Quick Response in Manufacturer-Retailer Channels

Ananth. V. Iyer • Mark E. Bergen
Kraner Graduate School of Management, 1310 Kraner Building, Purdue University,
West Lafayette, Indiana 47907-1310
Carlson School of Management, Department of Marketing and Logistics Management, 271 19th Avenue South,
University of Minnesota, Minneapolis, Minnesota 55455-0413

Quick Response (QR) is a movement in the apparel industry to shorten lead time. Under QR, the retailer has the ability to adjust orders based on better demand information. We study how a manufacturer-retailer channel impacts choices of production and marketing variables under QR in the apparel industry. Specifically, we build formal models of the inventory decisions of manufacturers and retailers both before and after QR. Our models allow us to address who wins and who loses under QR, and suggest actions such as service level, wholesale price and volume commitments that can be used to make QR profitable for both members of the channel, i.e., Pareto improving. Detailed discussions with a major retailer, and information from industry sources provide supporting evidence for the structure and conclusions of the model.

(Inventory; Channels, Lead Time, Service Level, Bayesian Models)

1. Introduction

Traditionally, ordering fashion merchandise in the apparel industry has been characterized by long lead times between order placement by the retailer and delivery by the manufacturer. Specific lead times faced by retailers are reported to be seven months for Oxford shirts ordered by J.C. Penney (Skinner 1992), eight months for Jaymar Ruby slacks (Hammond 1991), five months for Liz Claiborne apparel (Dalby and Flaherty 1990) and five months for Benetton apparel (Signorelli and Heskett 1984). With long lead times, retailers have to place orders with manufacturers regarding individual items long before demand is actually realized. As a result, the retailer can have either: (i) too little inventory, which results in product stockouts and low service levels, or (ii) too much inventory, resulting in forced markdowns, disposal costs, or expediting costs. The magnitude of the industry-wide impact of these costs is estimated to be around 25 percent of sales (Frazier 1986).

An industry response to this problem is Quick Response (QR), a strategy focusing on providing shorter lead times. QR enables orders to be placed closer to the start of the selling season. We focus on QR for fashion products—products with a short selling season for which only one order is placed before the start of the season with no opportunity for reorder. Typically, for fashion products, QR lowers lead time by one to four months (Skinner 1992). The essential benefit from QR for fashion products is that information gathered regarding sales of related items can be used to decrease forecast error. The magnitude of this reduction in forecast error can be substantial. For example, reducing lead time for ordering from eight months to four months was estimated to decrease forecast error from 65% to 35% (Crafted With Pride, Inc 1990). Similarly, Blackburn...
(1991) reports forecast error reductions from 40% to 20% as lead time is decreased from six months to four months.

Inspection of the apparel industry indicates that it is extremely fragmented, with more than 15,000 apparel manufacturers and an even larger number of apparel retailers. Manufacturers and retailers are often separate companies, with individual goals and incentives. The literature on channels of distribution suggests that the individual incentives of manufacturers and retailers can have large effects on wholesale prices, service levels, contractual terms, etc. required to coordinate the channels members (Bergen et al. 1992; Stern and El-Ansary 1988). To understand the impact of this channel on QR, we build formal models of the inventory decisions and profits of manufacturers and retailers both before and after QR. An extensive bibliography of the inventory literature for style goods is provided in Eppen and Iyer (1995).

We show that QR may not always make both manufacturers and retailers better off. In particular, we show that manufacturers may not be better off under QR. Thus, manufacturers and retailers may be required to undertake actions to make QR Pareto improving. An action makes QR Pareto improving if both parties are at least as well off, and one party is better off than before QR. We study a wide variety of actions that can make QR Pareto improving, ranging from commitments to higher service levels or higher wholesale prices for single products to volume commitments across products.

We use three sources of information to provide supporting evidence for the structure and conclusions of the model. Our best source of primary data is derived from extensive discussions with retailer A, a national retailer. We talked to people at all levels of the company, from the senior vice president of apparel to QR coordinators to buyers of individual products. Second, we contacted QR managers from 10 other apparel retailers and discussed our models and their conclusions. Third, we used a variety of secondary data sources including industry reports, the proceedings of the Conferences on QR from 1992 through 1995, and academic sources.

2. Model Structure

In the old system, orders are placed at time 0 by the retailer. These orders are delivered $L_1$ units of time later to the retailer for sale in the season beginning at time $L$. In the QR system, at time 0 the retailer starts collecting information regarding sales of related products. At $L_1 (\leq L)$ the retailer places an order with the manufacturer. The manufacturer produces the order for delivery $L_2 = L - L_1$ units of time later for sale during the season beginning at time $L$. The values of $L_1$ are reported to range from five to eight months. The values of $L_1$ range from two to five months.

2.1. Demand Uncertainty Structure

In our model we have two levels of demand uncertainty. The first source of uncertainty concerns inherent demand uncertainty about the product, i.e., even if we know the mean demand $\theta$, the demand during the season is normally distributed with mean $\theta$ and variance $\sigma^2$, i.e., $f(x|\theta) \sim N(\theta, \sigma^2)$. This model captures uncertainty regarding the number of people who come to the store, their propensity to buy, etc. The second source of uncertainty models uncertainty about the demand at time 0. Information regarding $\theta$ at time 0 is modeled as a Normal distribution with mean $\mu$ and variance $\tau^2$, i.e., $g(\theta) \sim N(\mu, \tau^2)$. Thus, at time 0, our model of demand is a Normal distribution with mean $\mu$ and variance $\sigma^2 + \tau^2$ (Berger 1985), i.e., $m(x) \sim N(\mu, \sigma^2 + \tau^2)$.

