w/b - 1)$ . As before, the optimization problem can be solved sequentially as follows:

1. For a fixed  $z$ , the optimal price is determined uniquely as  $p^* = p(z) = p^1 + (b/b - 1) \cdot [(C_o \cdot \Theta(z) + \beta \cdot \Omega(z))/E(\varepsilon) - \Omega(z)]$ .

2. The optimal stocking quantity  $Q^*$  is given by  $Q^* = y(p^*) \cdot z^*$ , where  $p^*$  is determined as above and  $z^*$  is determined as follows:
- If  $F(\cdot)$  is an arbitrary distribution, an exhaustive search over all possible values of  $z$  in the region  $[L, H]$  will determine  $z^*$ .
  - If  $F(\cdot)$  satisfies the condition that  $2r(z)^2 + (dr(z)/dz) > 0$ , where  $r(z) = (f(z)/[1 - F(\cdot)])$  is the hazard rate and  $b \geq 2$ , then  $z^*$  is the unique  $z$  that satisfies  $(dE(\Pi(z, p(z))))/dz = 0$ .

The authors provide a unifying interpretation of the above results by introducing the notion of a base price and showing that the optimal price in both the additive and multiplicative cases can be interpreted as the base price plus a premium.

Su and Zhang (2008) extend the standard newsvendor problem by allowing the retailer to choose the price and customers to choose their timing of purchase (either during the selling season or at the end of the season). In their model setup, the firm sells the product during the selling season at price  $p_r$  (or regular price) and at the end of the selling season at a salvage price  $p_s$  (the latter is exogenously fixed). Customers have valuation  $V$  for the product. They form an expectation of product availability and believe that they will obtain the product with a probability  $E_\phi$  at the end of the selling season. Therefore, customers will only buy the product during the selling season if  $V - p_r \geq (V - p_s) \cdot E_\phi$  or, equivalently,  $p_r \leq V - (V - p_s) \cdot E_\phi$ .

The retailer holds a rational expectation that all customers will not buy the product during selling season unless  $p_r \leq V - (V - p_s) \cdot E_\phi$ . Given this expectation, the retailer sets  $p_r^* = V - (V - p_s) \cdot E_\phi$ . Also, it chooses  $Q^*$  to maximize its profit given below:

$$\pi(Q|p_r^*) = p_r^* \cdot \min\{d, Q\} - w \cdot Q + p_s \cdot [Q - d]^+ \quad (26.28)$$

$$\text{or } \pi(Q|p_r^*) = (p_r^* - p_s) \cdot \min\{d, Q\} - (w - p_s) \cdot Q \quad (26.29)$$

where, as before,  $d$  is the demand and has a probability and distribution function given by  $f(\cdot)$  and  $F(\cdot)$  respectively.

If customers' expectation is rational (i.e.  $E\theta = F(Q^*)$  or the expectation of product availability equals to the actual fill rate during the selling season), then it can be shown that the optimal regular price and stocking quantity are

$$p_r^* = p_s + \sqrt{(V - p_r) \cdot (w - p_r)} \quad (26.30)$$

$$F(Q^*) = 1 - \sqrt{\frac{w - p_s}{V - p_s}} \quad (26.31)$$

Note that the optimal regular price is between  $w$  and  $V$ . Interestingly, the equilibrium stocking quantity  $Q^*$  is lower than that of the standard newsvendor problem. The following inequality shows this result:

$$F(Q^0) = 1 - \frac{w - p_s}{V - p_s} > 1 - \sqrt{\frac{w - p_s}{V - p_s}} = F(Q^*) \quad (26.32)$$

where  $Q^0$  is the optimal stocking quantity in the standard newsvendor problem.

Using the same numerical example for the optimal ordering quantity (26.20) in

the basic newsvendor model above, we have  $p_s = s = \$1$  and  $V = p = \$4$ . Hence  $F(Q^*) = 1 - \sqrt{(2 - 1/4 - 1)} = 0.5774$ , implying an optimal ordering quantity of 94.12 units (i.e.  $100 - 0.294 \cdot 20$ ). Notice that this optimal quantity is smaller than 108.7 units obtained before.

Dana and Petruzzi (2001) extend Petruzzi and Dada's (1999) model by allowing customers to actively choose whether or not to visit the retailer depending on its price  $p$  and stocking quantity  $Q$ , which are assumed to be known to the customers before they make their visit decision. Customers are heterogenous in two ways. First, they have a value of either  $V > 0$  or  $V = 0$  for the product. Second, each customer has an outside option  $u$ . The number of customers with a positive value  $V$  (assumed to be continuous),  $d$ , has a support of  $[L, H]$  and a probability and distribution function of  $f(\cdot)$  and  $F(\cdot)$  respectively. The expected number of customers with positive value is  $\mu = E(d)$ . Similarly, the outside option value has a probability and distribution function of  $g(\cdot)$  and  $G(\cdot)$ . The retailer must choose its price and stocking quantity *ex ante* (i.e. before observing the realizations of  $d$  and  $u$ ).

Let  $\hat{u}$  be the outside option of the marginal customer, the person who is indifferent between the outside option and visiting the store. Then the retailer's total demand is  $d \cdot G(\hat{u})$ . Its total sales is  $\min(Q, d \cdot G(\hat{u}))$ , so the expected demand and sales are

$$E[\text{Demand}(\hat{u})] = \int_L^H x \cdot G(\hat{u}) \cdot f(x) \cdot dx = \mu \cdot G(\hat{u}) \tag{26.33}$$

$$E[\text{Sales}(Q, \hat{u})] = \int_L^H \min(Q, x \cdot G(\hat{u})) \cdot f(x) \cdot dx = \mu \cdot G(\hat{u}) - E[d \cdot G(\hat{u}) - Q]^+ \tag{26.34}$$