The lower lead time under QR enables data collected during $L_1$ regarding sales of related items to be used to decrease forecast error for the item being ordered. Consider the practices of retailer A in forecasting demand for women’s blouses for the Spring 1996 season. Retailer A stores extensive Point of Sale data that can be analyzed along any one of over 1,000 possible attributes, such as color, silhouette, pattern, fabric, buttons or zippers to name a few. Thus, during the summer and fall seasons in 1995, a fashion buyer can access information along any particular attribute under consideration, for instance the color red. The buyer examines data regarding sales of other items that share the same color red over the last one to two months and can compare that to recent sales in the previous few months, or compare them to sales in the previous year. These comparisons enable the buyer to assess whether the color red exhibits a fashion trend for the upcoming Spring 1996 season. Similar analyses are undertaken for other attributes, including patterns, silhouettes, and styles. The buyers at retailer A stressed that this capability to learn about
trends from related items was a key component for deriving the benefit of QR for fashion items. This example is representative of buying for fashion items across retailers. All ten QR managers we contacted agreed that information gathered from related items closer to the start of the season reduced forecast error. In fact, the standard training for fashion buyers requires that historical data be supplemented with current trend analysis at the style, color, fabric and styling level including dress styles as well as fashion magazine trend proclamations, etc. (Drew 1992).

Data collected during $L_1$ is converted by the buyer into an estimate of the demand for the item under consideration, i.e., $d_1$. We note, that because the data is generated from similar (but not the same) items the remaining uncertainty, $\sigma^2$, reflects uncertainty both due to inherent demand uncertainty as well as due to the use of similar but not the same items for error reduction before the start of the season. This $d_1$ can be used to generate a posterior distribution for demand. Thus, given $d_1$, the conditional distribution of demand during the season is $g(\theta | d_1) \sim N(\mu(d_1), 1/\rho)$, where

$$\rho = \frac{1}{\sigma^2} + \frac{1}{\tau^2} \quad \text{and} \quad \mu(d_1) = \frac{\sigma \mu}{\sigma^2 + \tau^2} + \frac{\tau^2 d_1}{\sigma^2 + \tau^2}$$

This implies that $m(x | d_1) \sim N(\mu(d_1), \sigma^2 + (1/\rho))$ (Berger 1985).

22. Costs and Revenue Parameters

The retailer faces the following costs: a cost of $c$ per unit to purchase the product from the manufacturer, a goodwill cost of $\pi$ per unit of demand not satisfied during the season, a holding cost of $h$ per unit of product left over at the end of the season. The manufacturer experiences a cost per unit of $\pi$ to produce a unit of product. Each unit of product sold by the retailer to customers generates a revenue of $r$ per unit.

3. QR and the Channel

We now generate the inventory levels and expected profits of both the manufacturer and the retailer under the old system without QR. We then show the inventory levels and expected profits for both parties under a QR system.

3.1. The Old System

In the old system, the inventory level has to be chosen at time $0$. The retailer chooses an inventory level, $Q$, to maximize the following expected profit which is

$$\int_{x_0}^{Q} r(x) m(x) dx + \int_{Q}^{\infty} r(x) m(x) dx - h \int_{x_0}^{Q} (Q - x) m(x) dx - \pi \int_{x_0}^{Q} (x - Q) m(x) dx - c Q$$

Since this is the standard Newsboy model, it can be verified that this function is concave in $Q$ and the optimum initial inventory is

$$l_{old, opt} = \mu + Z(s) \sqrt{\sigma^2 + \tau^2}$$

where the optimal service level ($s$) is defined as $(r + \pi - c)/(r + \pi + h)$ and $Z(s)$ is the $Z$ value of a standard Normal distribution that generates a cumulative probability of $s$. The associated maximum expected profit in the Old System is

$$EP_{old, opt} = \mu - (\sigma + h) Z(s)$$

$$+ (r + h + \pi) b_r(Z(s)) \sqrt{\sigma^2 + \tau^2}$$

where $b_r(Z(s))$ is the right linear loss function of a standard Normal distribution at $Z(s)$. The expected quantity sold during the season in the old system ($l_{old, sold}$) is

$$l_{old, sold} = \int_{x_0}^{Q} x m(x) dx + \int_{x_0}^{\infty} l_{old, opt} m(x) dx$$

which simplifies to
\[ I_{\text{old}} = \mu - b_i(Z(s))\sqrt{\sigma^2 + \tau^2}. \]

The expected quantity left over at the end of the season in the old system is

\[ I_{\text{old-left}} = \int_{x}^{I_{\text{old}}} (I_{\text{old}} - x)m(x)dx \]

\[ = (Z(s) + b_i(Z(s)))\sqrt{\sigma^2 + \tau^2}. \]

3.2. The QR System

In the QR System, the retailer’s inventory choice involves two steps:

1. Observe the demand for related items during \( L_1 \) and use it to generate an estimate for the item under consideration, i.e., \( d_i \). Use that demand to generate the posterior distribution.

2. Choose the optimal inventory to maximize expected profits given the posterior distribution. This would imply different inventory levels depending on \( d_i \). Intuitively, if the retailer estimates a low \( d_i \) then he would be more likely to choose low inventory levels because he expects a low probability of high demand in the season.

It can be verified (details in the Appendix) that the expected quantity ordered by the retailer under QR, the corresponding expected profit, the expected quantity sold and the expected quantity left over are as follows:

\[ \text{EL}_{\text{QR}} = \mu + Z(s)\sqrt{\frac{\sigma^2}{\rho} + \frac{1}{\rho}}, \]  

\[ \text{EP}_{\text{QR}} = (r - c)\mu - ((c + h)Z(s) + (r + h + \pi)b_i(Z(s))) \times \sqrt{\frac{\sigma^2}{\rho} + \frac{1}{\rho}}, \] 

\[ \text{EL}_{\text{QR-sold}} = \mu - b_i(Z(s))\sqrt{\frac{\sigma^2}{\rho} + \frac{1}{\rho}}, \] 

\[ \text{EL}_{\text{QR-left}} = [Z(s) + b_i(Z(s))\sqrt{\frac{\sigma^2}{\rho} + \frac{1}{\rho}}. \]

With QR, we note that \( \text{EL}_{\text{QR-sold}} \geq \text{EL}_{\text{old-sold}} \) and \( \text{EL}_{\text{QR-left}} \leq \text{EL}_{\text{old-left}} \). Thus QR enables the retailer to decrease left over inventory and yet increase the customer fill rate.

The manufacturer-expected profit under the old system is obtained using Equation (1) as

\[ \text{EP}_{\text{old}} = (c - w)(\mu + Z(s)\sqrt{\sigma^2 + \tau^2}). \]  

(5)

The expected profit for the manufacturer under QR is obtained using Equation (3) as

\[ \text{EP}_{\text{QR}} = (c - w)\left\{ \mu + Z(s)\sqrt{\frac{\sigma^2}{\rho} + \frac{1}{\rho}} \right\}. \]  

(6)

3.3. When Is QR Pareto Improving?

A channel view requires that we compare the expected profits of manufacturers and retailers both before and after Quick Response. Our goal is to study if QR is in each individual channel member’s self-interest. We operationalize this concept with the Pareto criterion, which says that a new situation is Pareto improving if both parties are at least as well off, and one party is better off, after adopting QR. We begin by studying when QR will, and will not, be Pareto improving by itself.

**Lemma 1.** If \( s \leq \frac{1}{\rho} \) then under QR we have an increase in manufacturer and retailer expected profits, i.e., \( \text{EP}_{\text{QR}} \geq \text{EP}_{\text{old}} \) and \( \text{EP}_{\text{QR}} \geq \text{EP}_{\text{old}} \).

**Proof.** Observe that \( \sqrt{\sigma^2 + \frac{1}{\rho}} \leq \sqrt{\sigma^2 + \tau^2} \) and that when \( s = \frac{1}{\rho} \), \( Z(s) \leq 0 \). Hence the result using Equations (5) and (6). □

In Lemma 1 we show that under low service levels QR improves both manufacturer and retailer profits and is therefore Pareto improving. We would not expect to observe such low service levels in practice. Yet, our discussions with QR managers revealed two preexisting channel arrangements, consignment inventory and markdown money, that provide conditions which make QR Pareto improving. With consignment inventory, manufacturers are held responsible for product left over at the end of the season at the retailer. In this case, retailers pay holding costs through the season, but now have a salvage value for the product equal to its wholesale price \( c \). In the presence of this system, the manufacturer’s expected profit in the old system is \( (c - w)\mu - (wZ + c b_i(Z))\sqrt{\sigma^2 + \tau^2} \). Under QR, this
expected profit changes to \((c - w)\mu - (wZ_c + ch(Z_c))\sqrt{\sigma^2 + (1/p)}\) which is clearly larger than the expected profit in the old system. Markdown money is an amount the manufacturer pays to the retailer per item left over at the end of the season, and is used by the retailer to subsidize markdowns. It is usually set at a value below \(c\). If markdown money is high enough, then the savings for the manufacturer from reduced retailer end of season inventory may be large enough to make QR Pareto improving.

**Lemma 2.** If \(s \geq \frac{1}{2}\), then QR results in a decrease in expected manufacturer profit but an increase in expected retailer profits, i.e., \(EP_{QR,old} > EP_{old,old}\) and \(EP_{QR} > EP_{old}\).

**Proof.** If \(s \geq \frac{1}{2}\) then \(Z(s) \geq 0\). Verify from Equations (5) and (6) that \(EP_{QR,old} > EP_{old,old}\).

In Lemma 2, we show that when the service level for products is higher than \(\frac{1}{2}\), then QR will not be Pareto improving without some additional action. Specifically, QR is not profitable for the manufacturer. Retailer A as well the other QR managers we contacted stated that they explicitly considered manufacturer revenues under QR. Industry observers have also noticed manufacturer concerns with QR, stating that "(QR) activity in the (apparel) retail sector... has not sparked the same breadth of participation among apparel manufacturers" (Hammond 1990).

Similarly, the literature on fashion products suggests that the benefits of QR should be greatest for higher fashion items (Istasorda 1992), yet it is exactly those higher fashion products where manufacturers have been least interested in QR (Fisher and Raman 1992, Hammond 1990). This can be better understood in the context of our model. Pashigian (1988) defines fashion items as having greater uncertainty about the popularity of colors, patterns, and fabrics. In our model, because \(\tau\) reflects the ability to learn about customer demand, this implies that high fashion items would have a lower \(\sigma/\tau\) ratio than items with a lower fashion content. Note that as \(\sigma/\tau\) decreases, the total channel (i.e., sum of manufacturer and retailer) profits increase, which is consistent with the benefits being greatest for higher fashion products. Yet, as \(\sigma/\tau\) decreases the manufacturer expected profits decrease at an increasing rate. Thus higher fashion items pose the greatest potential loss on the manufacturer under QR.