Note that  $\hat{u}$  is a function of the retailer's price and stocking quantity. In fact  $\hat{u}$  solves the following implicit function:

$$\hat{u} = \phi(Q, \hat{u}) \cdot (V - p) \tag{26.35}$$

where  $\phi(Q, \hat{u})$  is the probability that a random customer is served (i.e. fill rate). It is the ratio of the expected sales and demand and is given by  $(E[\text{Sales}(Q, \hat{u})] / E[\text{Demand}(\hat{u})]) = 1 - (E[d - (Q/G(\hat{u}))]^+ / \mu)$ . The retailer's expected profit is

$$p \cdot E[\text{Sales}(Q, \hat{u})] - w \cdot Q \tag{26.36}$$

Let  $z = (Q/G(\hat{u}))$ . Then the fill rate can be rewritten as  $\Phi(z) = 1 - (E[d - z]^+ / \mu)$  and  $\hat{u}$  can be solved using the revised implicit function  $\hat{u} = \Phi(z) \cdot (V - p)$ . Since  $p = V - (\hat{u}/\Phi(z))$ , the retailer's optimization problem is to choose  $\hat{u}$  and  $z$  to maximize the expected profit given below:

$$\pi(z, \hat{u}) = \left[ V - \frac{\hat{u}}{\Phi(z)} \right] \cdot G(\hat{u}) \cdot \left[ \mu - \int_z^H [1 - F(x)] dx \right] - w \cdot z \cdot G(\hat{u}) \tag{26.37}$$

The authors first consider the case where the price is set exogenously. Here they show that a retailer that takes into account the effect of its stocking quantity on customers' visit decision carries more inventories, attracts more customers, and earns a higher expected

profit than a retailer that ignores this effect. When price is set endogenously, they show that the two-dimensional optimization problem can be reduced to two, sequential, one-dimensional optimization problems by first solving for  $z^*$  locally and then given  $z^*$  solving for  $u^*$  globally.

Deneckere and Peck (1995) extend the standard newsvendor problem by incorporating competition. In their model setup, there are  $n$  firms and each firm  $i$  simultaneously chooses a stock quantity  $Q_i$  and the price  $p_i$ . Customers know both the stocking quantities and prices of all firms before making their store visit decision. It is assumed that leftovers are disposed at zero cost and there is no loss in goodwill associated with shortages (i.e.  $\alpha = \beta = 0$  or  $C_o = w$  and  $C_u = p - w$ ). Note that the outside option of a customer for each firm is defined by other firms' strategies. As before, the number of customers has a probability and distribution function of  $f(\cdot)$  and  $F(\cdot)$  respectively. The mean number of customers is denoted by  $\mu$ . At equilibrium, customers choose a mixed strategy  $(ms_1, \dots, ms_n)$  and  $ms_i$  represents the probability that a customer visits firm  $i$ . The fill rate of firm  $i$  given a stocking quantity  $Q_i$  and  $ms_i$  is

$$\phi(Q_i | ms_i) = E \left[ \frac{\min(Q_i, ms_i \cdot d)}{ms_i \cdot d} \right] \quad (26.38)$$

Hence, firm  $i$ 's expected profit is given by

$$\Pi_i(p_i, Q_i) = p_i \cdot \mu \cdot ms_i \cdot \phi(Q_i | ms_i) - w \cdot Q_i \quad (26.39)$$

The authors show that an equilibrium in which all firms choose pure strategies, if it exists, is unique and is characterized as follows:

$$Q_i^* = \frac{F^{-1}\left(1 - \frac{w}{V}\right)}{n} \quad (26.40)$$

$$p_i^* = \frac{w \cdot F^{-1}\left(1 - \frac{w}{V}\right) \cdot \frac{n}{n-1}}{\mu \cdot \phi\left(F^{-1}\left(1 - \frac{w}{V}\right)\right) + w \cdot \frac{\phi\left(F^{-1}\left(1 - \frac{w}{V}\right)\right)}{v \cdot (n-1)}} \quad (26.41)$$

$$ms_i = \frac{1}{n} \quad (26.42)$$

The authors show that if  $n$  is sufficiently large, the above equilibrium always exists. In addition, they show that the optimal stocking quantity decreases with the purchase cost  $w$ , increases with the customer value  $V$  and is independent of the number of firms  $n$ . The optimal price, on the other hand, increases with  $V$  and decreases with  $n$  and is ambiguous with respect to  $w$ .

Dana (2001) extends Deneckere and Peck's (1995) model by making stocking quantities unobservable before customers visit a firm. As before, let  $ms_i$  denote the probability with which a random customer visits firm  $i$ . Customers visit only one firm. There is no loss in goodwill for shortages and leftovers can be disposed at zero cost. The author consider

two closely related models. In the first model (Bertrand model), firms commit to observable prices before they choose their stocking levels. In the second model (Cournot model), firms commit to their stocking levels before they choose their prices. Here, a firm's price acts as a 'signal' of the stocking level it chooses.

In the Bertrand model, taking prices and consumers' subgame-perfect equilibrium strategies  $(ms_1, \dots, ms_n)$  as given, firm  $i$  solves

$$\max_{Q_i} p_i \cdot \int_L^H \min(Q_i, ms_i \cdot d) \cdot f(x) \cdot dx - w \cdot Q_i \tag{26.43}$$

to determine the stocking quantity. The optimal stocking quantity  $Q^*(p_i, ms_i)$  is solved by the standard newsvendor condition given by  $F(Q^*((p_i ms_i)/ms_i)) = (p_i - w/p_i)$ . Consequently, it can be shown that each firm sets a price to maximize consumer surplus. That is,  $p^* = \arg \max_{p \geq w} (V - p) \cdot \phi(F^{-1}((p - w/p)))$ , where  $\phi(\cdot)$  is the fill rate as defined above.

In the Cournot model, it is assumed that customers conjecture that each firm has chosen the optimal stocking level given the firm's observed price and its competitors' equilibrium prices. Given this assumption, the author proves that there exists a unique symmetric pure-strategy equilibrium in which every firm offers a common price and consumers' equilibrium strategies satisfy  $ms_i = (1/n)$  (similar to the results of Deneckere and Peck's 1995 model.) The authors then show that the Cournot equilibrium price is always higher than the Bertrand price and that the difference depends on the number of firms. As the number of firms increases, the equilibrium price of the Cournot model converges to that of the Bertrand model. In both cases, it is shown that industry profits are strictly positive even when there are arbitrarily many firms.