The response of retailer A under QR was to undertake actions, often times commitments at time 0, in order to make QR Pareto improving. In the rest of the paper we focus on various commitments at time 0 that can be made by manufacturers and retailers to make Quick Response Pareto improving.

**4. Commitments Using Service Level**

We begin with a discussion of the role of service level commitments in making Quick Response Pareto improving. A service level commitment at time 0 implies that a higher service level \(s, s' > s\) is offered by the retailer, with the actual quantities ordered conditional on the estimated demand \(d_t\). These commitments are common in the apparel industry. A classic example is a commitment by Dayton Hudson to a planned 100% service level for vendors on a QR partnership (Dayton 1995). This is in contrast to the usual planned 90% service level for most other products.

A commitment is not considered credible unless it will be in the incentives of the retailer to take those actions at a later time. We thus model the change in service level by changing the goodwill cost \(\pi\). Commitments to variables that change goodwill costs to the retailer make the commitment to a higher service level credible from the manufacturer's perspective. Although negotiated changes in goodwill costs are not traditional in the production literature, it is often possible to influence the goodwill cost component \(\pi\) either contractually, or by making the in stock position of the product more salient to the consumer (see §4.1). Formally, we

\[\pi = \gamma(\pi_u, \pi_v)\]  

\[\pi_u = \pi_u + \pi_v\]  

\[\gamma(\pi_u, \pi_v)\]  

\[\pi_u\]  

\[\pi_v\]  

\[\gamma(\pi_u, \pi_v)\]
show that there always exists a service level arrangement that makes QR Pareto improving.

**Theorem 1.** If

\[ s \geq \frac{1}{2}, \quad d = \frac{\sigma}{\tau} \quad \text{and} \quad y = \sqrt{\frac{1 + (1/(1 + d^2))}{1 + (1/d^2)}} \]

then increasing \( \pi \) to \( \pi' \), where \( \pi' = (s'(r + h) + c - r)/(1 - s') \) and \( s' \) such that \( Z(s') = Z(s)/y \) makes QR Pareto improving.

**Proof.** See Appendix.

Note, however, that the effect of this service level arrangement is to make the retailer place the same expected order size as in the old system. The following lemma shows that under QR, even though the expected order size is the same as in the old system, the expected quantity sold to the customer increases and the expected quantity left over decreases, making the retailer better off.

**Lemma 3.** If

\[ s \geq \frac{1}{2}, \quad d = \frac{\sigma}{\tau} \quad \text{and} \quad y = \sqrt{\frac{1 + (1/(1 + d^2))}{1 + (1/d^2)}} \]

then increasing \( \pi \) to \( \pi' \), where \( \pi' = (s'(r + h) + c - r)/(1 - s') \) and \( s' \) such that \( Z(s') = Z(s)/y \), results in \( \text{EIQ} \text{sold}(s') = \text{EIQ sold} \text{old} \) and \( \text{EIQ left}(s') = \text{EIQ left} \text{old} \).

**Proof.** See Appendix.

4.1. Discussion of Industry Evidence

Apparel industry studies suggest service level increases from 80 to 95 percent associated with QR (Muller 1990). QR Pilot studies at Saks Fifth Avenue stores (a retailer of fashion merchandise) indicate service level improvements from 80 to 97 percent (Millson-Whitney 1993). In addition, discussions with managers at retailer A and QR managers at other retailers verified that service level improves under QR. We interpret these observed service level improvements as support for Theorem 1, i.e., the existence of service level commitments between manufacturers and retailers. This is because if the retailer could place orders under QR without having to worry about the channel then Equation 3 shows that it is optimal for the retailer to maintain the original service level \( s \). Thus any observed increases in service level do not automatically arise from QR, but must be the result of deliberate actions by the manufacturers and retailers.

Industry practice suggests a variety of ways to commit credibly to these higher service levels. One approach is a contractual commitment between the manufacturer and the retailer to provide a higher service level. Recall that Dayton Hudson’s QR contract with manufacturers states a planned service level of 100% for products supplied by vendors under a QR program. In other cases, the contracts provide additional discounts to the retailer for increasing service level. 4

Another approach is the use of cooperative advertising that uses service level guarantees. For example, Marshall Field’s (Chicago Tribune 1993) advertisements for Liberty shoes and Levi’s Dockers pants guarantee that the store will have every size and advertised color in stock, and if they do not they will get a pair for the customer at no charge. Carson Pirie Scott (Chicago Tribune 1995) ran similar advertisements for men’s blazers. Such advertisements increase the goodwill costs associated with a stockout and thus increase the retailer’s incentives to provide a higher service level. Managers at retailer A as well as all the QR managers we contacted found this scheme to be the most interesting, and stressed that these service level commitments could only be done profitably in situations with QR. 5

A third approach is to increase the salience of a stockout to consumers. This increased salience should increase the goodwill costs associated with a stockout and thereby increase the retailer’s incentives to provide a higher service level. 6 For example, Mercantile Stores

---

4 A specific provision might read as follows, “Value added discounts will be awarded upon field verification for services as follows: Maintain an inventory of our products that represents no less than sixty (60) days of estimated or historical annual sales—discount of 5%.” (Stern and El-Ansary 1988)

5 If advertising and promotion increase demand, this only increases manufacturer’s revenues and thus directly makes Quick Response Pareto improving

6 Increased attention to an item leads to increased memory of whether that item was stocked out, and therefore increased importance on the customers perception of that store’s service level (Dreze 1993, Kelly et al 1991)
provided their QR partner Warren Featherbone with a renewed emphasis on children's dresses by giving their items more prominent shelf locations and display space (Heinsen 1992). A second example is provided by the Norwalk furniture company, where Norwalk guarantees quicker response (in this case 30 days delivery) in return for the dealers displaying Norwalk furniture (Stern and El-Ansary 1988). This kind of negotiation was echoed by a buyer at retailer A who stated categorically that location in the store, attention and display were parameters negotiated with the manufacturer under QR.

5. Commitments Regarding the Wholesale Price

We now discuss the role of price commitments in making QR Pareto improving. Price commitments have often been suggested as ways of achieving channel coordination (Jeuland and Shugan 1983, Bergen et al. 1992, Kouvelis and Gutierrez 1992). We formally show how nonlinear price commitments can make QR Pareto improving in Lemma 4. We also show the limitations of simple price commitments under QR in Lemma 5.

Quantity discount schemes, a general form of nonlinear pricing, have been studied by Kouvelis and Gutierrez (1992) and Jeuland and Shugan (1983) as mechanisms to achieve channel coordination. We therefore examine the role of a Quantity discount scheme to make QR Pareto improving. The Quantity discount scheme we consider is an all units quantity discount that requires the retailer to pay a price of $c_i \geq c$ if quantity purchased is $< I_{d_{i1}}$ and a price of $c$ if quantity purchased is $\geq I_{d_{i1}}$. Under QR, the manufacturer essentially increases the wholesale price for quantities purchased below $I_{d_{i1}}$ to such a level that the retailer would prefer to buy $I_{d_{i1}}$ units at the price of $c$ rather than purchase the optimal quantity (given the estimate $d_i$) at a price of $c_i$.

**Lemma 4.** There exists a quantity discount scheme that makes QR Pareto improving.

**Proof.** See Appendix.

We note, however, that simple linear wholesale price commitments may not be sufficient to guarantee a Pareto improvement under QR. This is because as the wholesale price $c$ increases to $c'$ and all other parameters remain the same, the optimal service level offered to the customer by the retailer decreases. Lemma 5 shows that there exist sets of parameters for which increasing $c$ will not increase manufacturer expected profit.

**Lemma 5.** There exists a wholesale price $c^*$ that depends on system parameters such that if $c \geq c^*$, then there exists no flat wholesale price commitments that can make QR Pareto improving.

**Proof.** See Appendix.

These results are consistent with previous findings in the marketing literature. For example, previous research has shown that simple wholesale price changes are not sufficient to overcome traditional channel coordination problems, more complicated solutions such as nonlinear pricing are required (Bergen et al. 1992, and Jeuland and Shugan 1983).

5.1. Discussion of Industry Evidence

In our discussions with retailers, price commitments have played a limited role in QR. Most of the buyers contacted claimed no change in wholesale prices paid to QR manufacturers. Further, it is rare to find discussions of direct wholesale price increases for QR manufacturers in the literature.

About the only industry evidence we found supporting price negotiations was a quote by Bravman (1992) which states that under QR, potential "relief in manufacturer prices can be expected" QR managers also reported schemes whereby manufacturers now did more of the tasks such as tagging items, etc. and were paid for these tasks. This payment to the manufacturer to perform additional tasks may be considered a way to effectively pay manufacturers for QR. Industry reports confirm that many QR manufacturers are routinely performing these tasks (Stumhofer 1993) and claim that manufacturers were expected to be more efficient at these tasks than the retailer (Nannery 1995).

Our conclusion from the available industry evidence is that among the possible actions that can be used to
make QR Pareto improving, price adjustments are less likely to be used in the apparel industry. Discussions with managers at retailer A suggested that manufacturers often chose to negotiate on other dimensions, such as those discussed in the section on service level commitments or the section on volume commitments following, rather than negotiate on price.

6. Volume Commitments Across Products

We now discuss the role of volume commitments in making QR Pareto improving. Retailer A, as well as most of the QR managers we talked to, emphasized the use of volume commitments. Until this point in the paper the models have focused on one product. In order to model how volume commitments make QR Pareto improving we will need to incorporate additional products into the model.

Consider $M$ products, each of which has the same cost structure, the same values of $\mu$, $\sigma$ and $\tau$ and independent product demands. Under a volume commitment, the retailer at time 0 commits to buy $I(M)$ units across $M$ products. For the next $L_i$ periods, data is collected that allows the retailer to better assess demand by color, style, line, etc. Data estimated during $L_i$ (for item $i$ where $i = 1, 2, \ldots, M$) is used to choose the number of units of each item $i$ that will be purchased by the retailer. At time $L_i$, the retailer has to specify the exact quantities of each of the $M$ items such that the total number of units purchased is equal to $I(M)$.

To keep the supplier as well off as in the old system, the buyer offers commitments that are at least as large as in the old system.\(^8\) In the old system, the retailer would have purchased the same quantity for each item, i.e., $\mu + Z_i \sigma^2 + \tau^2$. Suppose the retailer commits to buying the same quantity across $M$ items at time 0 (i.e., commits $M\{\mu + Z_i \sigma^2 + \tau^2\}$). Formally, we will describe how the retailer can use this flexibility to order items to improve expected profit under QR, thereby making volume commitments Pareto improving.