#### 4. Dynamic pricing

In many situations, pricing decisions can be revised periodically in response to current information or market conditions. Static models that yield a single fixed price would not be adequate in providing guidance on how prices should be adjusted over time. In this section, we review several dynamic pricing models and highlight the managerial insights that they provide.

We first discuss the dynamic pricing model developed by Gallego and van Ryzin (1994). They consider a monopolist seller operating in a finite time horizon of length  $T$ . The seller has a finite inventory of  $N$  units to sell over the time horizon. The seller may adjust prices  $p(t)$  dynamically over time  $t \in [0, T]$ . Demand arrives according to a Poisson process with rate  $\lambda$ . Each arriving customer has i.i.d. (independent and identically distributed) valuations for the product that follow distribution  $F$ . Therefore sales occur (at prevailing prices) according to a variable-rate Poisson process with intensity  $\lambda(p) = \lambda(1 - F(p))$  dependent on the current price. In other words, during the small time interval  $[t, t + \varepsilon)$ , a customer arrives with probability  $\lambda\varepsilon$ , and given that the current price is  $p(t)$ , this arrival purchases with probability  $1 - F(p(t))$ . Units remaining at the end of the time horizon have no value. This model captures constraints in both inventory and time, and can be applied to travel industries (airlines and hotels), fashion retailing, as well as other seasonal or perishable items.

For this model, the seller's pricing problem can be formulated as follows. Let  $J(n, t)$  denote the value function, representing the seller's optimal continuation payoff at time

$t$  and with  $n$  units of inventory remaining. Consider a small time interval  $[t; t + \varepsilon)$ . The Hamilton–Jacob–Bellman (HJB) equation for this stochastic control problem can be written as

$$J^*(n, t) = \sup_p \{ \lambda(p)\varepsilon[p + J^*(n - 1, t + \varepsilon)] + (1 - \lambda(p)\varepsilon)J^*(n, t + \varepsilon) + o(\varepsilon) \} \quad (26.44)$$

Rearranging and taking limits, we obtain

$$-\frac{\partial J^*(n, t)}{\partial t} = \lim_{\varepsilon \rightarrow 0} \frac{J^*(n, t) - J^*(n, t + \varepsilon)}{\varepsilon} \quad (26.45)$$

$$= \sup_p \lambda(p) [p + J^*(n - 1, t) - J^*(n, t)] = 0 \quad (26.46)$$

where we have assumed regularity conditions (to interchange limits and supremums) and the last equality follows from the zero-derivative first-order condition. The boundary conditions are  $J^*(n, T) = 0$  (since remaining units have zero value) and  $J^*(0, t) = 0$  (there is nothing to sell). With these boundary conditions and the HJB equation, we can numerically compute the optimal price  $p^*(n, t)$  corresponding to having  $n$  units on hand at time  $t$ . However, for a general demand intensity  $\lambda(p)$ , the optimal price does not admit an explicit characterization. Nevertheless, there is an intuitive interpretation of equation (26.46). The rate of change of the value function  $J^*(n, t)$  is determined by two terms: the revenue accrued from consummating a sale at price  $p$ ; and a loss of  $J^*(n, t) - J^*(n - 1, t)$ , which can be interpreted as the option value of retaining the  $n$ th unit for sale in the future.

In their analysis, Gallego and van Ryzin provide additional results. They show that the optimal prices  $p^*(n, t)$  are decreasing in both  $n$  and  $t$ . Put differently, as the inventory level  $n$  increases, the optimal price drops; similarly, as we have less time to sell, the risk of having unsold units increases and thus the optimal price also falls. In addition, the authors consider a deterministic version of the problem. In this version, the instantaneous demand rate is now deterministic at  $\lambda(p)$ ; that is, given price  $p$ , units are sold at this constant rate. They demonstrate that for this deterministic problem, the optimal solution is to set a fixed price for the entire time interval. This optimal price is the maximum of  $p^*$  and  $p^0$ , where the  $p^*$  is the price that maximizes the revenue rate  $p\lambda(p)$  and  $p^0$  is the ‘run-out’ price under which all  $N$  units will be sold out at exactly time  $T$ , i.e.  $\lambda(p^0) = N/T$ . This result is intuitive because when there is sufficient inventory, charging  $p^*$  maximizes revenue, but when inventory is too low, it is preferable to sell all units at a higher price  $p^0$ . (Here, an assumption that the revenue function  $p\lambda(p)$  is quasi-concave is used.)

Let us now summarize the insights from the Gallego and van Ryzin model. First, optimal prices can be determined by assessing the tradeoff between sales at the current price and the option value of unsold units. Since this option value decreases with more inventory and also as time passes, optimal prices should also follow these trends. Second, the price dynamics in this model are driven primarily by demand uncertainty. In an analogous setup with deterministic demand, we see that a single fixed price is optimal. Therefore, this model is useful in isolating the price dynamics that are important in environments with high demand uncertainty as well as other operational considerations such as inventory and time horizon constraints.

Next, we turn to another class of dynamic pricing models. These models study intertemporal price discrimination by durable goods monopolists. The basic setup involves a monopolist firm selling a durable product to a fixed market of consumers with heterogeneous valuations. The monopolist's problem is to set prices optimally over time so that consumers are willing to buy. In particular, consumers form rational expectations over future prices and thus are not willing to buy if they anticipate more attractive purchase opportunities in the future. Therefore, while the previous class of models focuses on managing uncertainties in the demand process, intertemporal price discrimination models focus on the strategic interactions with rational consumers.