**Lemma 6** The impact of a volume commitment of $I(M)$ on expected profit for the retailer is given by

$$\text{EP}_{QR}(I(M)) = (r + h)M\mu - (c + h)I(M)$$

$$- (r + h + \pi)M \sqrt{\frac{\sigma^2}{\rho} + \frac{1}{\rho} \int_{-\infty}^{\infty} \phi_\pi(z) b_i(z) dz}$$

where $\phi_\pi(z)$ is Normally distributed with mean $\mu_\pi$ equal to $(I(M) - M\mu) / M(\sigma^2 + (1/\rho))$ and a variance $\sigma^2$ equal to $\tau^2 / M(\sigma^2 + \tau^2)$. This implies a volume commitment of $M_i \mu + Z_i \sigma^2 + \tau^2$ across $M$ products makes QR Pareto improving.

**Proof.** See Appendix.

We will show that these volume commitments suggest observed service level increases and are therefore consistent with industry practice reported in §4.1. Volume commitments maintain the manufacturer’s profits but generate an expected service level which is below that required for a service level commitment. As $M$ increases, however, the volume commitment generates a service level improvement that tends to the service level under a service level commitment.

**Lemma 7.** The expected service level $s(M)$ under a volume commitment of $M_i \mu$ (as in Lemma 6) across $M$ products, $s(M)$, is $s(M) = s'$, where $s'$ is defined as in Theorem 1.

**Proof.** See Appendix.

It should be clear that without the constraint on the retailer to maintain manufacturer revenues under QR, it is optimal for the retailer to commit to an order that is lower than the old order quantity. This is intuitive because the volume commitment across products enables the retailer to get the benefit of demand pooling.

**Lemma 8.** The slope of

$$\frac{d\text{EP}_{QR}(I)}{dI} \bigg|_{I=I_{old}} \leq 0.$$ 

Thus the optimal value of $I(M)$ is $I_{old}$.

\(^8\) Managers at retailer A did suggest that another pricing aspect of QR was that it permitted manufacturers to offer value added service in lieu of price concessions to the retailer. In that case, QR enables the manufacturer to maintain wholesale prices.

\(^9\) An additional manufacturer benefit from a volume commitment under QR is the ability to order material which may have long lead times and which may be shared by the different SKUs. In some cases these commitments enabled the manufacturer to use the order as collateral to proceed with purchase of material, capacity etc. in advance of the season.

566 MANAGEMENT SCIENCE / Vol. 43, No. 4, April 1997

Copyright © 2001 All Rights Reserved
PROOF. See Appendix

6.1. Discussion of Industry Evidence
All of the buyers we contacted emphasized the use of explicit volume commitments to the manufacturer under QR. For example, at retailer A these commitments were made using purchase orders at time 0 specifying total volume but allowing specific attributes of the order such as color, silhouette or size, to be decided at L0. Other QR managers confirmed that they too explicitly considered the manufacturer’s revenues under QR when setting volume commitments at time 0. For example, at Boscov department stores, under QR they commit the same dollar volume to Philips Van Heusen as in the old system (Barrett 1993).

Further, all buyers we contacted agreed that the order commitment under QR was at least as large as the order commitment in the old system. This is particularly interesting given Lemma 8 shown earlier. Without the need to generate a Pareto scheme, the retailer might prefer to order a smaller quantity of the manufacturer’s products. This may explain why in some cases, the buyers added additional SKUs that were not formerly sold by the retailer so as to generate the same volume commitment (Foster 1993). For example, retailer A claimed to increase the number of SKUs carried from manufacturers in exchange for becoming QR providers in order to keep the total order commitment the same as in the old system.

7. Conclusions and Future Research
We have examined the impact of a channel view on Quick Response for the fashion apparel industry. Using formal inventory models that separate the effects of QR on each participant in the channel, we show that manufacturers may not be better off under QR. We discuss the use of service level commitments, wholesale price commitments, and volume commitments to make QR Pareto improving.

For a QR manager, we illustrate the impact of a channel view on the specific actions that have to be taken to implement QR. This view suggests channel constraints that must be explicitly considered. Often, these constraints lead to actions that are different from a system in which the channel does not play a significant role. We aid managers by modeling the role of three specific actions—service level, price, and volume commitments as tools for making QR Pareto improving. Further, our model allows an estimate of the extent of changes required under QR by linking them to specific data stored by the retailer. For example, the changes required to make QR Pareto improving get larger as the fashion content increases

This is a first step in understanding the impact of QR on a manufacturer-retailer channel. Future research issues include the impact of competition between manufacturers and retailers, the impact of multiple selling seasons and the role of relationships between manufacturers and retailers under QR.10

10 The authors acknowledge useful discussions with Professors Robert L. McCulloch and Jianming Ye at the University of Chicago and Professor Dean Foster at the University of Pennsylvania. We thank participants at workshops at Washington University at St. Louis, Northwestern University, University of Chicago, and the University of Minnesota for useful comments. We acknowledge feedback provided by Professor Marshall Fisher as well as three anonymous reviewers that greatly improved the presentation of the paper. The authors were also provided research support by the GSB, University of Chicago.

Appendix
The QR System
Under QR, given estimate  ̄, the retailer would choose the inventory level to maximize expected profits (denoted as  ̄ ) by setting  ̄ = ( ̄ ) + ( ̄ ). The expected profit realized by the retailer given  ̄ (denoted by  ̄ ) is

\[ EP_\text{QR}(d_i) = (r + h) \mu(d_i) + (r + h) \lambda_0(d_i) \]

The corresponding retailer expected order quantity (ELR) is

\[ ELR = \int_0^\infty \mu(d_i) m(d_i) \, dd_i \]

This simplifies to

\[ ELR = \mu + Z(\bar{\sigma}) \sqrt{\frac{\sigma^2 + 1}{\rho}} \]

Similarly, the expected profits under QR (EPQR) are obtained as

\[ EP_{QR} = \int_0^\infty EP_{QR}(d_i) m(d_i) \, dd_i \]

This simplifies to

\[ EP_{QR} = (r + h) \mu + (r + h + \tau) Z(\bar{\sigma}) \sqrt{\frac{\sigma^2 + 1}{\rho}} \]

The expected quantity sold in the season given the estimate  ̄ during  ̄ is

\[ L_{QR-\text{est}}(d_i) = \int_{y_0}^{y_0 + h} u m(d_i) \, dy + \int_{y_0}^{y_0 + h} \lambda_0(d_i) m(d_i) \, dy \]

\[ = \mu(d_i) - \lambda_0(Z(\bar{\sigma}) \sqrt{\frac{\sigma^2 + 1}{\rho}}) \]

MANAGEMENT SCIENCE/Vol. 43, No. 4, April 1997
Unconditioning on $d_i$, we get the expected quantity sold during the season as

$$E_{QR-nal} = \int_{s}^{\infty} l_{QR-nal}(d_i)m(d_i)dd_i = \mu - b_i(Z(s))\sqrt{\frac{\sigma^2 + 1}{\rho}}$$

Following a similar analysis, we get the expected quantity left over as

$$E_{QR-bn} = (Z(s) + b_i(Z(s)))\sqrt{\frac{\sigma^2 + 1}{\rho}}$$

Theorem 1 If

$$s \geq \frac{1}{2}, \quad d = \frac{\sigma}{\tau} \quad \text{and} \quad y = \sqrt{\frac{1 + (1/(1 + d^2))}{1 + (1/d^2)}}$$

then increasing $\pi$ to $\pi'$, where $\pi' = (s'(r + h) + c - r)/(1 - s')$ and $s'$ such that $Z(s') = Z(s)/y$ makes QR Pareto improving.

Proof To maintain manufacturer expected profits, we require that $E_{nal}\neq E_{QR-nal}$. We guarantee this relation by changing the service level from $s$ to $s'$ under QR such that $Z(s') = Z(s)/y$, where $s' = \Phi(Z(s'))$, $\Phi(Z(s'))$ is the cumulative density function of a standard Normal distribution,

$$y = \sqrt{\frac{1 + (1/(1 + d^2))}{1 + (1/d^2)}}, \quad d = \frac{\sigma}{\tau}$$

(note that $y \approx 1$)

We will implement this new service level at the retailer by changing $\pi$ to $\pi'$ such that $s' = (r + \pi' - c)/(r + \pi' + h)$. We have to verify whether the retailer expected profits under this service level $s'$ and QR are greater than in the old system. Verify that from Equations (2) and (4), we have to show that

$$\frac{b_i(Z(s))}{b_i(Z(s)/y)} \geq \frac{(r + h + \pi')y}{r + h + \pi}$$

which simplifies to showing that

$$v(Z(s)) = \frac{Z(s)}{y}$$

where $v(Z)$ is the probability density function of a standard normal distribution.

Note that it is sufficient to show that $d v(Z)/dZ \leq 0$ for all $Z \geq 0$.

Also, from Chung (1968), we know that

$$1 - \Phi(Z) \geq \frac{Z\Phi(Z)}{1 + Z^2} \quad \text{for all } Z > 0$$

Thus we have

$$v(Z) = \frac{1 + Z^2}{Z^2} \quad \text{and} \quad \frac{dv(Z)}{dZ} \leq 0$$

for all $Z \geq 0$. The retailer expected profits under QR and the service level $s'$ are greater than or equal to the retailer profits under the old system. Hence QR is Pareto improving with this change in service level from $s$ to $s'$. □

Lemma 3 If

$$s \geq \frac{1}{2}, \quad d = \frac{\sigma}{\tau} \quad \text{and} \quad y = \sqrt{\frac{1 + (1/(1 + d^2))}{1 + (1/d^2)}}$$

then increasing $\pi$ to $\pi'$, where $\pi' = (s'(r + h) + c - r)/(1 - s')$ and $s'$ such that $Z(s') = Z(s)/y$, results in $E_{nal}(s') \geq E_{nal}(s)$ and $E_{QR-bn}(s') \geq E_{QR-bn}(s)$.

Proof To show that $E_{nal}(s') \geq E_{nal}(s)$ we have to show that

$$\mu - b_i(Z(s))\sqrt{\frac{\sigma^2 + 1}{\rho}} \geq \mu - b_i(Z(s)/y)\sqrt{\frac{\sigma^2 + \tau^2}{\rho}}$$

This implies that $b_i(Z(s))/b_i(Z(s)/y) \approx y$. This is clearly true from Theorem 1 (note that $\pi' \approx \pi$). Hence $E_{nal}(s') \geq E_{nal}(s)$. Since we know that $E_{nal}(s') = E_{nal}(s')$, we see that $E_{nal}(s') = E_{nal}(s')$. □

Lemma 4 There exists a quantity discount scheme that makes QR Pareto improving.