We first review the two-period model of Bulow (1982), which makes the analysis rather transparent. In this model, the monopolist faces demand curve of the form  $p = \alpha - \beta q$  and sells over two periods. In other words, if a quantity  $q_1$  is sold in the first period, the effective demand curve in the second period is  $p = (\alpha - \beta q_1) - \beta q$ , so the firm maximizes revenue from second-period sales by producing  $q_2 = (\alpha - \beta q_1)/(2\beta)$  units and selling them at price  $p_2 = (\alpha - \beta q_1)/2$ . Therefore rational consumers, upon observing that  $q_1$  units are sold in the first period, will expect the second-period price to fall to  $p_2$ . Now, the crucial step is to recognize that in order for  $q_1$  units to be sold in period 1, the price  $p_1$  must be chosen such that the marginal consumer (who has valuation  $\alpha - \beta q_1$ ) is indifferent between buying and waiting. In other words, assuming the discount factor  $\delta$ , we need

$$\alpha - \beta q_1 - p_1 = \delta(\alpha - \beta q_1 - p_2) = \delta(\alpha - \beta q_1)/2 \quad (26.47)$$

$$p_1 = (1 - \delta/2)(\alpha - \beta q_1). \quad (26.48)$$

Finally, we maximize the total revenue over both periods by solving

$$\text{Max}_{q_1, q_2} (1 - \delta/2)(\alpha - \beta q_1)q_1 + \delta(\alpha - \beta q_1 - \beta q_2)q_2 \quad (26.49)$$

The key to this type of analysis is to characterize the reservation prices of consumers who form rational expectations over future prices by predicting the monopolist's optimal actions. Given these reservation price constraints, the monopolist's dynamic pricing problem can then be formulated and solved. As an illustration, consider the following example with  $\alpha = \beta = \$1$  and  $\delta = 0.5$ . At the profit-maximizing solution, the monopolist sells  $q_1 = 0.4$  units at price  $p_1 = \$0.45$  in period 1 and sells  $q_2 = 0.3$  units at price  $p_2 = \$0.3$  in period 2. The total profit earned is \$0.225.

The two-period model above can be extended to infinite horizon settings. This was the setting considered by Coase (1972), who first proposed the durable goods monopoly problem. He conjectures that durability eliminates monopoly power because as long as prices remain above marginal cost, the monopolist will have the incentive to lower the price (to sell additional units) after some consumers have bought, so these consumers will not be willing to buy in the first place. Stokey (1981) solves the infinite-horizon pricing problem and characterizes the monopolist's optimal falling price path. In a related analysis, Stokey (1979) assumes that the monopolist commits to the temporal price schedule and finds that a single fixed price is optimal; this suggests that if the monopolist could commit, he would prefer not to price-discriminate over time. This point is evident from the numerical example above: if the monopolist could commit not to lower prices in

period 2, he essentially faces a single-period monopoly pricing problem, for which the optimal solution is to sell 0.5 units at price \$0.50, yielding profit \$0.25 (which is higher than \$0.225 above). Besanko and Winston (1990) isolate the effect of consumer rational expectations by comparing a model with strategic consumers (similar to Stokey, 1981) to a model with myopic consumers (in which consumers are not forward looking and purchase as soon as the price is below their valuations). They show that relative to the static monopoly price, prices are uniformly lower with strategic consumers and prices are uniformly higher with myopic consumers. In addition, prices start higher and fall faster when there are myopic consumers, as compared to strategic consumers.

A related model by Conlisk et al. (1984) studies intertemporal price discrimination when there is a continual influx of new consumers. (The models reviewed in the previous two paragraphs assume that all consumers are present at the start of the time horizon.) In some sense, this demand structure is similar to the customer arrival processes in the Gallego and van Ryzin setup, although customer inflows are deterministic here (i.e.  $N$  consumers enter the market each period). There are two customer types: a fraction  $\alpha$  are high-types with valuation  $V_H$  and the rest are low-types with valuation  $V_L$ . The discount factor is  $\beta$  per period. Consumer rational expectations and heterogeneous valuations continue to play a major role. In this environment, the authors show that the optimal solution involves cyclic pricing. Each price cycle, of duration  $n$  periods, is characterized by

$$p_j = (1 - \beta^{n-j}) V_H + \beta^{n-j} V_L \quad (26.50)$$

for  $j = 1, \dots, n$ . There are periodic ‘sales’ (when  $j = n$ ) through which the monopolist ‘harvests’ the low-valuation consumers that have accumulated in the market; prices subsequently return to the ‘regular price’ level at the start of each cycle ( $j = 1$ ) and gradually decline until the next sale at the end of the cycle ( $j = n$ ). In each price cycle, the price path is chosen so that high-valuation customers are willing to buy rather than wait for the sale; thus prices are higher the further away the anticipated sale is. Time discounting is quite important in driving the price cycles in this model. Another important element is the assumption that the monopolist is unable to commit to future prices. Sobel (1991) analyzes the case in which commitment is feasible, and he shows that the seller’s optimal solution is to commit to a fixed price, similar to the above case where there is a fixed market of consumers. This suggests that commitment power diminishes the usefulness of dynamic pricing when selling to strategic consumers.

Observe that the durable goods monopoly models described above do not involve inventory or time constraints. The firm is able to sell as many units as demanded at each price. Moreover, most models have infinite time horizons. This is in contrast to the Gallego and van Ryzin setup, in which there is a finite inventory to sell over a limited time period, so each unsold unit has a dynamically evolving option value that shapes the optimal prices. The stark difference between these two strands of work motivates a third class of models, which incorporates both perspectives into a single framework. These relatively more recent models contain two main ingredients: operational constraints (i.e. in time and inventory) as well as strategic constraints (i.e. consumer rational expectations and sequential rationality). Aviv and Pazgal (2008) develop a model with Poisson customer arrivals and analyze the optimal timing of a single price discount. Su (2007) uses a simpler deterministic demand structure to study the firm’s dynamic pricing problem.

We review the basic model by Su (2007). The setup is based on the deterministic setting in Gallego and van Ryzin, where there is a finite inventory  $N$  to sell over a finite time horizon  $T$ . This puts the operational constraints in place. Next, the main ingredient of durable goods monopoly models is added: strategic customers, who arrive according to a deterministic flow process, form rational expectations of future prices and optimally choose between buying versus waiting. This behavior generates incentive constraints that prices must satisfy in order to induce purchases. The derivation of these incentive constraints is similar in spirit to the analysis in (26.47)–(26.48). With both operational and incentive constraints in place, the firm's pricing problem can then be formulated and the optimal prices characterized. In this model, there is a mixture of strategic customers (with rational expectations) and myopic customers (who purchase as soon as prices are below their reservation values). Moreover, consumers may have high valuations ( $V_H$ ) or low valuations ( $V_L$ ). This creates four consumer segments: strategic-highs, strategic-lows, myopic-highs and myopic-lows. Depending on the relative sizes of these four segments, the optimal price path may take one of two forms: (1) prices start at  $V_L$  and jump to  $V_H$  at some point during the selling season; or (2) prices start high at  $V_H$  and drop to an intermediate price  $p \in (V_L, V_H)$  before the end of the season, when prices fall to  $V_L$ . These results can also be extended to incorporate time discounting in the form of waiting costs (i.e. customers may face waiting costs when delaying purchases).

Two surprising results emerge from this model. First, note that a remarkably robust finding from the stream of revenue management literature following Gallego and van Ryzin is that optimal prices (on average) tend to fall as time passes. This is intuitive, given that the option value of unsold units declines over time. However, once strategic customer behavior is added to the picture, the optimal price path may either increase or decrease over time. This endogenous structure of optimal price paths depends on the composition of the customer population, and in particular, on the correlation between strategic behavior and reservation prices. The main result is that, when strategic customers have higher reservation prices than myopic customers, optimal prices increase over time. However, in the reverse case, decreasing prices (e.g. markdowns) serve as an intertemporal price discrimination device because the seller is able to extract higher revenues from myopic customers who do not wait. This suggests that option value considerations (which generate declining price paths) are no longer dominant when there are strategic interactions in the marketplace. Second, a common finding in the durable goods monopoly literature is that strategic customers with rational expectations hurt seller profits. In the extreme case of the Coase conjecture, monopoly profits are completely eroded. However, by incorporating operational constraints, the model demonstrates that strategic behavior may benefit the seller. This is because low-valuation customers, by competing with high-valuation customers for product availability, may increase their willingness to pay. This effect is driven by the operational constraints, because otherwise there would be no notion of 'product availability'. In essence, with limited inventory and limited time, rational consumers need not only consider future prices, but also future availability.

## **5. Queuing models**

For many situations, queuing models provide a useful way to capture capacity limitations. We begin this section with a description of a standard textbook queuing model, and then review the classic work that incorporates pricing effects into queuing models.

Consider a single-server queue facing a stream of customers who arrive over time according to a Poisson process of rate  $\lambda$ . Service is rendered in a first-come-first-served basis, and service times for each customer are independently, identically and exponentially distributed with rate  $\mu$ . We denote  $\rho = \lambda/\mu$  and assume  $\rho < 1$ . It is well known that the average time spent in the queue, including service time, is  $1/(\mu - \lambda)$ . Therefore we may interpret the customer arrival rate  $\lambda$  as demand and the service rate  $\mu$  as the capacity of the queue. Then, increasing demand generates congestion effects that lead to longer waits for all customers. Observe that queueing models exhibit ‘soft’ capacity constraints in the sense that all customers will eventually be served. This is in contrast to operations models with finite inventory, in which some customers may not obtain the product in the event of a stock-out, so there are ‘hard’ capacity constraints. For this reason, queueing models are often applied to service contexts where the consequences of capacity limitations are more subtle. The waiting time that customers face as a result of queueing capacity  $\mu$  may also be interpreted as a degradation in other dimensions of service quality.

The classic model that considers pricing in queueing models is proposed by Naor (1969). The fundamental premise is that customers, upon arrival and observing the current state of the queue, may choose whether to join. Customers earn a reward of  $R$  upon completion of service at the queueing station but incur waiting costs at a rate of  $C$  per unit time. Thus, if a customer arrives at a queue with  $n$  other customers in line, the expected payoff from joining is  $R - nC/\mu$  (compared to zero from departing). This implies that the customer leaves the queue whenever the queue length exceeds a particular threshold,  $k = \lfloor R\mu/C \rfloor$ . Now, suppose that the server charges some price  $p$ . By the same logic, customers will now adopt the threshold strategy

$$k(p) = \lfloor (R - p)\mu/C \rfloor \quad (26.51)$$

i.e. join if  $n \leq k(p)$  and leave if  $n > k(p)$ . This sets up the framework to study pricing effects in queues. Given a particular price  $p$ , we can characterize the resulting demand pattern. Consider the following numerical example. Suppose that customers arrive to a coffee stand at a rate of  $\lambda = 1$  customer per minute, and that they can be served at a rate of  $\mu = 2$  customers per minute. Further, customers value their coffee at  $R = \$10$  but assess waiting costs to be  $C = \$1$  per minute. Then, we see from equation (26.51) that customers are willing to wait in line for free coffee (i.e. when the price is  $p = 0$ ) as long as there are no more than  $k = 20$  customers in line. However, when the price  $p = \$5$  is charged, customers are no longer willing to wait in line when there are more than  $k = 10$  people waiting.

From a revenue maximization perspective, the queue manager should optimally balance the tradeoff between a high price and the consequent reduction in demand. Corresponding to price  $p$ , customers join the queue only when the queue length is less than or equal to threshold  $k(p)$ . From queueing theory (see, e.g., Gross and Harris, 1998; Wolff, 1989), the probability that a customer joins the queue is  $(1 - \rho^{k(p)})/(1 - \rho^{k(p)+1})$ . In other words, the demand function is

$$D(p) = \lambda \frac{1 - \rho^{k(p)}}{1 - \rho^{k(p)+1}}. \quad (26.52)$$

The manager should charge the optimal price  $p^*$  that maximizes the expected revenue rate  $pD(p)$ . This induces the customer threshold strategy  $k^* = k(p^*)$  that maximizes the firm's revenue.