Proof Consider the estimated demand $d_i = 0$ and the associated conditional mean of $\mu(0) = \sigma\mu/(\sigma^2 + \tau^2)$ and standard deviation $\sigma' = \sqrt{\sigma^2 + (1/\rho)}$. Identify the wholesale price $c_i$ such that it is better for the retailer to buy $l_{nal} = \mu + Z(s)\sqrt{\sigma^2 + \tau^2}$ at $c_i$ rather than buy $\mu(0) + Z(s)\sigma'$ at $c_i$. Hence $c_i = c_i$ is the minimum value of $c_i$ such that

$$\begin{align*}
(r - c_i) \times \mu(0) - (r + h) \times Z(s) + (r + \pi + h) \times b_i(Z(s)) \sigma' &
\leq (r + h) \times \mu(0) - (r + c + h) \times l_{nal} - (r + h + \pi) \sigma b_i(l_{nal} - \mu(0))
\end{align*}$$

Observe that if the manufacturer were to set the wholesale price as $c_i$ for orders $l_{nal}$ and $c$ for orders greater than $l_{nal}$, it is optimal for the retailer to always order at least $l_{nal}$ for any observed $d_i$. Hence this quantity discount scheme makes QR Pareto improving. □

Lemma 5 There exists a wholesale price $c_i$ that depends on system parameters such that if $c \approx c_i$, then there exists no flat wholesale price commitments that can make QR Pareto improving.

Proof Note that given that the retailer adjusts service level in response to the wholesale price $c_i$, the manufacturer expected profit is

$$(r + \pi) - (\Phi(Z) \times (r + \pi + h)) - w \times (\mu + (Z \times \sigma'))$$

where

$$\sigma' = \sqrt{\sigma^2 + \frac{1}{\rho}} \quad \text{and} \quad \Phi(Z) = \frac{r + \pi - c_i}{r + \pi + h}$$

Thus we have
To identify the optimum value $Z^*$ that maximizes this manufacturer profit, differentiate the above term with respect to $Z$ and set it equal to zero which yields

$$\Phi(Z^*) \sigma^* + \Phi(Z^*) \times (\mu + (Z^* \times \sigma^*)) = \frac{r + \pi}{r + \pi + h} \times \sigma^*,$$

then solve for $Z^*$

If the cost $c$ in the old system is greater than $c^*$, where $c^* = (r + \pi) - \Phi(Z^*) \times (r + \pi + h)$, then increasing the wholesale price $c$ will not increase manufacturer revenues and thus not make QR Pareto improving.

**Lemma 6** The impact of a volume commitment of $M$ on expected profit for the retailer is given by

$$\text{EP}_{\text{QR}}(I(M)) = (r + h)M \mu - (r + h)I(M) - (r + h + z)M \int \phi(z) h(z) dz$$

where $\phi(z)$ is normally distributed with mean $\mu$, equal to $(I(M) - M\mu)(\sigma^2 + (1/\rho))$ and a variance $\sigma^2$ equal to $r^2/M(\sigma^2 + \tau^2)(\sigma^2 + (1/\rho))$. Thus a volume commitment of $M(\mu + Z\sigma^2 + \tau^2)$ across $M$ products makes QR Pareto improving.

**Proof** Given that we observe $d_{i,t} = 1, 2, \ldots, M$, and a constraint that the total order should be $I(M)$, the optimal purchase quantity is that which generates the same service level for every $i$, i.e., $(I(M) - \Sigma_i d_{i,t})/M \sigma^2 + (1/\rho)$, where $\mu(d_{i,t})$ is as defined in §2.1. Given the distribution of $\Sigma_i d_{i,t} = N(M\mu, M(\sigma^2 + \tau^2))$, we see that the distribution of the $Z$ value, i.e., $\Phi(Z)$ is $N(\mu, \sigma^2)$ as defined in the lemma. Hence the expected profit is as in the statement of this lemma. If we set $I(M) = M(\mu + Z\sigma^2 + \tau^2)$, this can be verified to make QR Pareto improving by checking that $\int \phi(z) h(z) dz$ is nonincreasing in $M$.

**Lemma 7** The expected service level $s(M)$ under a volume commitment of $M\lambda d$ (as in Lemma 6) across $M$ products, $s(M)$, is $s = s(M) = s'$, where $s'$ is as defined in Theorem 1.

**Proof** Given the distribution of the $Z$, i.e., $\Phi(Z)$ as in Lemma 6, the associated average service level is $\int \phi(z) h(z) dz$ where $\phi(z)$ is the probability density function of a standard normal distribution and $\Phi(z)$ is the cumulative density function of a normal distribution with mean $\mu$ and variance $\sigma^2$. It is clear that this value is equal to $s$ when $M = 1$ and is equal to $s'$ when $M \to \infty$. It can be verified after some calculations that the inequalities hold.

**Lemma 8** The slope of

$$d\text{EP}_{\text{QR}}(I)/dt \bigg|_{\lambda = \lambda'} = 0$$

Thus the optimal value of $I(M)$ is $\lambda = \lambda'$.

**Proof** The slope with respect to $I$ evaluated at $M \times \lambda_{\text{opt}}$ is equal to

$$-(c + h) - (r + \pi + h) \pi \int \phi(z) dz$$

Since we know that when $M = 1$, this value is $0$, we have to show that this value is $\leq 0$ for $M > 1$. This requires that we verify (after some calculations) that $M \int \phi(z) dz$ decreases with $M$.

**References**


Nannery, Matt, "Distribution Center Shows Fred Meyer's QR Commitment," *DNR*, March 15, 1995
Signorelli, Sergio and James L. Heskett, "Benetton (A)," Harvard Business School, Case # 9-685-014, Cambridge, MA, 1984

Accepted by Marshall Fisher, received February 2, 1994. This paper has been with the authors 5 months for 2 revisions.