Alternatively, from a social planner perspective, we may write down the social welfare function as

$$SW(p) = RD(p) - CN(p) \tag{26.53}$$

$$= \lambda R \frac{1 - \rho^{k(p)}}{1 - \rho^{k(p)+1}} - C \left[ \frac{\rho}{1 - \rho} - \frac{(k(\rho) + 1)\rho^{k(p)+1}}{1 - \rho^{k(p)+1}} \right] \tag{26.54}$$

The first term is the rate at which reward is earned, and the second term is the rate at which waiting cost is incurred; here,  $N(p)$  denotes the average number of customers in the queue at any point in time and can be expressed in terms of  $k(p)$  as shown above. Therefore the first-best can be attained if the firm charges the price that maximizes social welfare  $SW(p)$ . Let us denote the first-best price as  $p^{FB}$  and the resulting customer threshold as  $k^{FB} = k(p^{FB})$ .

Apart from laying out the framework above, Naor (1969) also provides the following comparative results. He shows that

$$k^* < k^{FB} < k_0 \tag{26.55}$$

where  $k_0$  denotes customers' queue-joining threshold strategy in the absence of prices (i.e.  $p = 0$ ). Analogously, the following also holds:

$$p^* > p^{FB} > 0. \tag{26.56}$$

Using our numerical example above, we find that the revenue-maximizing price is  $p^* = \$8.50$  while the socially efficient price is  $p^{FB} = \$5$ . Equivalently, the revenue-maximizing and first-best queue-joining thresholds are  $k^* = 3$  and  $k^{FB} = 10$ .

There are two key insights. First, revenue maximization leads to prices above the socially efficient level, and second, achieving first-best requires positive prices (above marginal cost). The first is consistent with standard monopoly pricing models, but the second is not. This is due to the negative congestion externality that is present in queueing models. Customers make their joining decisions in self-interest even though their current decisions will influence the well-being of future arrivals. In this situation, pricing can be used to address such externalities and attain first-best.

Thus far, we have assumed that customers are served in a first-come-first-served order. In an interesting analysis, Hassin (1985) considers the opposite extreme of last-come-first-served priority rules. In this case, since all future arrivals will be placed in front of current customers, these current customers will take them into consideration. The consequence is that equilibrium-joining behavior will be socially optimal even in the absence of pricing. However, Hassin points out a strategic difficulty involved with last-come-first-served queues. Now, customers have the incentive to leave and re-enter the join, presumably disguised as a new arrival. In the analysis, such behavior is assumed away, but in practice, substantial monitoring may be required. Certainly, there are also equity and fairness issues that have not been accounted for.

This modeling framework has been shown to be robust along several dimensions. The restriction to customer threshold joining strategies is an important simplification but currently holds only under the assumption of Poisson customer arrivals. Yechiali (1971) extends this setup to general arrival processes and shows that threshold-type policies is without loss of generality. In another paper, Yechiali (1972) extends the analysis to multi-server systems. While we have considered only linear waiting costs and linear rewards, Knudsen (1972) analyzes general nonlinear cost and reward structures and shows that the basic insights still hold. Lippman and Stidham (1977) introduce discounting and also consider finite time horizons; they also find that the structure of the results remains unchanged. This line of research demonstrates that for economic and managerial analysis, it is usually sufficient to focus on simple models, such as single-server exponential systems. Nevertheless, even for such ‘simple’ models, there is usually a high degree of technical complexity involved. This is because for most dynamic queueing processes, characterizing performance measures (such as waiting time and number of customers in queue) is not an easy task.

Another modeling approach is to consider a static steady-state analysis of queueing systems. Here, we review the framework proposed by Mendelson (1985). The starting point is a value function  $V(\lambda)$ , which represents the total value of performing the service when the aggregate arrival rate is  $\lambda$ . We assume that  $V$  is concave, as this captures the decreasing marginal value of each additional unit of customers served. In other words, the value of service to the marginal customer is  $V'(\lambda)$ . Apart from the service rewards, there are also waiting costs due to capacity constraints. Letting  $W(\lambda)$  denote the average wait time and  $C$  denote the waiting cost (per unit time), we see that corresponding to arrival rate  $\lambda$ , the waiting cost incurred by each customer is  $CW(\lambda)$ . Therefore, in the absence of pricing, equilibrium arrival rates chosen by the customer population satisfies

$$V'(\lambda) = CW(\lambda) \quad (26.57)$$

Similarly, when each customer faces an admission fee  $p$ , the equilibrium arrival rate is  $\lambda(p)$  given implicitly by

$$V'(\lambda(p)) = p + CW(\lambda(p)) \quad (26.58)$$

We stress that for any price  $p$ , the equilibrium is unique since  $V$  is concave in  $\lambda$  (so  $V'$  is decreasing in  $\lambda$ ) and  $W$  is increasing in  $\lambda$ . Therefore, there is a one-to-one relationship between prices and equilibrium arrival rates, and we may use  $p(\lambda)$  to denote the price that can be used to induce arrival rate  $\lambda$ . In this setup, the implicit assumption is that customers do not observe queue lengths when making joining decisions, but instead base their decisions on the expected steady-state queue lengths. This simplifies the analysis since customer decisions no longer dynamically depend on the evolution of the stochastic queueing system.

Using this modeling approach, we may proceed to study profit-maximizing pricing schemes and compare them to the first-best. For a given price  $p$ , the firm’s revenue rate is

$$\Pi(p) = \lambda(p) \cdot p = \lambda(p) \cdot [V'(\lambda(p)) - CW(\lambda(p))] \quad (26.59)$$

In terms of  $\lambda$ , we have

$$\Pi(\lambda) = [V'(\lambda) - CW(\lambda)]. \quad (26.60)$$

Therefore, we may maximize this expression to find the profit-maximizing arrival rate  $\lambda^*$ . The firm's optimal price is then given by  $p^* = p(\lambda^*)$ . Similarly, we now characterize the first-best outcome. In terms of arrival rate  $\lambda$ , the social welfare function is

$$SW(\lambda) = V(\lambda) - \lambda CW(\lambda) \quad (26.61)$$

Maximizing this expression over  $\lambda$ , we obtain the first-best arrival rate  $\lambda^{FB}$ . This can be sustained in equilibrium by imposing the first-best price  $p^{FB} = p(\lambda^{FB})$ . Consistent with Naor's model, this setup also yields  $p^* > p^{FB} > 0$ .

This modeling framework has been extended to incorporate multiple customer classes. In Mendelson and Whang (1990), there are multiple customer classes, each with different value functions  $V_i$  and waiting costs  $C_i$ . Customer classes are unobservable, so this is a hidden information problem. The authors analyze a priority pricing mechanism; that is, paying different prices corresponds to receiving different priorities in the queue and thus incurring different waiting times. The pricing mechanism can be designed to be incentive compatible and socially optimal. This analysis highlights an interesting feature of queueing models: with just a single server (producing a single good), the addition of priorities essentially introduces multiple different goods, which can be used to price discriminate amongst different customer classes. Subsequently, Lederer and Li (1997) extend this analysis to a competitive setting. In another paper, Van Mieghem (2000) introduces methodology to treat the case of convex delay costs (rather than linear delay costs assumed above). This modeling framework is quite flexible and can be extended to include another dimension of choice by customers: apart from choosing whether to join the queue, customers may also choose how much service to request. In some sense, this resembles a quantity decision. Ha (2001) studies this scenario and derives optimal incentive-compatible pricing mechanisms.

## 6. Conclusions

In this chapter, we introduce four broad areas of research in operations management that relate to pricing. A central theme that cuts across all areas is that customers are active and strategic, and they maximize their utility by choosing an appropriate buying and/or operational action. In reviewing each area, we first describe a classic model in operations management and show how subsequent research extends these standard models. Our review is deliberately selective because we want to show how research and model development accumulates in the literature. Our primary goal is to expose marketing and economic researchers to the rapidly growing areas of research in operations management that relate to pricing.

Table 26.1 summarizes the main findings and insights in this chapter when consumers are strategic and actively engage in operational decision-making. In the EOQ inventory models, we show that pricing variability leads to higher shopping frequency and smaller average purchase quantities. Promotions can serve as an effective vehicle to transfer inventory-holding costs from the seller to consumers, and to price-discriminate between

Table 26.1 Summary of results and insights when buyers are customers

Area	Key operational consideration(s)	Consumer decisions	Role of pricing	Applications
EOQ inventory models	<ul style="list-style-type: none"> <li>holding cost</li> <li>fixed cost</li> </ul>	<ul style="list-style-type: none"> <li>how much to buy</li> </ul>	<ul style="list-style-type: none"> <li>Price variance induces consumers to shop more frequently at smaller quantities.</li> <li>Deals transfer holding costs to consumers who stockpile.</li> <li>Periodic promotions allow firms to price-discriminate between consumers with high and low holding costs.</li> </ul>	<ul style="list-style-type: none"> <li>Food products</li> <li>Household items</li> </ul>
Newsvendor inventory models	<ul style="list-style-type: none"> <li>availability</li> </ul>	<ul style="list-style-type: none"> <li>whether to visit the store</li> </ul>	<ul style="list-style-type: none"> <li>Low prices attract consumers to visit the store (and face the risk of stock-outs).</li> <li>High prices signal high availability.</li> </ul>	<ul style="list-style-type: none"> <li>Fashion</li> <li>Electronics</li> </ul>
Dynamic pricing models	<ul style="list-style-type: none"> <li>limited capacity</li> <li>limited time</li> </ul>	<ul style="list-style-type: none"> <li>when to buy</li> </ul>	<ul style="list-style-type: none"> <li>Stable prices discourages strategic timing of purchases and increases firm profits.</li> <li>Prices should be adjusted dynamically to reflect the option value of unsold units.</li> <li>Dynamic pricing serves as a price discrimination device when consumers have different propensities to wait.</li> </ul>	<ul style="list-style-type: none"> <li>Airlines</li> <li>Hotels</li> <li>Fashion</li> </ul>
Queueing models	<ul style="list-style-type: none"> <li>congestion</li> </ul>	<ul style="list-style-type: none"> <li>whether to join the queue</li> </ul>	<ul style="list-style-type: none"> <li>The price that attains first-best is higher than the marginal cost; at this price, existing customers will consider the externality they impose on future arrivals.</li> <li>The firm may price-discriminate by establishing priorities and charging different prices.</li> </ul>	<ul style="list-style-type: none"> <li>Services</li> </ul>

different consumers. In the newsvendor inventory models, we show that low prices attract more store visits by consumers while high prices signal high product availability. In the dynamic pricing models, it is shown that stable prices increase profits because they discourage strategic timing of purchases. Dynamic pricing can also be effective if consumers have different propensities to wait. Another consideration is that prices should be adjusted dynamically to reflect the option value of unsold units over the selling horizon. In the queuing models, we show that pricing above marginal cost induces customers to consider the externality they impose on future customer arrivals, and that firms can price-discriminate by establishing service priorities.

Besides making the problem contexts more realistic and richer, operational considerations often influence the optimal price of the firm significantly. Also, these considerations frequently generate more realistic equilibrium outcomes in competitive settings. While they are typically accompanied by more challenging analyses, the payoffs seem worthwhile because we begin to see an accumulation of knowledge and insights. The current approaches make a major step forward by focusing on making customers active (i.e. or in game-theoretic terms, they are players in the model). This is accomplished by making them more strategic and rational. Clearly, we do not need to restrict to these standard assumptions. In fact, research in psychology and experimental economics suggests that these assumptions are routinely violated even when customers are motivated by substantial monetary incentives.

A promising and perhaps more radical approach is to assume that active customers are boundedly rational. One can extend the equilibrium analysis to include situations where mistakes are allowed but in the way that more costly mistakes are made less frequently than less costly mistakes (see McKelvey and Palfrey, 1995), and where a lack of rational expectation in belief formation among players is possible (see Camerer et al., 2004). Also, consumers care both about the final outcomes as well as the changes in outcomes with respect to a target outcome, they are impatient in that they prefer instant gratification, and they care about being treated fairly (see Ho et al., 2006 for a comprehensive review).

## References

- Assunção, J.L. and R.J. Meyer (1993), 'The rational effect of price promotions on sales and consumption', *Management Science*, **39** (5), 517–35.
- Aviv, Y. and A. Pazgal (2008), 'Optimal pricing of seasonal products in the presence of forward-looking consumers', *Manufacturing and Service Operations Management*, **10** (3), 339–59.
- Bell, D.R., G. Iyer and V. Padmanabhan (2002), 'Price competition under stockpiling and flexible consumption', *Journal of Marketing Research*, **39** (3), 292–303.
- Besanko, D. and W.L. Winston (1990), 'Optimal price skimming by a monopolist facing rational consumers', *Management Science*, **36** (5), 555–67.
- Blattberg, R.C., G.D. Eppen and J. Lieberman (1981), 'A theoretical and empirical evaluation of price deals for consumer nondurables', *Journal of Marketing*, **45** (1), 116–29.
- Bulow, J.I. (1982), 'Durable-goods monopolists', *Journal of Political Economy*, **90** (2), 314–32.
- Camerer, C.F., T.H. Ho and J.K. Chong (2004), 'A cognitive hierarchy model of games', *Quarterly Journal of Economics*, **119** (3), 861–98.
- Coase, R.H. (1972), 'Durability and monopoly', *Journal of Law & Economics*, **15** (1), 143–49.
- Conlisk J., E. Gerstner and J. Sobel (1984), 'Cyclic pricing by a durable goods monopolist', *Quarterly Journal of Economics*, **99** (3), 489–505.
- Dana, J.D. (2001), 'Competition in price and availability when availability is unobservable', *RAND Journal of Economics*, **32** (3), 497–513.
- Dana, J.D. and N.C. Petruzzi (2001), 'The newsvendor model with endogenous demand', *Management Science*, **47** (11), 1488–97.

- Deneckere, R. and J. Peck (1995), 'Competition over price and service rate when demand is stochastic: a strategic analysis', *RAND Journal of Economics*, **26** (1), 148–62.
- Gallego, G. and G. van Ryzin (1994), 'Optimal dynamic pricing of inventories with stochastic demand over finite horizons', *Management Science*, **40** (8), 999–1020.
- Gross, D. and C.M. Harris (1989), *Fundamentals of Queueing Theory*, 3rd edn, New York: John Wiley & Sons.
- Ha, A.Y. (2001), 'Optimal pricing that coordinates queues with customer-chosen service requirements', *Management Science*, **47** (7), 915–30.
- Hassin, R. (1985), 'On the optimality of first come last served queues', *Econometrica*, **53** (3), 201–2.
- Ho, T.H., C.S. Tang and D.R. Bell (1998), 'Rational shopping behavior and the option value of variable pricing', *Management Science*, **44** (12), S145–S160.
- Ho, T.H., N. Lim and C.F. Camerer (2006), 'Modeling the psychology of firm and consumer behavior with behavioral economics', *Journal of Marketing Research*, **43**, 307–31.
- Jeuland, A.P. and C. Narasimhan (1985), 'Dealing – temporary price cuts – by seller as a buyer discrimination mechanism', *Journal of Business*, **58** (3), 295–308.
- Knudsen, N.C. (1972), 'Individual and social optimization in a multiserver queue with a general cost–benefit structure', *Econometrica*, **40** (3), 515–28.
- Kunreuther, H. and J.F. Richard (1971), 'Optimal pricing and inventory decisions for non-seasonal items', *Econometrica*, **39** (1), 173–5.
- Lederer, P.J. and L. Li (1997), 'Pricing, production, scheduling, and delivery-time competition', *Operations Research*, **45** (3), 407–20.
- Lippman, S.A. and S. Stidham (1977), 'Individual versus social optimization in exponential congestion systems', *Operations Research*, **25** (2), 233–47.
- McKelvey, R.D. and T.R. Palfrey (1995), 'Quantal response equilibria for normal form games', *Games and Economic Behavior*, **10**, 6–38.
- Mendelson, H. (1985), 'Pricing computer services: queueing effects', *Communications of the ACM*, **28** (3), 312–21.
- Mendelson, H. and S. Whang (1990), 'Optimal incentive-compatible priority pricing for the M/M/1 queue', *Operations Research*, **38** (5), 870–83.
- Naor, P. (1969), 'The regulation of queue size by levying tolls', *Econometrica*, **37** (1), 15–24.
- Petruzzi, N. and M. Dada (1999), 'Pricing and the newsvendor problem: a review with extensions', *Operations Research*, **47**, 183–94.
- Porteus, E.L. (2002), *Foundations of Stochastic Inventory Theory*, Stanford, CA: Stanford University Press.
- Sobel, J. (1991), 'Durable goods monopoly with entry of new consumers', *Econometrica*, **59** (5), 1455–85.
- Stokey, N.L. (1979), 'Intertemporal price discrimination', *Quarterly Journal of Economics*, **93** (3), 355–71.
- Stokey, N.L. (1981), 'Rational expectations and durable goods pricing', *Bell Journal of Economics*, **12** (1), 112–28.
- Su, X. (2007), 'Intertemporal pricing with strategic customer behavior', *Management Science*, **53** (5), 726–41.
- Su, X. and F. Zhang (2008), 'Strategic customer behavior, commitment, and supply chain performance', *Management Science*, **54** (10), 1759–73.
- Van Mieghem, J.A. (2000), 'Price and service discrimination in queueing systems: incentive compatibility of Gc scheduling', *Management Science*, **46** (9), 1249–67.
- Wolff, R.W. (1989), *Stochastic Modeling and the Theory of Queues*, Englewood Cliffs, NJ: Prentice Hall.
- Yechiali, U. (1971), 'On optimal balking rules and toll charges in the GI/M/1 queueing process', *Operations Research*, **19** (2), 349–70.
- Yechiali, U. (1972), 'Customers' optimal joining rules for the GI/M/s queue', *Management Science*, **18** (7), 434–43.